Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Plant Biology

Using mustard genomes to explore the genetic basis of evolutionary change

Lachezar A Nikolov and Miltos Tsiantis

Recent advances in sequencing technologies and gene manipulation tools have driven mustard species into the spotlight of comparative research and have offered powerful insight how phenotypic space is explored during evolution. Evidence emerged for genome-wide signal of transcription factors and gene duplication contributing to trait divergence, e.g., *PLETHORA5/7* in leaf complexity. Trait divergence is often manifested in differential expression due to cis-regulatory divergence, as in *KNOX* genes and *REDUCED COMPLEXITY*, and can be coupled with protein divergence. Fruit shape in *Capsella rubella* results from anisotropic growth during three distinct phases. Brassicaceae exhibit novel fruit dispersal strategy, explosive pod shatter, where the rapid movement depends on slow build-up of tension and its rapid release facilitated by asymmetric cell wall thickenings.

Address
Department of Comparative Development and Genetics, Max Planck Institute for Plant Breeding Research, Cologne 50829, Germany

Corresponding author: Tsiantis, Miltos (tsiantis@mpipz.mpg.de)

Current Opinion in Plant Biology 2017, **36**:119–128

This review comes from a themed issue on **Genome studies and molecular genetics**

Edited by **Ian Henderson** and **Korbinian Schneeberger**

For a complete overview see the Issue and the Editorial

Available online 8th March 2017

http://dx.doi.org/10.1016/j.pbi.2017.02.005

Introduction

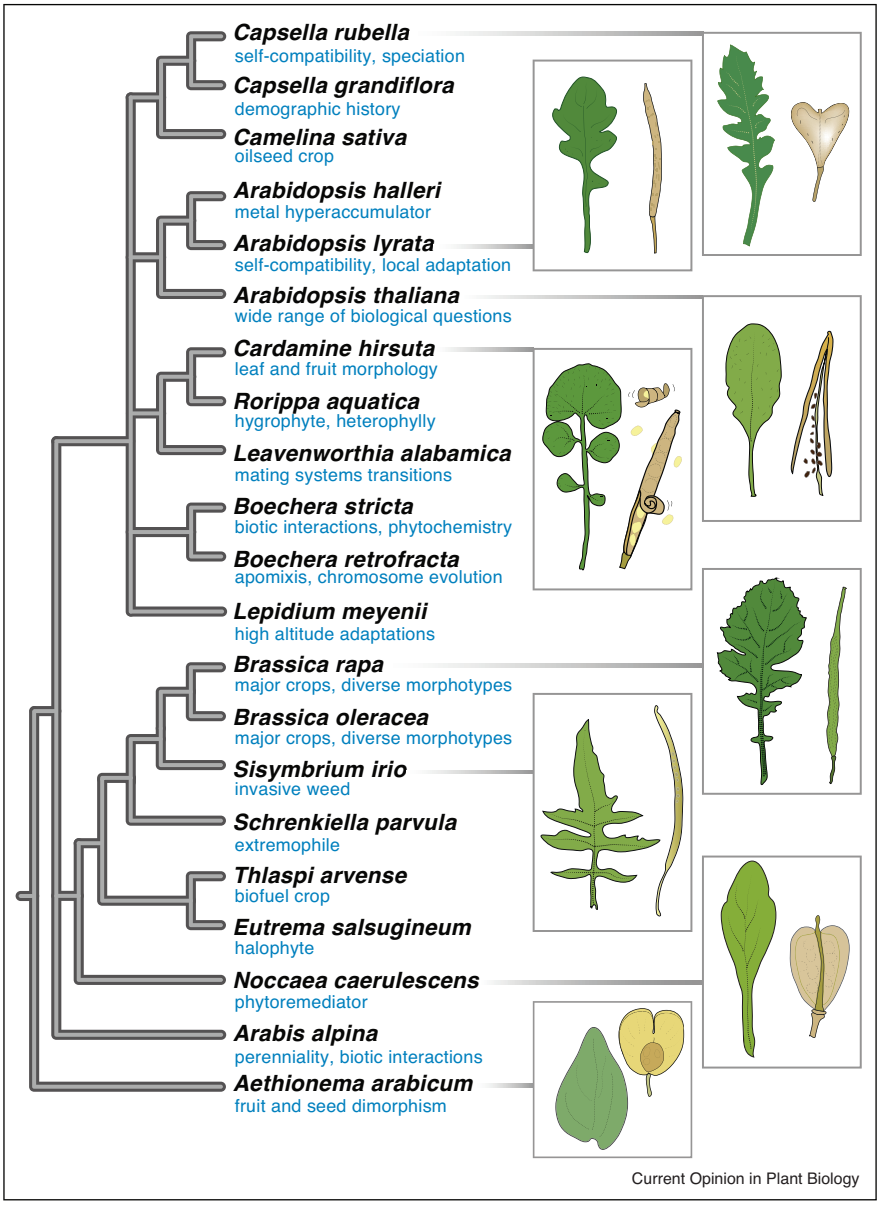
The mustard family (Brassicaceae) holds a special place in the plant kingdom as it contains *Arabidopsis thaliana*, a key reference species for the green evolutionary lineage [1]. Reference species provide powerful functional insight that, when put in a comparative perspective, can also illuminate novel aspects of biology not observed in the model alone. The mustards offer excellent opportunities to address the genetic basis of phenotypic change at a phylogenetic scale where unique biological questions not accessible in *A. thaliana* can be studied with mechanistic rigor using resources pioneered for this important model over the last four decades. The Brassicaceae exhibit considerable diversity at the morphological, physiological, and biochemical level [2] (Figure 1). A robust phylogenetic framework, which is a prerequisite for the comparative approach, is coming into focus and currently comprises 49 tribes [3] and several monophyletic subfamilial groupings [4]. Brassicaceae is the plant family with the highest number of species with sequenced genomes. Many Brassicaceae species are genetically and biochemically tractable. Collectively, these qualities create a powerful platform to link genotype to phenotype and address how diverse traits evolve. Here we focus on recent studies employing comparative genomics to summarize patterns of genomic variation and to understand the genetic basis of phenotypic change in mustards.

Genomic advances at different evolutionary scales

Mustards on the map: intra-specific and inter-specific variation

After the completion of the reference sequence of *A. thaliana* (Col-0) [5], much effort has been invested to re-sequence accessions throughout its geographic range to gain insight into the genomic variation and evolutionary history of the species. The most recent report included 1135 naturally inbred lines and featured the densest variant map available for any complex eukaryote [6•]. It provided unprecedented resolution of the pattern of polymorphisms, insight into the demographic history of the species, and highlighted the formative role of climate (in particular, the last glacial period) and geography (Iberian Peninsula as a hotspot of relict populations surviving since the last ice age) in shaping the current genetic variation of *A. thaliana*. The study revealed that although the relict populations continue to inhabit ancestral habitats, the descendants of one as-yet unidentified ancestral population have expanded broadly and this may have resulted in rapid fixation of genetic variants due to founder effects. Subsequent divergence was reinforced by species-wide genetic incompatibilities. *NLR* immune receptor genes were among the most influential players in such incompatibilities, and may have caused gene flow barriers through autoimmunity-driven hybrid necrosis in intraspecific hybrids [7]. Other immunity-related genes, the homologs of *NON-EXPRESSOR OF PR-GENES1* and *RECOGNITION OF PERENOSPORA PARASITICA 5*, interacted epistatically to cause an autoimmune response and gene flow barriers when combined in hybrids of different *Capsella* species [8•].

Figure 1



Phylogenetic relationships of Brassicaceae species with abundant sequencing resources. Specific questions addressed using these models are listed in blue. Panels depict diversity in leaf and fruit morphology of selected species.

At a higher taxonomic level, coding sequence variation in the genus *Arabidopsis* was harnessed to reveal support for seven clusters corresponding to the currently recognized seven *Arabidopsis* species [9•]. The relationship among species based on individual gene trees varied significantly likely due to extensive hybridization, and only *A. thaliana* could be reliably placed as a sister to the rest of the genus [10]. Interestingly, allele distribution test (ABBA-BABA) taking into account population structure showed that *A. thaliana* was more similar to *Arabidopsis lyrata* than to *Arabidopsis arenosa*, which is unexpected for its sister position. This suggests that *A. thaliana* became reproductively isolated from the outcrossing *Arabidopsis* species more recently than previously thought and its placement may reflect fixation of genetic variation derived from ancient admixture. In this case, the unique genomic features of *A. thaliana*, like the basic chromosome number of 5, and its derived traits, such as self-compatibility and annual habit, must have arisen more rapidly than previously thought. This study also identified a set of 129 genes containing shared ancestral polymorphisms in at least two of the three most common *Arabidopsis*

species, which were enriched for viral-related functions and could indicate balancing selection.

Gene discovery by *de novo* assembly

Although single nucleotide polymorphism data provide a powerful window into the distribution of small-scale sequence variation, some genomic features, like collinear and rearranged variation, require high-quality genomes assembled without reference guidance. The genome of another commonly used ecotype of *A. thaliana*, Landsberg *erecta*, was assembled from short-reads and PacBio sequencing data using a genetic map scaffold [11]. The complete assembly revealed large-scale rearrangements including 1.2-Mb inversion on chromosome 4 and several hundred unique genes, along with genes exhibiting accession-specific duplication or an asymmetric loss of a paralog. Such dynamic gene gain and loss processes have been documented broadly throughout Brassicaceae and underscore the importance of 'total evidence' in genomics: the pangenome as the sum of core genes shared by all accessions and accession-specific genes. In *A. thaliana* in particular, lineage-specific genes arose after duplication and subsequent divergence, retrotransposition, and unequal crossing over [12]. More rarely, they resulted from overprinting (creating new genes by accumulation of mutations after frameshift in an existing parental gene) and transposon exaptation. The relative contribution of these processes in shaping the genomes of other mustards remains unexplored.

Despite some degree of lineage-specific genic composition, sequence conservation, broadly homologous gene complement, and large-scale collinearity extending to massive genomic blocks characterize all sequenced mustard genomes [13–16]. These features facilitate comparative studies that tease apart the functional role of specific genomic regions. Phylogenetic footprinting across the family demonstrated that a significant portion of the Brassicaceae genome is under selection [17]. Approximately one fourth of the signal could be localized to *ca*. 90 000 conserved noncoding sequences, which are implicated in the regulation of gene expression, and were present in some of the nine analyzed genomes. As more genomes become available, identifying clade-specific conserved noncoding sequences that account for shared phenotypic traits becomes feasible. Phylogenetic genome-wide association approach (PhyloGWAS) that correlates phenotypes or broad-scale environmental factors with standing genetic variation regardless of phylogenetic relatedness also holds promise at shallower phylogenetic scales [18].

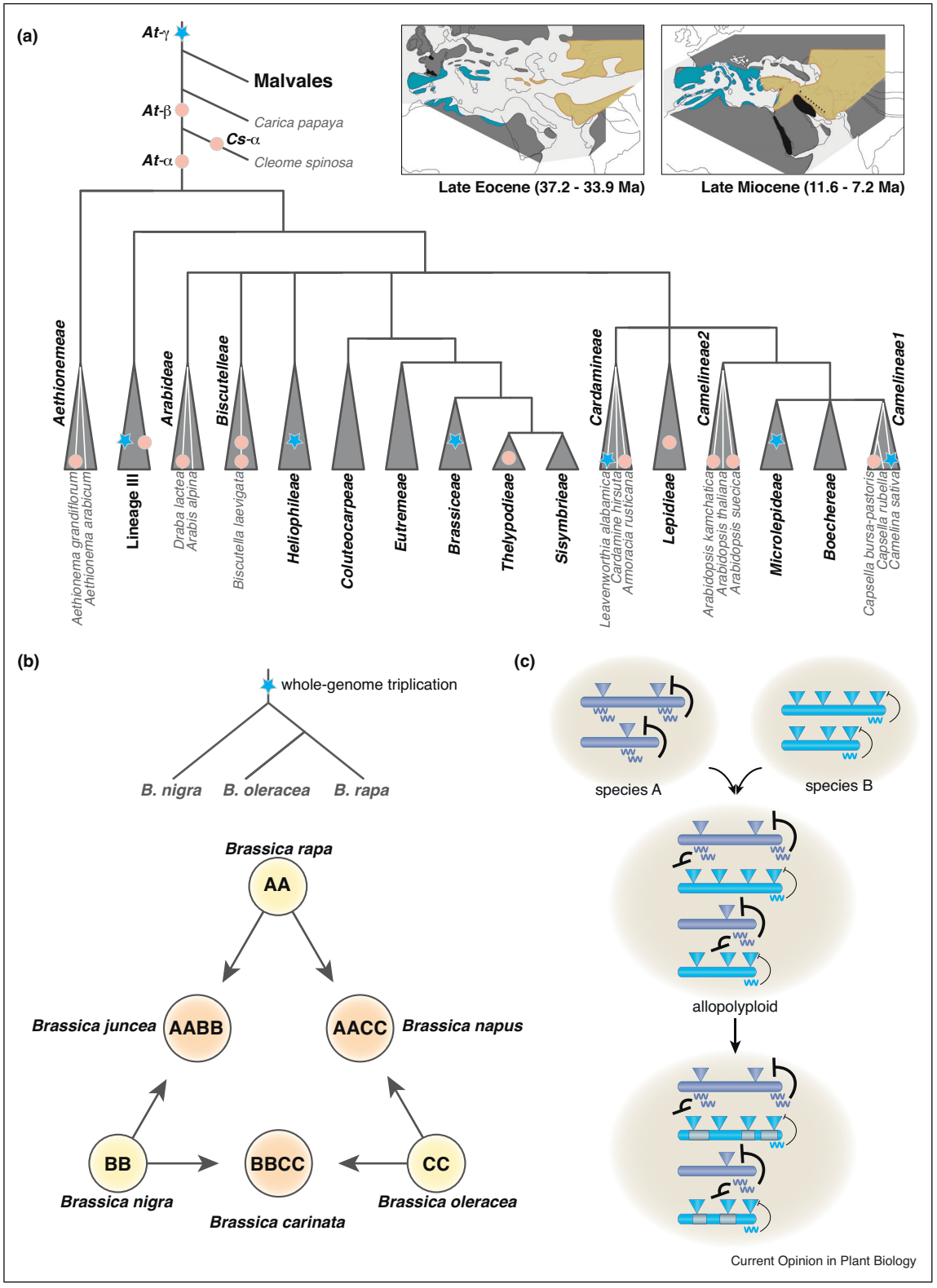
Comparable catalogs of the gene complements of mustard species with distinct phenotypes can be used to correlate traits with gene family expansions and contractions. For example, the genome of *Cardamine hirsuta* [19•] features an overrepresentation of transcription factors among its expanded or unique gene families compared to seven other mustard species. Comparative leaf transcriptomes of *A. thaliana* and *C. hirsuta* revealed similar overrepresentation of transcription factors and tandemly duplicated genes among the differentially expressed genes between these two species [19•]. *A. thaliana* and *C. hirsuta* exhibit markedly different leaf shapes and such changes may have contributed to trait divergence between these two species. This comparative transcriptomic approach led to the discovery of novel leaf shape regulators, the stem cell fate transcription factors *PLETHORA5 (PLT5)* and *PLT7*, which have not been previously implicated in interspecific divergence. *PLT5/7* were shown to be necessary (and *PLT7*—sufficient in *A. thaliana*) to increase leaf complexity in the complex leaved *C. hirsuta* [19•]. Clade-specific expansions with potentially adaptive roles were detected in the metal hyperaccumulation pathways [20], stress-tolerance genes [21], the photoperiod regulator *CONSTANS* [22], and some glucosinolate synthesis genes [23]. In the heavy metal hyperaccumular *Noccaea caerulescens*, there is also evidence that structural rearrangements may have played a role to group together functionally similar genes to unify the regulation of gene expression [24].

Genomic processes shaping Brassicaceae diversity

Series of genome duplications create nested diversity in the gene space

Brassicaceae is one of the largest families of angiosperms and it is relatively young [25]. There are some large genera (*e.g.*, *Draba*, *Erysimum*, *Lepidium*, *Cardamine*, *Alyssum*), which collectively account for about one third of the species but the overall phylogenetic diversity in the family is high and this correlates with elevated morphological disparity. The family's diversification coincides with cooling during the Miocene (23–5.3 Mya), which may have created new open and drier habitats (Figure 2a) [26,27]. The intrinsic mechanisms that enabled this diversification are the focus of intensive investigation. One possible contributing factor is polyploidy (Figure 2a) [28, but see 29]. In addition to the well-established paleopolyploidy events shared by all members of Brassicaceae: At-α (at the base of the Brassicaceae), At-β (within Brassicales after the divergence of papaya), and At-γ (hexaploidy or two consecutive genome duplication events at the base of the core eudicots), there are numerous lineage-specific mesopolyploidy and neopolyploidy events. This pattern of 'nested duplications,' when a duplicated genome undergoes a subsequent round of duplication, creates a hierarchy of gene copies with different degree of relatedness. At least five independent whole-genome duplication events during the Cenozoic may have conferred higher resilience during climate change [30]. The karyologically diploid species in the tribe Brassiceae are derived from whole-genome triplication, likely in a two-step process, and subsequent

Figure 2



The patterns and processes of polyploid evolution in Brassicaceae. **(a)** Polyploidization events mapped on the phylogeny of Brassicaceae (tree adapted from Ref. [4]). Three whole-genome polyploidization events (At-α, At-β, and At-γ) preceded the basal split in the family and are shared by all mustard species; Cs-α is an independent polyploidization in the sister family of Brassicaceae, Cleomaceae. Although almost all tribes include a significant number of more recent polyploids, the similarity in genome size and gene content among the sister lineage of the core mustards

fractionation to produce diploid karyotypes [31]. Similar results were obtained by chromosome painting for species in the South African tribe Heliophileae (mesopolyploid triplication) [32], and the Australasian Microlepideae (mesotetraploid event) [33] despite the low chromosome numbers (n = 4–6) observed in this group. *Biscutella laevigata* has undergone mesopolyploidization and a more recent neopolyploidization event, from which it retains a number of stress-responsive genes that may have adaptive value [21]. Cycles of polyploidization were implicated in the escalating evolutionary arms race between mustards and herbivorous butterflies, an interaction with broad consequences on diversity in both groups, by increasing the number of glucosinolate synthesis genes and the complexity of the pathway [34].

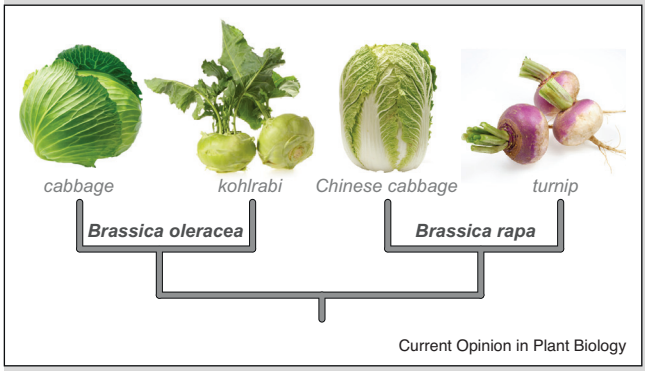
Genome dominance results in biased homeolog retention

Some species, for example the recent hexaploid *Camelina sativa*, exhibit little bias in terms of gene retention and gene expression of paralogs [35]. Other, more ancient polyploids, like the *Brassica* crops, have biased homeolog retention and expression [36]. *Brassica* crops include three diploid species (*Brassica rapa* [AA], *Brassica nigra* [BB], and *Brassica oleracea* [CC]), each a result of prior triplication, and their allopolyploid hybrids *Brassica napus* [AACC], *Brassica juncea* [AABB], and *Brassica carinata* [BBCC] (Figure 2b; Box 1). In *B. rapa*, the fractionation of each of the three ancestral subgenomes varied from 70% retained genes in the least fractionated subgenome to 36% and 46% retention in the other two subgenomes [37]. Similar asymmetric gene loss among the ancestral subgenomes was observed in *B. oleracea*. The least fractionated subgenome was dominant over the other two subgenomes. Among retained homeolog pairs, the genes from the dominant subgenome were more highly expressed than the homeologs from the more fractionated subgenomes and potentially experienced higher purifying selection. This phenomenon appeared heritable for millions of years [38]. One tantalizing hypothesis suggested that it resulted from gene silencing after transposon-guided DNA methylation via 24-nt RNAs, which preferentially targeted regions in the recessive subgenomes (Figure 2c) [38]. Because separate subgenomes harbor unique transposon populations and are adapted to maintain a tradeoff between their silencing and the silencing of nearby genes (positional effect) [39], combining two subgenomes into one polyploid genome led to preferential targeting of the subgenome with weaker anti-transposon defense or more transposable elements. The genes in this subgenome were silenced preferentially, and subsequently deteriorated due to relaxed selection.

Epigenetic differences within and between species may play a role in other processes beyond genome dominance.

Box 1 Convergence in morphology after domestication.

Parallel leaf-heading and tuber-forming morphotypes in *B. rapa* and *B. oleracea* have been domesticated separately in different cultures and at different localities. Genome scans of heading and non-heading accessions in both species and the subgenomes of each species separately identified a number of shared genomic regions with reduced diversity as a possible sign of selection [68]. Some of these regions included putative regulators of adaxial-abaxial leaf polarity and leaf cell proliferation, which may have contributed to leaf curling in the heading phenotypes, cabbage and Chinese cabbage, and which were expressed at higher level compared to their paralogs within each genome. Evidence for convergence was also detected in genes associated with sugar metabolism and root development in the tuber-forming varieties. For example, members of the expansin gene family may have been selected separately in turnip (*B. rapa*) and kohlrabi (*B. oleracea*). These observations can be explored further by explicitly simulating the demographic history of *Brassica* crops to obtain a null model of the neutral processes shaping nucleotide variation patterns. Such models can be used to determine the probability of shared regions with reduced diversity arising under neutrality and under selection [69,70].



(Figure 2 Legend Continued) (*Aethionema arabicum*), *A. thaliana* (Lineage I), *Sisymbrium irio* (Lineage II), and *Arabis alpina*, which collectively span the phylogenetic breadth of the family suggests that (1) there have not been additional polyploidization events along the backbone of the phylogeny leading to these species, and (2) the additional polyploidization events likely occurred along more terminal branches (leading to tribes and possibly subtribal groupings). Broader sampling, especially in Lineage III, will reveal whether certain lineages share polyploidization events. Single polyploidization events are marked by a pink dot and polyploidization events of higher order, for example the core eudicot hexaploidization event At-γ, are denoted with blue stars. Maps in the upper right corner show changes in the landmasses forming the Irano-Turanian (yellow) and Mediterranean (blue) floristic regions, both suspected cradles of mustard diversity, between the Late Eocene and the Late Miocene when many mustard species evolved and diversified (adapted from Ref. [71] after Ref. [72]). **(b)** Phylogenetic relationships among the diploid *Brassica* crops, which nonetheless share an older genome triplication, and the Triangle of U depicting the ancestry of the three allopolyploid species. **(c)** Model explaining the origin of genome dominance. Hybridization/polyploidization combines two parental genomes (light and dark blue) adapted to different balance of transposon activity and silencing of neighboring regions driven by small RNAs (blocking arrows). Over time, a dominant subgenome (dark blue) emerges, which is more successful in preventing transposon population expansion and the subsequent spread of heterochromatin and gene silencing. The preferentially silenced genome deteriorates more rapidly due to relaxed selection (gray boxes). Triangles represent transposon insertions and wavy lines—small RNAs.

Comparing the epigenomes of *A. thaliana* accessions throughout its geographic range revealed changes in methylation status of approximately one-fourth of the genes [40•]. Some of the most variable genes were immunity-related, with sequence variation, epigenetic status changes, and transposon insertions all contributing to their differential expression. Epigenetic differences are particularly pronounced between species. In *A. thaliana* and *C. rubella*, rapid evolution of epigenetic status was mirrored by high diversity and rapid divergence of miRNAs and miRNA targets, and DNA methylation, which was largely attributed to lineage-specific expansion and contraction of transposable elements [41–43]. The significance of these findings for phenotypic diversity is still unclear because variation was confined to the hypervariable genomic regions and thus hardly accessible to natural selection. Evidence from *Capsella grandiflora* suggests that whether gene body methylation constrains *cis*-regulatory variation, the proximity of transposable elements and tissue-specific expression increase the likelihood of allele-specific expression in natural populations of the species [44]. The role of methylation in shaping the genome is profound. In *Arabis alpina*, reduced symmetric CG and CHG DNA methylation may have resulted in increased retrotransposition activity and the transformation of previously euchromatic regions into repeat-rich pericentromeric regions [45•]. *DNA METHYLTRANSFERASE 1* and *VARIANT IN METHYLATION 1*, two essential components of the CG methylation maintenance pathway in *A. thaliana*, lacked clear orthologs in *A. alpina* and this finding may explain the observed DNA methylation maintenance deficiency. Mobile shoot-derived small RNAs regulated the DNA methylation status of thousands of loci in the root of *A. thaliana* [46]. Such shoot-derived RNA signals of one accession moved across a graft union and could target normally non-methylated sites in roots of different accession for DNA methylation, suggesting the coordination of methylation status in different plant organs [46].

Phenotypic divergence at different evolutionary scales

In addition to providing a more comprehensive view of the phenotypic landscape, the comparative approach reconstructs how the landscape is explored to shed light on the repeatability of evolution. Empowered by more robust and more extensive phylogenetic estimates, character mapping showed that similar leaf and fruit morphologies have arisen independently in different lineages of Brassicaceae, highlighting the prevalence of homoplasy and suggesting similar solutions to common adaptive pressures in the family [4].

Leaf diversity

Leaf morphology is extensively studied in a comparative context because the trait is remarkably diverse, from simple leaves with nearly smooth margins, as in *A. thaliana*, to complex leaves with individual leaflets attached to a rachis by defined petiolules, as in *C. hirsuta* [47]. Growth at pre-patterned foci along the leaf margin combined with growth repression between the foci sculpts complex leaves [48]. Growth at the margin requires local peaks of auxin activity that are organized by the *KNOX* genes *SHOOT MERISTEMLESS* (*STM*) and *BREVIPEDICELLUS* (*BP*), and *CUP-SHAPED COTYLEDON* (*CUC*) genes [48,49]. Separating these protrusions to form individual leaflets depends on additional growth repressors, such as *REDUCED COMPLEXITY* (*RCO*) [50].

Cis-regulatory divergence of *STM* and *BP* in *A. thaliana* eliminated their expression in simple leaves but when reintroduced into leaves, both STM and BP proteins modify the leaf shape of *A. thaliana*, albeit to a different extent [51]. The STM protein changes leaf shape more strongly than the BP protein, and at the same time, *stm* mutants are more pleiotropic than *bp* mutants. However, the entire genomic *BP* locus of *C. hirsuta*, which includes the coding sequence under the expression of endogenous regulatory sequences, elicits a more profound morphological change when moved by transformation into *A. thaliana* than the *STM* locus of *C. hirsuta*. Taken together, these findings suggested an inverse relationship between gene's pleiotropy and its potential to change leaf shape. Moreover, *C. hirsuta BP* provided a functional link between two largely isolated regulatory modules in *A. thaliana*, that of *ASYMMETRIC LEAVES1* and of *CUC2/MIR164A*, which diversified genetic interactions and opened new functional space for exploration during morphological evolution [51]. *KNOX* genes were used repeatedly to modify leaf shape in other lineages besides the mustards, including tomatoes [52] and tropical lianas in the family Bignoniaceae [53].

KNOX gene expression is not confined to leaves and imposes a pleiotropic burden on plant development, which may restrict their ability to affect morphology [51]. Therefore, the discovery of *RCO*, a HD-ZIP transcription factor specifically expressed in developing leaves of *C. hirsuta*, provided a prominent example for a major effect molecular player driving the morphological divergence between species [50]. *RCO* arose after tandem duplication following the earliest split in Brassicaceae, which gave rise to a cluster of two to three genes in different mustard species, including *RCO*'s paralogs *LATE MERISTEM IDENTITY1* (*LMI1*) and *LMI1*-like in *C. hirsuta*. After species-specific deletion, *A. thaliana* retained only *LMI1*, which may have contributed to its simple leaf shape. *LMI1* genes are expressed more distally in the leaf, similarly to the pre-duplication gene in the sister lineage of core Brassicaceae, strongly suggesting functional conservation. In contrast, *RCO* acquired a novel expression domain at the base of the leaf, via the evolution of a novel enhancer element in its *cis*-regulatory region [54•].

Although both RCO and LMI1 proteins are potent growth repressors, RCO has been modified during evolution to be less stable and thus biochemically attenuated [50,54•]. Such dual changes in both expression and protein properties may reflect functional tradeoff between the specificity of gene activation and the level of transcriptional activity [55]. In addition to sculpting the compound leaves of *C. hirsuta*, *RCO* contributes to the complex leaf shape of *C. rubella* [56], highlighting the potential of homologous genetic modules to contribute to the repeatability of morphological evolution. The wealth of genetic and genomic resources available in the crucifers can also help understand whether equivalent genetic processes underlie trait variation at different evolutionary scales, for example within and between species. Interestingly, interspecific variation in *C. hirsuta* leaf shapes appears to be caused by variation of the age-dependent progression of leaf form as opposed to variation in growth and patterning which underlies differences between *A. thaliana* and *C. hirsuta* leaf form [48,50,57•]. This difference may reflect tinkering with pleiotropic links between leaf and flowering time pathways to support optimal leaf physiology and adaptive variation in flowering time, which is likely to be prevalent in winter annuals like *C. hirsuta*.

Fruit diversity

Another trait that varies significantly in Brassicaceae is fruit morphology. The genetic tractability of mustard species with divergent fruit morphologies has permitted the development and explicit testing of computational models, which have provided a framework to describe fruit development and diversity. Three phases in regional anisotropic growth generated the two broad fruit categories in the family: an elongated silique, as in *A. thaliana*, and a shorter silicle, as in *Capsella* and *Lepidium* species [58•]. Clonal and morphological analyses in *Capsella* and *Arabidopsis* were employed to build a tissue-level model with variable growth rates in space and time [58•]. The model identified tissue-specific and temporally specific parameters, which when varied could explain much of the fruit shape diversity in Brassicaceae. The valve identity gene *FRUITFULL* was implicated in one such activity along a distal direction of a linear gradient in late phase fruits [58•].

Fruits exhibit different seed dispersal strategies depending on changes in the dehiscence zone at the valve margin, giving rise to dehiscent and indehiscent fruits [59]. A particularly striking dispersal strategy evolved uniquely in the genus *Cardamine*. Many *Cardamine* species are widespread pioneer species that use explosive pod shatter, where seeds are ballistically shot away from the fruit at a distance of up to two meters, to successfully colonize ruderal and disturbed habitats [60•]. A combination of genetics and computational modeling revealed that a slow build-up of tension in the fruit depends on turgor pressure, cellular geometry and anisotropy, while the rapid release of this tension depends on unique, asymmetric cell wall thickenings [60•]. The evolutionary novelty of asymmetric lignin deposition in *Cardamine* fruits profoundly altered their mechanical properties and dispersal strategy. Interestingly, the cell walls of a seed coat layer in *C. hirsuta* are also asymmetrically thickened when compared to the seed coat of *A. thaliana* [19•]. The difference in cell wall properties was attributed to a higher degree of pectin methyl-esterification in *C. hirsuta* compared to *A. thaliana* seeds, which correlated with expansion and differential expression of several pectin methylesterase genes and pectin methylesterase inhibitor genes. This may reflect a functional link between explosive fruit shattering for distant propagule dispersal and aerodynamic seed morphology adapted to reduce drag.

Concluding remarks

Although much of the morphological variation in Brassicaceae remains underappreciated, the diversity in the family is a fertile ground to study the genetic basis of phenotypic change, especially featuring species that manifest truly novel biology compared to *A. thaliana*. There are numerous examples in addition to the discussed explosive pod shatter in *Cardamine*. *Chamira circaeoides* has persistent cotyledons that function as the main photosynthetic organs. Climbing habit with its adaptations evolved in the genera *Lepidium*, *Heliophila*, and *Cremolobus*. The branches are modified into thorns in some dry-adapted species (*Vella* spp., *Zilla* spp., *Hormatophylla spinosa*). Trichome morphology, inflorescence shape, and nectary arrangement are very diverse in Brassicaceae [2,61]. Glandular multicellular trichomes evolved in the divergent Lineage III (*e.g.*, *Anchonium*, *Bunias*, *Chorispora*, *Hesperis*). There are exceptions to the largely stereotypical floral bauplan in the family (*e.g.*, sepal fusion in *Sisymbrium*, petal reduction in *Lepidium* and *Subularia*, stamen reduction in *Lepidium*, and stamen multiplication in *Megacarpaea*) [62]. Floral monosymmetry with two petals of unequal size evolved several times independently in the Brassicaceae [63]. There is also a hidden floral diversity in the family because several distinct pollinator guilds could be identified as faithful visitors of different Brassicaceae species [64]. At least part of this hidden diversity may be attributed to differences in the floral volatile profiles among species, an area of research that receives increasing attention [65,66]. Geocarpy, where fruits bury into the soil after fertilization, evolved several times independently in the family (*e.g.*, *Cardamine chenopodiifolia*, *Geococcus pusillus*, *Lignarella* spp.). Some *Aethionema* species exhibit phenotypic plasticity under different environmental conditions by producing two distinct fruit and seed morphs [67]. Genomic information throughout the family, combined with functional studies and field experiments to explore trait significance and species interactions in the

wild is a promising way forward to understand how such traits evolve and diversify.

Acknowledgements

We are grateful to Angela Hay for critically reading the manuscript. We thank Sara Manafzadeh and Elena Conti for sharing geological maps, and Adam Runions for comments of the figures. Work on crucifer development and diversity in the Tsiantis Laboratory is supported by Deutsche Forschungsgemeinschaft "Adaptomics" grants TS 229/1-1 and SFB (Sonderforschungbereit) 680, and core grant by the Max Planck Society. LN is supported by a Humboldt-Bayer Research Fellowship.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest

1. Provart NJ, Alonso J, Assmann SM, Bergmann D, Brady SM, Brkljacic J, Browse J, Chapple C, Colot V, Cutler S *et al.*: **50 years of *Arabidopsis* research: highlights and future directions**. *New Phytol* 2016, **209**:921-944.

2. Appel O, Al-Shehbaz IA: **Cruciferae**. In *Families and Genera of Flowering Plants V*. Edited by Kubitzki K. Springer; 2003:75-174.

3. Al-Shehbaz IA: **A generic and tribal synopsis of the Brassicaceae (Cruciferae)**. *Taxon* 2012, **61**:931-954.

4. Huang CH, Sun R, Hu Y, Zeng L, Zhang N, Cai L, Zhang Q, Koch MA, Al-Shehbaz I, Edger PP *et al.*: **Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution**. *Mol Biol Evol* 2016, **33**:394-412.

5. The Arabidopsis Genome Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana***. *Nature* 2000, **408**:796-815.

6. • The 1001 Genomes Consortium: **1135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana***. *Cell* 2016, **166**:1-11.
The latest report on global genomic variation in *A. thaliana* throughout its geographic range.

7. Chae E, Bomblies K, Kim S-T, Karelina D, Zaidem M, Ossowski S, Martin-Pizarro C, Laitinen RAE, Rowan BA, Tenenboim H *et al.*: **Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis**. *Cell* 2014, **159**:1341-1351.

8. • Sicard A, Kappel C, Josephs EB, Lee YA, Marona C, Stinchcombe JR, Wright SI, Lenhard M: **Divergent sorting of a balanced ancestral polymorphism underlies the establishment of gene-flow barriers in *Capsella***. *Nat Commun* 2015, **6**:7960.
This study identifies two loci that engage in epistatic interactions resulting in gene-flow barriers in three shepherd's-purse species, and their differential lineage sorting.

9. • Novikova PY, Hohmann N, Nizhynska V, Tsuchimatsu T, Ali J, Muir G, Guggisberg A, Pappe T, Schmid K, Fedorenko OM *et al.*: **Sequencing the genus *Arabidopsis* identified a complex history of nonbifurcating speciation and abundant trans-specific polymorphism**. *Nat Genet* 2016, **48**:1077-1082.
Exhaustive study of the distribution of coding sequence variation in all currently recognized taxa in the genus *Arabidopsis*, which revealed extensive gene flow and reticulate evolution.

10. Bomblies K, Madlung A: **Polyploidy in the *Arabidopsis* genus**. *Chromosom Res* 2014, **22**:117-134.

11. Zapata L, Ding J, Willing E-M, Hartwig B, Bezdan D, Jiao W-B, Patel V, James GV, Koornneef M, Ossowski S *et al.*: **Chromosome-level assembly of *Arabidopsis thaliana* Ler reveals the extent of translocation and inversion polymorphism**. *Proc Natl Acad Sci U S A* 2016, **113**:E4052-E4060.

12. Donoghue MTA, Keshavaiah C, Swamidatta SH, Spillane C: **Evolutionary origins of Brassicaceae specific genes in *Arabidopsis thaliana***. *BMC Evol Biol* 2011, **11**:47.

13. Lysak MA, Mandáková T, Schranz ME: **Comparative paleogenomics of crucifers: ancestral genomic blocks revisited**. *Curr Opin Plant Biol* 2016, **30**:108-115.

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

15. Dassanayake M, Oh D-H, Hass JS, Hernandez A, Hong H, Ali S, Yun D-J, Bressan RA, Zhu J-K, Bohnert HJ *et al.*: **The genome of the extremophile crucifer *Thellungiella parvula***. *Nat Genet* 2011, **43**:913-918.

16. Wu HJ, Zhang Z, Wang JY, Oh DH, Dassanayake M, Liu B, Huang Q, Sun HX, Xia R, Wu Y *et al.*: **Insights into salt tolerance from the genome of *Thellungiella salsuginea***. *Proc Natl Acad Sci U S A* 2012, **109**:12219-12224.

17. Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-Lopez Z, Steffen JG, Hazzouri KM *et al.*: **An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions**. *Nat Genet* 2013, **45**:891-898.

18. Pease JB, Haak DC, Hahn MW, Moyle LC: **Phylogenomics reveals three sources of adaptive variation during a rapid radiation**. *PLoS Biol* 2016, **14**:e1002379.

19. • Gan X, Hay A, Kwantes M, Haberer G, Hallab A, Dello Ioio R, Hofhuis H, Pieper B, Cartolano M, Neumann U *et al.*: **The *Cardamine hirsuta* genome highlights the pervasive role of transcription factors and tandem gene duplications in generating morphological diversity**. *Nat Plants* 2016, **2**:16167.
The study reports the genome assembly and annotation of the hairy bittercress, a model species employed extensively to study the genetic basis of morphological change, including leaf shape diversity and explosive pod shatter.

20. Hanikenne M, Kroymann J, Trampczynska A, Bernal M, Motte P, Clemens S, Krämer U: **Hard selective sweep and ectopic gene conversion in a gene cluster affording environmental adaptation**. *PLoS Genet* 2013, **9**:e1003707.

21. Geiser C, Mandáková T, Arrigo N, Lysak MA, Parisod C: **Repeated whole-genome duplication, karyotype reshuffling and biased retention of stress-responding genes in Buckler Mustard**. *Plant Cell* 2016, **28**:17-27.

22. Simon S, Rühl M, de Montaigu A, Wötzel S, Coupland G: **Evolution of *CONSTANS* regulation and function after gene duplication produced a photoperiodic flowering switch in the Brassicaceae**. *Mol Biol Evol* 2015, **32**:2284-2301.

23. Hofberger J, Lyons E, Edger PP, Pires JC, Schranz ME: **Whole genome and tandem duplicate retention facilitated glucosinolate pathway diversification in the mustard family**. *Genome Biol Evol* 2013, **5**:2155-2173.

24. Mandáková T, Singh V, Krämer U, Lysak MA: **Genome structure of the heavy metal hyperaccumulator *Noccaea caerulescens* and its stability on metalliferous and nonmetalliferous soils**. *Plant Physiol* 2015, **169**:674-689.

25. Franzke A, Koch MA, Mummenhoff K: **Turnip time travels: age estimates in Brassicaceae**. *Trends Plant Sci* 2016, **21**:554-561.

26. Franzke A, Lysak MA, Al-Shehbaz IA, Koch MA, Mummenhoff K: **Cabbage family affairs: the evolutionary history of Brassicaceae**. *Trends Plant Sci* 2011, **16**:108-116.

27. Hohmann N, Wolf EM, Lysak MA, Koch MA: **A time-calibrated road map of Brassicaceae species radiation and evolutionary history**. *Plant Cell* 2015, **27**:2770-2784.

28. Tank DC, Eastman JM, Pennell MW, Soltis PS, Soltis DE, Hinchliff CE, Brown JW, Sessa EB, Harmon LJ: **Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications**. *New Phytol* 2015, **207**:454-467.

29. Kellogg EA: **Has the connection between polyploidy and diversification actually been tested?** *Curr Opin Plant Biol* 2016, **30**:25-32.

30. Kagale S, Robinson SJ, Nixon J, Xiao R, Huebert T, Condie J, Kessler D, Clarke WE, Edger PP, Links MG *et al.*: **Polyploid evolution of the Brassicaceae during the Cenozoic era**. *Plant Cell* 2014, **26**:2777-2791.

31. Cheng F, Mandáková T, Wu J, Xie Q, Lysak MA, Wang X: **Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa***. *Plant Cell* 2013, **25**:1541-1554.

32. Mandáková T, Mummenhoff K, Al-Shehbaz IA, Mucina L, Mühlhausen A, Lysak MA: **Whole-genome triplication and species radiation in the southern African tribe Heliophileae (Brassicaceae)**. *Taxon* 2012, **61**:989-1000.

33. Mandáková T, Joly S, Krzywinski M, Mummenhoff K, Lysak MA: **Fast diploidization in close mesopolyploid relatives of *Arabidopsis***. *Plant Cell* 2010, **22**:2277-2290.

34. Edger PP, Heidel-Fischer HM, Mekaert M, Rota J, Glöckner G, Platts AE, Heckel DG, Der JP, Wafula EK, Tang M *et al.*: **The butterfly plant arms-race escalated by gene and genome duplications**. *Proc Natl Acad Sci U S A* 2015, **112**:8362-8366.

35. Kagale S, Koh C, Nixon J, Bollina V, Clarke WE, Tuteja R, Spillane C, Robinson SJ, Links MG, Clarke C *et al.*: **The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure**. *Nat Commun* 2014, **23**:3706.

36. Cheng F, Wu J, Fang L, Sun S, Liu B, Lin K, Bonnema G, Wang X: **Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa***. *PLoS One* 2012, **7**: e36442.

37. Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F *et al.*: **Genome sequencing project consortium: the genome of the mesopolyploid crop species *Brassica rapa***. *Nat Genet* 2011, **43**:1035-1039.

38. Woodhouse MR, Cheng F, Pires JC, Lisch D, Freeling M, Wang X: **Origin, inheritance, and gene regulatory consequences of gene dominance in polyploids**. *Proc Natl Acad Sci U S A* 2014, **111**:5283-5288.

39. Hollister JD, Smith LM, Guo YL, Ott F, Weigel D, Gaut BS: **Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata***. *Proc Natl Acad Sci U S A* 2011, **108**:2322-2327.

40. • Kawakatsu T, Huang SC, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y *et al.*: **Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions**. *Cell* 2016, **166**:492-505.

Detailed study of the global epigenomic diversity and expression of *A. thaliana* accessions throughout its geographic range.

41. Hatorangan MR, Laenen B, Steige KA, Slotte T, Köhler C: **Rapid evolution of genomic imprinting in two species of the Brassicaceae**. *Plant Cell* 2016, **28**:1815-1827.

42. Smith LM, Burbano HA, Wang X, Fitz J, Wang G, Ural-Blimke Y, Weigel D: **Rapid divergence and high diversity of miRNAs and miRNA targets in the Camelineae**. *Plant J* 2015, **81**:597-610.

43. Seymour DK, Koenig D, Hagmann J, Becker C, Weigel D: **Evolution of DNA methylation patterns in the Brassicaceae is driven by differences in genome organization**. *PLoS Genet* 2014, **10**:e1004785.

44. Steige KA, Laenen B, Reimegård J, Scofield DG, Slotte T: **Genomic analysis reveals major determinants of *cis*-regulatory variation in *Capsella grandiflora***. *Proc Natl Acad Sci U S A* 2017, **114**:1087-1092.

45. • Willing E-M, Rawat V, Mandáková T, Maumus F, Velikkakam James G, Nordström KJV, Becker C, Warthmann N, Chica C *et al.*: **Genome expansion of *Arabis alpina* linked with retrotransposon and reduced symmetric DNA methylation**. *Nat Plants* 2015, **1**:e14023.

The study reports the genome assembly and annotation of the most advanced perennial mustard model, the Alpine rock-cress, which also exhibits unique DNA methylation pattern.

46. Lewsey MG, Hardcastle TJ, Melnyk CW, Molnar A, Valli A, Urich MA, Nery JR, Baulcombe DC, Ecker JR: **Mobile small RNAs regulate genome-wide DNA methylation**. *Proc Natl Acad Sci U S A* 2016:E801-E810.

47. Hay AS, Pieper B, Cooke E, Mandáková T, Cartolano M, Tattersall AD, Ioio RD, McGowan SJ, Barkoulas M, Galinha C *et al.*: ***Cardamine hirsuta*: a versatile genetic system for comparative studies**. *Plant J* 2014, **78**:1-15.

48. Bilsborough GD, Renions A, Barkoulas M, Jenkins HW, Hasson A, Galinha C, Laufs P, Hay A, Prusinkiewicz P, Tsiantis M: **Model for the regulation of *Arabidopsis thaliana* leaf margin development**. *Proc Natl Acad Sci U S A* 2011, **108**:3424-3429.

49. Bar M, Ori N: **Leaf development and morphogenesis**. *Development* 2014, **141**:4219-4230.

50. Vlad D, Kierzkowski D, Rast MI, Vuolo F, Dello Ioio R, Galinha C, Gan X, Hajheidari M, Hay A, Smith RS *et al.*: **Leaf shape evolution through duplication, regulatory diversification, and loss of a homeobox gene**. *Science* 2014, **343**:780-783.

51. Rast-Somssich MI, Broholm S, Jenkins H, Canales C, Vlad D, Kwantes M, Bilsborough G, Dello Ioio R, Ewing RM, Laufs P *et al.*: **Alternate wiring of a KNOXI genetic network underlies differences in leaf development of *A. thaliana* and *C. hirsuta***. *Genes Dev* 2015, **29**:2391-2404.

52. Bar M, Ori N: **Compound leaf development in model plant species**. *Curr Opin Plant Biol* 2015, **23**:61-69.

53. Sousa-Baena MS, Lohmann LG, Rossi M, Sinha NR: **Acquisition and diversification of tendrilled leaves in Bignonieae (Bignoniaceae) involved changes in expression patterns of *SHOOTMERISTEMLESS* (*STM*), *LEAFY*/*FLORICAULA* (*LFY*/*FLO*), and *PHANTASTICA* (*PHAN*)**. *New Phytol* 2014, **201**:993-1008.

54. • Vuolo F, Mentink RA, Hajheidari M, Bailey CD, Filatov DA, Tsiantis M: **Coupled enhancer and coding sequence evolution of a homeobox gene shaped leaf diversity**. *Genes Dev* 2016, **30**:1-6.

The study outlines the interplay between regulatory and protein divergence in the evolution of *RCO*, a major regulator of leaf shape in mustards, and provides for the first time evidence linking causal genetic variation for leaf shape between species to plant performance.

55. Farley EK, Olson KM, Zhang W, Brandt AJ, Rokhsar DS, Levine MS: **Suboptimization of developmental enhancers**. *Science* 2015, **350**:325-328.

56. Sicard A, Thamm A, Marona C, Lee YW, Wahl V, Stinchcombe JR, Wright SI, Kappel C, Lenhard M: **Repeated evolutionary changes of leaf morphology caused by mutations to a homeobox gene**. *Curr Biol* 2014, **24**:1880-1886.

57. • Cartolano M, Pieper B, Lempe J, Tattersall A, Huijser P, Tresch A, Darrah PR, Hay A, Tsiantis M: **Heterochrony underpins natural variation in *Cardamine hirsuta* leaf form**. *Proc Natl Acad Sci U S A* 2015, **112**:10539-10544.

Reports that regulatory variation at the MADS box gene *Flowering Locus C* underlies a major effect quantitative trait locus affecting leaflet number in *C. hirsuta*, and highlights the role of heterochronic shifts as drivers of morphological change.

58. • Eldridge T, angowski L, Stacey N, Jantzen F, Moubayidin L, Sicard A, Southam P, Kennaway R, Lenhard M, Coen ES *et al.*: **Fruit shape diversity in the Brassicaceae is generated by varying patterns of anisotropy**. *Development* 2016, **143**:3394-3406.

Detailed clonal and modeling work is used to describe the heart-shaped fruits of shepherd's-purse and extends these findings to the broader fruit shape diversity in mustards.

59. Muhlhausen A, Lenser T, Mummenhoff K, Theissen G: **Evidence that an evolutionary transition from dehiscent to indehiscent fruits in *Lepidium* (Brassicaceae) was caused by a change in the control of valve margin identity genes**. *Plant J* 2013, **73**:824-835.

60. • Hofhuis H, Moulton D, Lessinnes T, Routier-Kierzkowska AL, Bomphrey RJ, Mosca G, Reinhardt H, Sarchet P, Gan X, Tsiantis M *et al.*: **Morphomechanical innovation drives explosive seed dispersal**. *Cell* 2016, **166**:1-12.

A combination of modeling and genetics reveals the mechanism of explosive pod shatter in *Cardamine*.

61. Erbar C: **Nectar secretion and nectaries in basal angiosperms, magnoliids and non-core eudicots and a comparison with core eudicots**. *Plant Divers Evol* 2014, **131**:63-143.

62. Endress PK: **Evolution and floral diversity: the phylogenetic surroundings of *Arabidopsis* and *Antirrhinum***. *Int J Plant Sci* 1992, **153**:S106-S122.

63. Busch A, Horn S, Mühlhausen A, Mummenhoff K, Zachgo S: **Corolla monosymmetry: evolution of a morphological novelty in the Brassicaceae family**. *Mol Biol Evol* 2012, **29**:1241-1254.

64. Gómez JM, Torices R, Lorite J, Klingenberg CP, Pefectti F: **The role of pollinators in the evolution of corolla shape variation, disparity and integration in a highly diversified plant family with a conserved floral bauplan**. *Ann Bot* 2016, **117**:889-904.

65. Muhlemann J, Kleimpien A, Dudareva N: **Floral volatiles: from biosynthesis to function**. *Plant Cell Environ* 2014, **37**:1936-1949.

66. Sas C, Müller F, Kappel C, Kent TV, Wright SI, Hilker M, Lenhard M: **Repeated inactivation of the first committed enzyme underlies the loss of benzaldehyde emission after the selfing transition in *Capsella***. *Curr Biol* 2016, **26**:3313-3319.

67. Lenser T, Graeber K, Cevik ÖS, Aligüzel N, Dönmez AA, Grosche C, Kettermann M, Mayland-Quellhorst S, Mérai Z, Mohammadin S *et al.*: **Developmental control and plasticity of fruit and seed dimorphism in *Aethionema arabicum***. *Plant Physiol* 2016, **172**:1691-1707.

68. Cheng F, Sun R, Hou X, Zheng H, Zhang F, Zhang Y, Liu B, Liang J, Zhuang M, Liu Y *et al.*: **Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea***. *Nat Genet* 2016, **48**:1218-1224.

69. Jensen JD, Foll M, Bernatchez L: **Introduction: the past, present, and future of genomic scans for selection**. *Mol Ecol* 2016, **25**:1-4.

70. Huber CD, Nordbord M, Hermission J, Hellmann I: **Keeping it local: evidence for positive selection in Swedish *Arabidopsis thaliana***. *Mol Biol Evol* 2014, **31**:3026-3039.

71. Manafzadeh S, Salvo G, Conti E: **A tale of migrations from east to west: the Irano-Turanian floristic region as a source of Mediterranean xerophytes**. *J Biogeogr* 2013, **41**:366-379.

72. Rögl F: **Mediterranean and Paratethys. Facts and hypotheses of an Oligocene to Miocene paleogeography (short overview)**. *Geol Carpathica* 1999, **50**:339-349.