Complex Evolutionary Events at a Tandem Cluster of Arabidopsis thaliana Genes Resulting in a Single-Locus Genetic Incompatibility

The Harvard community has made this article openly available. **Please share** how this access benefits you. Your story matters

| Citation       | Smith, Lisa M., Kirsten Bomblies, and Detlef Weigel. 2011. Complex evolutionary events at a tandem cluster of Arabidopsis thaliana genes resulting in a single-locus genetic incompatibility. PLoS Genetics 7(7): e1002164. |
|----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Published Version | doi:10.1371/journal.pgen.1002164                                                                                                                                                                    |
| Citable link    | http://nrs.harvard.edu/urn-3:HUL.InstRepos:9976281                                                                                                                                                   |
| Terms of Use    | This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP |
Complex Evolutionary Events at a Tandem Cluster of Arabidopsis thaliana Genes Resulting in a Single-Locus Genetic Incompatibility

Lisa M. Smith, Kirsten Bomblies*, Detlef Weigel*
Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany

Abstract

Non-additive interactions between genomes have important implications, not only for practical applications such as breeding, but also for understanding evolution. In extreme cases, genes from different genomic backgrounds may be incompatible and compromise normal development or physiology. Of particular interest are non-additive interactions of alleles at the same locus. For example, overdominant behavior of alleles, with respect to plant fitness, has been proposed as an important component of hybrid vigor, while underdominance may lead to reproductive isolation. Despite their importance, only a few cases of genetic over- or underdominance affecting plant growth or fitness are understood at the level of individual genes. Moreover, the relationship between biochemical and fitness effects may be complex: genetic overdominance, that is, increased or novel activity of a gene may lead to evolutionary underdominance expressed as hybrid weakness. Here, we describe a non-additive interaction between alleles at the Arabidopsis thaliana OAK (OUTGROWTH-ASSOCIATED PROTEIN KINASE) gene. OAK alleles from two different accessions interact in F1 hybrids to cause a variety of aberrant growth phenotypes that depend on a recently acquired promoter with a novel expression pattern. The OAK gene, which is located in a highly variable tandem array encoding closely related receptor-like kinases, is found in one third of A. thaliana accessions, but not in the reference accession Col-0. Besides recruitment of exons from nearby genes as promoter sequences, key events in OAK evolution include gene duplication and divergence of a potential ligand-binding domain. OAK kinase activity is required for the aberrant phenotypes, indicating it is not recognition of an aberrant protein, but rather a true gain of function, or overdominance for gene activity, that leads to this underdominance for fitness. Our work provides insights into how tandem arrays, which are particularly prone to frequent, complex rearrangements, can produce genetic novelty.

Introduction

Both evolutionary biologists and breeders have long been interested in non-additive interactions among alleles at the same locus. For example, explanations for heterosis or hybrid vigor, a staple of modern agriculture, share many conceptual formalities with models proposed by Bateson, Dobzhansky and Muller to explain how negative heterosis could result from two or more genes that accumulate different changes in separate lineages. The associated phenotypes of hybrid weakness, sterility or lethality in turn may ultimately lead to reproductive isolation and hence speciation [1–3], reviewed in [4,5]. Hybrid incompatibilities form a continuum from the grey zone of developmental abnormalities through the clearer phenotype of F1 sterility to the severest form, lethality, and it is important to understand the genetic and molecular causes for the entire spectrum of incompatibilities.

F1 incompatibilities have been found in as many as 2% of Arabidopsis thaliana intra-specific hybrids [6]. Several similar cases in A. thaliana and other species involve interactions between alleles of disease resistance genes with other loci in the genome, which cause an autoimmune syndrome known as hybrid necrosis [6–8]. That hybrid necrosis is such a relatively common phenomenon is easily explained, since genes involved in plant defense are highly variable between different individuals of the same species [9,10], and thus make a perfect substrate for causing problems when different genomes are combined. Moreover, several important classes of defense genes, including those encoding nucleotide binding-leucine rich repeat (NB-LRR) proteins and receptor-like kinases (RLKs), commonly occur in tandem arrays, and new alleles are easily created through gene duplication, illegitimate recombination and gene conversion [11–19].

In addition to inappropriate activation of the immune system or sterility, aberrant development is often observed in incompatible plant hybrids [20,21]. Both Triticum and Nicotiana interspecific hybrids frequently suffer from tumor-like tissue proliferation [22,23]. In Nicotiana hybrids, wounding and physiological stresses enhance tumor formation, and tumors may differentiate into recognizable tissues [24]. Genetically-induced tumors have also
Heterozygous Disadvantage in *A. thaliana*

Here, we report on an intraspecific *A. thaliana* F1 hybrid, where heterozygosity at a single locus causes a pleiotropic syndrome that includes smaller stature and reduced seed set as well as ectopic outgrowths on leaf petioles. The causal receptor-like kinase (RLK) gene, *OUTGROWTH-ASSOCIATED PROTEIN KINASE (OAK)*, is found in a structurally hypervariable tandem cluster of related RLK genes. During duplication of the ancestral RLK gene, coding sequences were recruited to form a promoter with a new expression domain. Divergence in the extracellular domain of the gene led to evolution of alleles that now interact in the Bla-1/Shahybrid to produce phenotypes not seen in the parents, making this a case of underdominance for fitness caused by overdominance for gene expression.

### Results

**Ectopic petiole outgrowths and reduced biomass of Bla-1/Shahybrids**

The aberrant phenotype of Blanes-1 (Bla-1)/Shahdara (Sha) F1 hybrids was identified in a survey of more than 1,300 crosses among over 300 *A. thaliana* accessions from the world-wide range of the species [6]. Bla-1/Sha F1 plants had a range of phenotypes that were not normally seen in inbred accessions, including the Bla-1 and Sha parents, or in other F1 hybrids: outgrowths on the adaxial surface of the petioles, leaf twisting, leaf lesions, and loss of apical dominance reflected by precocious and increased release of side shoots (Figure 1a–1c). These phenotypes were observed regardless of the direction of the cross. Raising plants in long days at 23°C instead of 16°C restored apical dominance and largely suppressed leaf twisting and lesioning. This partial suppression of the hybrid phenotype at higher temperatures is similar to the suppression of necrosis seen in the Uk-1/Uk-3 and other hybrids with autoimmune defects [6].

Because the ectopic outgrowth phenotype was particularly striking and reliably observed in all F1 plants, we decided to investigate it in detail. The same phenotype with little variation was seen in approximately 50% of all F2 progeny, compatible with a single-gene, heterozygous genetic basis. The outgrowth phenotype segregated independently of the lesioning in the F2 and subsequent generations.

Outgrowths were occasionally noted in the Bla-1 parent, but with incomplete penetrance that varied greatly between experiments (Table S1). Onset of outgrowth formation in Bla-1, when it occurred, was much later than in the F1 hybrids. Crosses of each parental line to the reference accession Col-0 did not produce any progeny with outgrowths, but they were, as expected, seen in about one quarter of progeny after Col-0/Bla-1 and Sha/Col-0 F1 hybrids were crossed to each other.

Analysis of transverse sections revealed that outgrowths originated from proliferating parenchyma and/or epidermal cells on the adaxial surface of the petiole (Figure 1d–1f). The vascular system of the petioles appeared normal. Because of their determinate nature, we concluded that the outgrowths did not constitute undifferentiated callus.

We also asked whether the gene(s) causing the hybrid phenotypes of outgrowth and lesioning might affect overall plant performance. In a segregating F2 population of five-week old plants, we found that outgrowths alone were correlated with a 29% reduction in rosette weight, while lesioning or lesioning plus outgrowths reduced growth by over 50% (Table S2; 2-way ANOVA outgrowths $p = 0.0003$, lesioning $p<0.0001$). In addition, we assessed seed set as a proxy for lifetime fitness. Due to confounding factors such as differential flowering times in Sha and Bla-1, we measured seed set after the incompatibility was...
reconstituted in the Col-0 reference background [see below for further details]. Seed set was reduced by 90% in F1 hybrids that were phenotypically comparable to the natural hybrids (two-tailed, unequal variance t-test: \( p < 0.001 \); Figure S1). In two other independent crosses that resulted in a more severe incompatibility phenotype, all the hybrids died within two months, and thus did not produce any seeds at all. This indicates that the Bla-1/Sha phenotype, all the hybrids died within two months, and thus did not produce any seeds at all. This indicates that the Bla-1/Sha OAK incompatibility greatly reduces lifetime fitness.

Because wounding and physiological stresses enhance the formation of tumors in Nicotiana, where these may differentiate into recognizable tissues [24], we examined the effects of wounding, by prickng the petioles of Bla-1/Sha F1 plants with a fine needle. Outgrowth formation was not enhanced, but we found that increased humidity suppressed outgrowth formation (Figure S2). This is reminiscent of the suppression of constitutive activation of disease resistance in the \( ssi4 \) mutant by high humidity [34].

Compared to normal tissue, induction of callus from Nicotiana hybrid tumors requires less auxin [35]. Some \( A. \) thaliana tumor forming lines also produce callus tissue that can continue to proliferate on hormone-free media [36]. To test auxin response in our system, transverse sections of leaf and petiole tissue were induced to form callus. Although the Bla-1 parent had a relatively higher auxin requirement for callus formation, there was no difference between the Sha parent and the Bla-1/Sha hybrids (Figure S3). Thus, the outgrowths are probably genetically distinct from the \( A. \) thaliana tumors.

Genome-wide expression studies

Microarray analysis with triplicate Affymetrix ATH1 arrays using RNA extracted from three-week-old aerial tissue identified 356 genes differentially expressed in the hybrids compared to the parents. There was no significant up- or down-regulation of any particular known pathways or reactions based on the SkyPainter tool [37], but several, often overlapping, Gene Ontology (GO) categories were enriched among the differentially expressed genes, most notably several related to pathogen response (Table S3; [38]). Whether this reflects a link to disease resistance remains unclear, since some well-known markers for pathogen response, such as \( PRI \) or the defensin gene \( PDF1.2(b) \), were down-regulated in the hybrids (Tables S4 and S5). In any case, as with the morphological phenotype, there was no overwhelming connection to the hybrid necrosis syndrome as seen in many other incompatible \( A. \) thaliana F1 hybrids [21].

Ectopic outgrowths caused by a hypervariable protein kinase gene cluster

Using \( F_2 \) and \( F_3 \) progeny, we mapped the outgrowth phenotype to a single genomic region on chromosome 5 containing 17 genes in the reference accession Col-0 (\( At5g59560 \) to \( At5g59700 \); Figure S4). A tandem array of four genes that encode a distinct clade of closely related receptor-like kinases (RLKs; \( At5g59650 \) to \( At5g59680 \)) [17] were of particular interest, because RLKs are one of the most variable gene families in the \( A. \) thaliana genome [9].

We recovered the genomic regions from \( At5g59616 \) (encoding a protein kinase-related protein) to \( At5g59690 \) (histone H4) by long-range PCR from Bla-1 and Sha, and found the RLK cluster to be highly variable (Figure 2a). In Col-0 only, there are two transposons and a pseudogene upstream of the RLK genes. In Sha, the first RLK gene in the cluster, \( At5g59650 \), is missing and the upstream gene \( At5g59616 \) is only partially present. In both Bla-1 and Sha, a 150 bp remnant of the second RLK gene, \( At5g59660 \), indicates that a deletion likely occurred in the Bla-1/Sha lineage. Also in both Bla-1 and Sha, the third RLK gene of the cluster, \( At5g59670 \), has been duplicated to give rise to \( At5g59670a \) and \( At5g59670b \) (Table S6). In addition to Bla-1 and Sha, the \( At5g59670 \) duplication was detected by PCR analysis of the \( OAK \) promoter in 36 of 87 diverse \( A. \) thaliana accessions (Table S7), while a Col-0 like promoter was found in 45 accessions. Assays for both promoter types were positive in two accessions, indicating either illegitimate recombination or a different duplication event. The PCR assays failed in the remaining four accessions.

Reconstruction of the ancestral state of the tandem array, by comparison with the close relative \( A. \) lyrata [39], suggested the presence of three tandem RLK genes in the last common ancestor of \( A. \) thaliana and \( A. \) lyrata. The central gene was duplicated in the \( A. \) thaliana lineage to produce \( At5g59660 \) and \( At5g59670 \), whereas in \( A. \) lyrata, there have been subsequent duplications of the two flanking RLK genes, resulting in a cluster with six genes. Given the
Two alleles of a single RLK cause novel growth phenotypes

To determine whether any of the RLK genes contribute to the outgrowth phenotype, a genomic copy of each gene from Bla-1 and Sha was individually introduced into the Bla-1, Sha, and Col-0 backgrounds. Only plants transformed with At5g59670b from Bla-1 or Sha developed outgrowths (Figure 3a). Unexpectedly, while At5g59670b from Bla-1 induced outgrowths most effectively in Sha, and At5g59670b from Sha in Bla-1, outgrowths were also seen, albeit at lower frequency, upon transformation of either gene into the recurrent parent or into Col-0. This suggests a dosage effect, perhaps due to elimination of negative regulatory elements or epigenetic marks in the transgene that normally suppress the effect, perhaps due to elimination of negative regulatory elements or epigenetic marks in the transgene that normally suppress the hybrid phenotype (outgrowths, leaf twisting and apical dominance; Figure 2b and Figure S5). We therefore refer to At5g59670b as OUTGROWTH-ASSOCIATED PROTEIN KINASE (OAK).

Comparison of Bla-1 and Sha OAK alleles

The Bla-1 and Sha OAK primary transcripts are each 3.9 kb long, with 13 exons, and a 5' untranslated region of 92 nt (expressed in Bla-1 and Sha petioles) or up to 125 nt (expressed in Sha pedicels and peduncles), as determined by 5' RACE-PCR. Both OAK alleles encode proteins of 873 amino acids, with 9% of residues being different. The majority of polymorphisms are located in a 152 amino acid region, between positions 180 and 331, where 55 residues differ (Figure 3c). Among the remaining 721 residues, there are only 19 replacements.

Like many other plant RLKs, the OAK proteins include a signal peptide, potential leucine-rich repeats (LRRs; in OAK, four to five), a transmembrane domain, and a cytoplasmic kinase domain (Michael Hothorn, personal communication; Figure 3d and Figure S6). In addition, two related regions with similarity to a carbohydrate-binding domain in ER-localized mal lectin proteins from animals [41] are found between the signal peptide and the LRRs (http://toolkit.tuebingen.mpg.de/hhpred/; Michael Hothorn, personal communication). Interestingly, the region that is very different between the Bla-1 and Sha proteins, from residue 180 to 331, coincides almost perfectly with the second predicted mal lectin-like domain, from residue 169 to 331. An analysis of OAK and its homologs (OAKSha, OAKBla-1, At5g59670Bla-1, At5g59670Sha and At5g59670Col-0), using the Codeml program of PAML, to assess dN/dS ratios, did not provide evidence for directional or diversifying selection across the entire protein [42,43]. However, an Bayesian Posterior Probability analysis of positive selection at individual residues, using At5g59670Col-0 as a reference, suggested that several codons in the second mal lectin-like domain are under positive selection [44]. A broader analysis of 34 accessions from which OAK sequences could be recovered supported these conclusions (Figure 3d).

To determine if the second mal lectin-like region in OAK homologs is generally hypervariable, we performed a sliding window analysis of all eleven RLKs in the Col-0, Bla-1 and Sha clusters (Figure S7). Most highly conserved are the LRR and kinase domains. We also examined in detail the duplicated genes encoding the At5g59670 proteins. At5g59670Sha and OAKSha stood out, because they are identical across the first 598 amino acids of the protein. At the nucleotide level, the two genes include an identical 2.7 kb fragment, which most likely reflects a recent gene conversion event that extends from 13 bp upstream of the translational start site to the first 60 bp of the kinase encoding sequences. In conclusion, the divergence between the second mal lectin-like domain of OAKBla-1 and OAKSha is not representative of the variation between RLKs encoded by orthologs and paralogs in this cluster.

Role of divergent promoter sequences in causing the OAK hybrid phenotype

To determine the contribution of non-coding and coding sequences of OAK to the outgrowth phenotype, we performed a series of domain swaps between OAKSha, OAKBla-1, OAKSha, and OAKCol-0 (Figure 4a). Similar to plants transformed with the non-chimeric fragments, T1 transformants frequently showed more severe phenotypes than were observed in the F1 hybrids. This indicated that divergent OAK alleles have the potential to cause even stronger incompatibilities than seen between the accessions Bla-1 and Sha.

The first major conclusion from the experiments with the chimeric transgenes was that the promoter region contributed to the outgrowth phenotype, because outgrowths were only observed when a particular recombinant protein was expressed from either the OAKBla-1 or OAKSha promoter, but never with the
At5g59670Col-0 promoter (Figure 4b). GUS reporter experiments demonstrated that the OAK promoters from Bla-1 and Sha were active in the vascular system of the petioles, in a pattern consistent with the location of the outgrowths (Figure 5). In contrast, the At5g59670Col-0 promoter drove expression in the leaf lamina, explaining why it could not cause petiole outgrowths. The activity domain of the At5g59670aBla-1 promoter was similar to that of the At5g59670Col-0 promoter, but with additional expression in the lamina of the cotyledons. Finally, the At5g59670aSha promoter was active in all seedling tissues, but in isolated patches that differed from plant to plant. Thus, despite the encoded proteins being closely related, the promoters

**Figure 3. Identification of At5g59670b homologs as sufficient and necessary for outgrowths.** (a) Fraction of T1 plants (n≥90, except for Bla-1 transformed with Sha At5g59680 where n = 56) with outgrowths. (b) Suppression of outgrowths with amiRNAs against OAK (At5g59670b) from Bla-1 or Sha. (c) Divergence between OAK (At5g59670b) alleles from Bla-1 and Sha (sliding windows of 60 bp and 20 amino acids, respectively). (d) Identification of individual sites in the N-terminal part of OAK protein under positive selection (as determined by Bayesian Posterior Probability) across 34 accessions using PAML [43]. The second maelectin-like domain is enriched for such sites.

doi:10.1371/journal.pgen.1002164.g003

At5g59670Col-0 promoter (Figure 4b). GUS reporter experiments demonstrated that the OAK promoters from Bla-1 and Sha were active in the vascular system of the petioles, in a pattern consistent with the location of the outgrowths (Figure 5). In contrast, the At5g59670Col-0 promoter drove expression in the leaf lamina, explaining why it could not cause petiole outgrowths. The activity domain of the At5g59670aBla-1 promoter was similar to that of the At5g59670Col-0 promoter, but with additional expression in the lamina of the cotyledons. Finally, the At5g59670aSha promoter was active in all seedling tissues, but in isolated patches that differed from plant to plant. Thus, despite the encoded proteins being closely related, the promoters
conditioned a surprisingly wide spectrum of expression patterns, with differences both between duplicates within an accession and among orthologs from different accessions.

**Diversity and origin of promoters in the OAK cluster**

The OAK<sub>Bla-1</sub> and OAK<sub>Sha</sub> promoters are more similar to each other than are the coding regions, being 97% identical in the 1,238 bp upstream of the start codon. OAK promoter sequences could be recovered from a further 32 accessions. Pairwise identity for all 34 accessions including Bla-1 and Sha was between 97 and 100%. Given the high similarity of the promoter region, the duplication of At5g39670 to form OAK is unlikely to have occurred more than once. Therefore while the change in expression domain has determined how the incompatibility is expressed, the causative changes for the incompatibility are not within the promoter region. In comparison, over the first 1,077 bp of the coding region, the pairwise identity for the 34 accessions ranged from 87 to 100%, with a mean of 94%. One accession that was identical to Sha throughout both the promoter and coding region was Kondara, which we found to be incompatible with Bla-1 as well. Across the entire RLK cluster, there were only two nucleotide differences in 17.5 kb, and both were in non-coding sequences. Kondara was therefore not considered separately in any of the sequence analyses. Further crosses of Bla-1 and Sha to other accessions with the OAK gene revealed that while most accessions are compatible, a similar

---

**Figure 4. Contribution of both the OAK promoter and extracellular domain to outgrowths.** (a) Overview of domain swaps. (b) Phenotypic distribution of T<sub>1</sub> plants (n=90). Three-letter code indicates composition of chimeras. E.g., BBS, promoter and extracellular domain from Bla-1, kinase domain from Sha. Examples of phenotypic classes are shown at the bottom: mild (outgrowths, but otherwise normal leaves), moderate (outgrowths, shortened petioles, mild leaf twisting, normal lamina size) or severe (stunted plants, petioles almost absent, reduced lamina surface, seed rarely obtained). Scale bar = 1 cm. 

doi:10.1371/journal.pgen.1002164.g004
incompatibility phenotype is seen in Sha x Bak-2, Sha x Leo-1, Mer-6 x Bla-1 and Leb-3 x Bla-1 hybrids (all incompatibilities between Bla-1-like and Sha-like haplotype groups based on the second malectin domain; Figure S8). Less severe incompatibilities with a late onset of outgrowth formation were found in crosses of Bla-1 to a number of accessions with a second malectin domain that fell into a different haplotype group (ICE91, ICE92, ICE152, ICE153, Vash-1 and Valsi-1).

Using NeighborNet implemented in SplitsTree [45], we examined the relationship between the RLKs from the 34 accessions based on the promoter sequences and the extracellular domains (amino acids 1 to 360; Figure 6a, 6b). Similarity in the coding region was not always reflected in promoter similarity, and vice versa, suggesting a history of recombination or gene conversion events. The SplitsTree analysis suggested four major haplotypes at the OAK locus. Analysis with STRUCTURE [46], where we treated polymorphisms in the OAK locus as linked markers on a chromosome, confirmed that there are four major haplotype groups, with half of the accessions studied showing contributions from more than one haplotype group (Figure S6c). Within-locus switching between haplotype groups was confirmed by visual inspection of sequence alignments between individual accessions. This likely reflects high levels of gene conversion or recombination within the OAK gene.

A search of the Col-0 reference genome for the possible origin of the OAK promoter revealed that most of it probably arose from the coding region of one of the RLK genes, spanning intron 2 to exon 7 (encoding amino acids 207 to 383 of At5g59670). Although these regions are only 60 to 70% identical to the OAK promoter (BLASTN v2.2.25, E-value 1 × 10^{-61}), they present the best matches in the Col-0 genome (second best hit is to LRR-RLK gene At3g46330, E-value 3 × 10^{-13}) indicating that this is the most likely origin of the OAK promoter. While the promoter includes potential coding sequences, there are several in-frame stop codons upstream of the predicted OAK translation start. The OAK_Bla-1 and OAK_Sha promoters show similar levels of identity with RLK coding sequences across the cluster, but it seems most likely that the duplication of the At5g59670 gene involved an additional duplication that led to conversion of the region coding largely for the second malectin-like domain into a promoter. Interestingly, this is also the portion of the coding sequence that is most different between Bla-1 and Sha. The 260 bp promoter region immediately upstream of the start codon of OAK is most similar to sequences found in triplicate in the At5g59670_Col-0 promoter (Figure S9).
Role of the protein and kinase activity in causing the OAK hybrid phenotype

A second conclusion of the chimeric transgene experiments was that in addition to the promoter, the protein, and the extracellular domain in particular, contributed to the outgrowth phenotype (Figure 4a, 4b). The At5g59670Col-0 protein did not cause an incompatibility phenotype even when expressed under the OAKBla-1 or OAKSha promoters. Swapping the extracellular and cytoplasmic domains between the OAKBla-1 and OAKSha proteins showed that the cytoplasmic domains were broadly equivalent. However, introduction of the extracellular domain of OAKBla-1 into the Sha genotype, or vice versa, greatly increased the proportion of affected T1 plants. This result is supported by the incompatibility between Leo-1 and Sha, where Leo-1 has an extracellular domain identical to Bla-1, but only two amino acid differences in the cytoplasmic domain compared to Sha (Figure S10). Further attempts to narrow down the causal region within the extracellular domain with additional chimeras were not successful.

We tested the hypothesis that the outgrowth phenotype resulted from ectopic activation of a kinase-dependent signaling pathway by mutating key residues in the kinase catalytic domain [47]. Double mutants of D693N and K695R should lack all kinase activity. In the Sha background, over 80% of T1 plants carrying the Bla-1 kinase-active construct had a moderate or severe phenotype, while only one third of T1 plants transformed with the Bla-1 kinase-dead construct had any phenotype, and this was always mild. When the Sha kinase-dead construct was transformed back into the Sha accession, all T1 transformants were wild type in appearance, which contrasts with 30% of T1 plants expressing the Sha kinase-active construct having a mild to severe phenotype (Figure 7a). Results were comparable with Bla-1 transformants, although in this case some plants with a moderate phenotype were observed after transformation with the Sha kinase-dead construct.

Because RLKs can form homo- and heterodimers [48], we tested the effects of combining Bla-1 and Sha kinase-dead versions...
in the neutral Col-0 reference background. We transformed both kinase-active and -dead versions individually into Col-0 and then generated the four possible combinations by crossing (Figure 7b, 7c). The F1 hybrids in which only one of the transgenes expressed a kinase-active version had a less severe phenotype than those carrying both Bla-1 and Sha kinase-active versions. All F1 progeny from five crosses using OAK kinase-dead forms of both Bla-1 and Sha were wild type in appearance. This finding not only confirmed that kinase activity of OAK is required for its function, but also suggested that OAK can act as a heteroallelic dimer or multimer, because a kinase active version of one OAK allele can at least partially complement a kinase-dead version of the other OAK allele. In addition, these data indicated that other RLKs present at the OAK cluster in Col-0 are unlikely to be involved in the outgrowth phenotype.

Further circumstantial evidence suggesting that OAK proteins form dimers or multimers was obtained by expressing only the extracellular domain of OAKBla-1 or OAKSha in hybrid plants. Expression under the native promoter in particular suppressed the outgrowth phenotype in many OAKBla-1/OAKSha heterozygous plants (Figure S11). We propose that by binding to OAK proteins, the extracellular domains reduce the number of active OAKBla-1 or OAKSha heterodimers. The OAK kinase can couple to the salicylic acid pathway Curiosity led us to examine the consequences of mis-expressing the incompatible OAK alleles from the Col-0 promoter in the putative ancestral domain of the leaf lamina. We introduced ProAt5g59670-Col:OAKBla and Pro At5g59670-Col:OAKSha chimeric transgenes into the Col-0 reference background, and crossed the transformants, which were wild type in appearance, to each other. As described above, performing this experiment with the OAK wild-type alleles from Bla-1 and Sha reproduced the Bla-1/Sha hybrid phenotype with petiole outgrowths. Co-expressing the Bla-1 and Sha OAK proteins from the Col-0 promoter resulted in a new incompatibility phenotype, ranging from patches of cell death visible to the naked eye on the leaf lamina and abbreviated inflorescences, to severely stunted plants (Figure 7d–7f). It is striking that the altered expression domain leads essentially to a diametrically opposite phenotype, ectopic cell death instead of ectopic cell proliferation.

Tissue necrosis and ectopic cell death are typical responses to pathogen infection that rely on salicylic acid signaling [49]. To determine whether the cell death we observed was associated with increased activity of this pathway, we used a transgene that drives constitutive expression of a bacterial salicylate hydroxylase, nahG, which converts salicylic acid to catechol [50]. The Pro35S:nahG transgene suppressed the cell death phenotype caused by co-expression of OAKBla-1 and OAKSha proteins from the Col-0 promoter, but had no effect on the ectopic outgrowths and other phenotypes seen when the proteins were expressed from their own promoters in Col-0 (Figure S12). This not only indicated that OAK proteins can couple to alternative downstream signaling pathways (as is known for the BAK1 RLK [51]), but also that the ancestral function might have involved detection of microbes, a known function of different RLKs [52–54]. Mutation of other key genes in disease resistance pathways (PAD4, EDS1, and NDR1) [49] had no effect on the aberrant phenotypes caused by co-

Figure 7. Requirement of OAK kinase activity and expression domain for hybrid phenotype. (a) Phenotypic distribution of T1 plants (n=90) expressing kinase dead (KD) or wild-type (WT) versions of OAK. (b) Crosses of Col-0 plants carrying Bla-1/Sh a POAK:OAK KD constructs. Representative F1 plants from crosses among five pairs of independent, phenotypically normal T1 plants are shown with alongside the parental lines. Scale bar = 1 cm. (c) Crosses of five pairs of phenotypically normal Col-0 plants transformed with P OAK:OAKSh a and P OAK:OAKBla-1, or (d,e) with P At5g59670:OAKSha and P At5g59670:OAKBla-1. Plants in (b-d) are 4-weeks old, in (e) 6-weeks old. Arrows in (f) indicate regions of cell death visible to the naked eye on a close-up of the F1 plant in (d). doi:10.1371/journal.pgen.1002164.g007
expression of the OAK alleles under either the OAK or the Col-0 At5g59670 promoter.

Discussion

We have identified a case of a single-gene incompatibility interaction that leads to multiple aberrant phenotypes in hybrids between A. thaliana accessions Bla-1 and Sha. The phenotypes include reduced stature, leaf twisting, a loss of apical dominance and ectopic outgrowths on the petioles in addition to a decrease in lifetime fitness as measured by seed set. In the genetic sense, the Bla-1 and Sha OAK alleles can be thought of behaving in an overdominant fashion, since the action of either allele (which can cause milder versions of the hybrid phenotype in a foreign background on their own) is enhanced by the other allele. However, considering that the phenotypes are not normally seen in the parents or in other hybrids, and that one of them is reduced growth, the alleles behave in an underdominant fashion when it comes to fitness, as measured by seed set under laboratory conditions.

The causal gene for the Bla-1/Sha incompatibility, OAK, is an RLK that is part of a highly variable tandem array, with evidence of gene conversion, duplications and deletions in the recent evolutionary past. OAK was formed by a whole-gene duplication event in a common ancestor of Bla-1 and Sha, with the additional duplication of a segment of coding DNA that now forms most of the OAK promoter. This gene duplication is present in approximately one third of A. thaliana accessions sampled, but the Bla-1 and Sha alleles themselves are rare. The new promoter changed the OAK expression domain from the leaf lamina to the leaf petiole. Although this change expression domain is required for manifestation of the OAK incompatibility, it is not in itself causal as the new promoter probably arose only once, and most accessions carrying the OAK gene are compatible with Bla-1 and Sha. Notably, the coding sequences that became part of the promoter include those coding for the second malecin-like domain, which has diverged between Bla-1, Sha and other accessions after the initial duplication. Changes in cis-regulatory sequences are an important source of interspecific variation [53], but such drastic intraspecific shifts in expression domains as we have observed are rare.

A function for OAK in disease resistance or development?

The A. thaliana genome encodes over 600 RLKs. Approximately two thirds of A. thaliana RLKs are predicted to contain structurally diverse extracellular domains [15], which often include LRRs [56]. These extracellular domains are involved in perceiving a wide range of ligands, including small proteins, steroids, and carbohydrates. The function and ligands of most plant RLKs are unknown, but known activities of LRR-RLKs include both control of plant development (e.g., BRI1 in brassinosteroid response [57], CLV1 in meristem maintenance [58] and ERECTA in pleiotropic patterning processes [59]) and microbe detection (e.g., Xa21, FLS2 and GmNARK [52–54]). The RLK genes constitute one of the most variable gene families in A. thaliana, which has been interpreted as many RLKs evolving in response to pathogen pressure [9]. Local and genome-wide duplications, along with gene conversion, have contributed to the expansion and diversification of RLKs in plants [12], and RLK genes are overrepresented in tandem arrays [15,60], although those with known roles in plant development are generally not located in tandem arrays [17].

Circumstantial evidence that might point to an interaction of OAK-like RLKs with microbes include the microarray results and the high variability of the OAK gene cluster. OAK does not appear to be required for normal development, since amiRNA-mediated knockdown of OAK activity has no obvious adverse effects. However, it is also possible that OAK acts redundantly in plant development given that the incompatibility phenotype manifests itself primarily as morphological abnormalities. In addition, the mis-expression experiments using the Col-0 promoter revealed that OAKs can trigger typical SA-like dependent cell death as is often seen in response to pathogen attack, although OAK coupling to downstream signaling pathways may be dependent on the expression pattern of alternative interactors. Following the BAK1 paradigm [51], it is conceivable that the availability of OAK interaction partners determine its activity in plant development versus microbe-interactions. The similarity of the OAK extracellular domains to the carbohydrate-binding protein malectin [41] might indicate that OAK-like RLKs interact with carbohydrates found on the surface of microbes. Alternatively, their function might be detection of damaged self, according to the concept of indirect recognition of pathogens through damage-associated molecular patterns (DAMPs) [61]. A role for OAK in plant immunity through perception of self damage would be reminiscent of previously reported cases of hybrid incompatibility that involve disease resistance genes [6–8,62].

Causes for increased OAK activity in hybrids

Some RLKs function as hetero- or homodimers, with auto- and trans-phosphorylation required for function of the complex. For example, BAK1 and BRI1 form heteromultimers, and a multi-step pathway involving auto- and trans-phosphorylation events activates downstream signaling [63]. Our experiments with kinase-dead versions demonstrated that kinase activity is important for OAK function. The limited effects of the kinase-dead Sha allele in the Bla-1 background, and vice versa, indicate partial complementation by the opposite kinase-active allele, which is suggestive of heterooligomeric or multimer formation. In addition, the suppression of the hybrid phenotypes by expression of the Bla-1 or Sha OAK extracellular domain alone provides further support for this scenario.

We do not know whether the change in expression pattern associated with the acquisition of a new promoter by the Bla-1 and Sha OAK alleles subsequently became subject to positive selection, or whether these alleles lack a beneficial function all together. However, the fact that the unusually high divergence in sequence between the two alleles is largely restricted to the second malecin-like domain suggests positive selection or a gene conversion event. We speculate that these sequence changes also altered the affinity for potential ligands. The fact that the Bla-1 and Sha proteins on their own can cause a hybrid-like phenotype, albeit less effectively than when they are combined, suggests that each protein on its own can interact with this potential, unknown ligand. We speculate that OAK heterodimers have increased affinity for such a ligand, leading to ectopic activation of the downstream signaling pathway and aberrant development.

Evolution of incompatible OAK alleles

Several incompatibilities in F1 and F2 hybrids have recently been linked to disease resistance (R) genes. At least one of the A. thaliana factors, and likely another in A. thaliana and rice each, appears to be encoded in a highly polymorphic cluster of NB-LRR genes, the most common class of R genes, and at the same time the most polymorphic gene family in plants [6,8,9,62,64,65]. Indeed, more broadly, copy number variation is a recurring factor in reproductive isolation [66]. It has been proposed that the occurrence of disease resistance genes in clusters is critical for
generating diversity of resistance specificities, because the tandem arrays support high rates of gene conversion and illegitimate recombination [67]. Indeed, complex histories of transposon insertions, translocations, and gene duplications and rearrangements have also contributed to the formation of \textit{NB-LRR} gene clusters [11,13,16,18,19]. \textit{RLK} genes share with \textit{NB-LRR} genes the frequent occurrence in tandem arrays and extreme diversity [9,12,13]. The complex evolutionary history of the \textit{OAK} cluster is thus not atypical for this gene family.

Most hybrid incompatibilities described so far involve multiple loci and as such are classical examples of the Bateson, Dobzhansky and Muller model where derived alleles of two or more genes interact to produce underdominant fitness outcomes (e.g.[8,21, 62,68]). In contrast, the incompatibility we describe here is due to interaction of two different alleles at a single locus. Due to the high level of polymorphisms, it is difficult to know what the ancestral allele at the \textit{OAK} locus looked like immediately after duplication. The incompatible \textit{OAK} alleles may have evolved through mutations within both the Sha and Bla-1 lineages, with the current alleles remaining compatible with the ancestral allele. Alternatively, all important mutation and gene conversion events may have occurred in only one lineage, through multiple intermediate allelic forms that were never incompatible with the immediately ancestral allele [69]. Either way, evolution of the current situation would not require that plants passed through a fitness valley with heterozygosity for the two incompatible \textit{OAK} alleles.

Conclusions

Not many cases of single-gene hybrid incompatibility have been described in plants: in rice, incompatible alleles of the \textit{S5} locus cause most hybrids between the japonica and indica varieties to be female sterile [33]. It is not inconceivable that heterodimers are involved, similar to what appears to be the case for \textit{OAK}, and dimer formation may be an important pre-condition for evolution of single-gene incompatibilities. We note that passage through a fitness valley is not required so long as the genetic changes causing incompatibility evolve in multiple steps within separate genetic backgrounds. In this way, two alleles could cause underdominance for fitness and reduce or abolish gene flow, but only upon crossing of lines that have diverged independently from a common ancestor. If there were strong positive selection for two different alleles that caused underdominance or sterility in hybrids, then they could eventually contribute to a speciation event.

In animals single-gene single-generation speciation occurs in snails, where shell chirality is maternally determined, with opposite chirality forming a strong pre-mating barrier [70,71]. Extenuating factors that could allow rapid speciation based on a single locus, even after one generation, include transient silencing of genes, for example, by parental imprinting, or incomplete sterility of the hybrid. If an incompatible allele arises, but is silenced for one generation, this would allow for the production of multiple offspring that are pre-or post-zygotically incompatible with individuals carrying the ancestral allele. Offspring with the new allele can self or interbreed to establish a subpopulation before this allele is lost again by genetic drift. Similarly, if the heteroallelic combination is sublethal, then F2 offspring homozygous for the new allele can be produced. If, in turn, the homozygous form is subject to positive selection, the allele may become established in the population [70]. Such a scenario is particularly applicable to self-fertilizing species such as \textit{Arabidopsis thaliana}.

Whether the sort of developmental abnormalities we have observed in Bla-1/Sha F1 hybrids can contribute to cumulative reproductive isolation is of course not known. Nevertheless, that \textit{OAK} has the potential to greatly reduce reproductive success can be inferred from the severe phenotypes in some plants transformed with active \textit{OAK} constructs, the necrosis seen when incompatible \textit{OAKs} are co-expressed from the Col-0 promoter, and the decrease in lifetime fitness as measured via seed set. All together, we propose that the occurrence of genes in variable tandem repeats, such as \textit{NB-LRR} genes in several hybrid necrosis cases [6,8,62], or \textit{RLKs} as in the present case, predisposes them to being sources for the creation of novel hybrid phenotypes. Whether, as with other mutations, these are normally disadvantageous or not, will require further systematic analyses of hybrid incompatibilities in a broad range of taxa.

Materials and Methods

Plant material

Bla-1 (N28079) and Sha (N28735) were obtained from the European Arabidopsis Stock Centre. Plants were grown at 16°C with 16 hours light, or 23°C with 8 or 16 hours of light, as indicated. Transgenic seedlings were selected on soil by BASTA resistance, and at least 90 T1 plants phenotyped, unless otherwise indicated.

Transgenic plants

Genomic constructs spanned sequences from immediately downstream of the translational stop codon of the preceding gene to 200 bp downstream of the predicted translational stop. AmiRNAs were designed using WMD3 (http://wmd3.weigel-world.org/). Constructs were transformed into plants by the \textit{Agrobacterium tumefaciens} floral-dip method [72] using strain GV3101 pMP90RK or ASE. For reporter gene analysis, the promoter region between the stop codon of the previous gene and the translational start codon of the \textit{OAK} homolog was inserted into pGWB433 using Gateway LR clonase (InVitrogen, Darmstadt, Germany).

Seed set

Independent \textit{PROOAK}:OAK\textit{Bla-1} and \textit{PROOAK}:OAK\textit{Sha} T1 plants in Col-0 that did not show any morphological defects were crossed to each other to create F2 populations, which were raised in randomly distributed individual pots without selection for the transgenes. Plants were genotyped, and seeds collected from each plant after three months of growth and weighed. The weight of individual seeds was determined by weighing 500 seeds for each of three plants per genotype, and total and individual seed weight were used to calculate total seed number per plant.

Humidity assay

Plants were grown in 23°C (long days) at 65% ambient humidity; or under mild drought-stress with minimal watering (but equal ambient humidity); or in saturated humidity with water surrounding the pots and the tray covered.

Histology

Bla-1 and Bla-1/Sha petioles were fixed in 3.7% formaldehyde, 5% acetic acid, 50% ethanol, embedded in an ASP300 (Leica, Nussloch, Germany) tissue processor in paraffin. Transverse sections of 8 µm thickness, stained with 0.02% Toluidine Blue after dewaxing, were examined with a Zeiss Axioskop 2 microscope.

Callus assay

Seeds were stratified for one week on ½ strength MS plates. Seedlings were grown in Percival LE Intellus chambers (Perry, IA,
USA) under 23°C long days until the 4-6 leaf stage. At least 40 transverse sections per genotype of leaves (1 mm thick) and petioles (2 mm thick) were placed on callus induction medium (3.1 g/L Gamborg’s B5 salts, 2% glucose, 2.6 mM MES, pH 5.7, 0.8% agar) with 2.2 μM to 22 nM 2,4-dichlorophenoxyacetic acid (2,4-D) and 200 nM to 200 pM kinetin. Callus formation was assessed after 12 days.

Expression analysis
RNA was extracted from leaves of individual plants using the Qiagen (Hilden, Germany) Plant RNAeasy Mini kit. One μg of RNA was DNaseI treated, and cDNA synthesized with hexamer primers (Fermentas RevertAid kit, St. Leon-Rot, Germany). qRT-PCR was performed with Invitrogen (St. Louis, MO, USA) SYBR Green PCR Mastermix and the MJR Opticon Continuous Fluorescence Detection System (Bio-Rad, Hercules, CA, USA). Two technical replicates were performed per sample. Expression was normalized to β-TUBULIN-2 (At5g62690) and an amplification efficiency of 2.0 per cycle was used in the calculations. The average across three biological replicates is shown with standard deviation. Two technical replicates were performed per sample. Expression was normalized to β-TUBULIN-2 (At5g62690) and an amplification efficiency of 2.0 per cycle was used in the calculations. The average across three biological replicates is shown with standard deviation.

GUS staining
Twelve-day old seedlings grown on ½ strength MS plates with kanamycin selection were fixed in 90% acetone on ice for 20 minutes. X-gluc stained tissue [72] was examined with a Leica MZFLIII microscope.

Microarrays
Affymetrix (Santa Clara, CA, USA) ATH1 microarrays were probed as described [73].

Genetic mapping
Coarse mapping was performed with the Sequenom (San Diego, CA, USA) MassARRAY platform. For high-resolution mapping, approximately 750 F2 and F3 plants were genotyped with microsatellite and CAPS markers [72].

Phylogenetic and statistical analyses
For the sliding window analysis of divergence, amino acid sequences were aligned with MUSCLE (http://www.ebi.ac.uk/Tools/muscle/) and nucleotide sequences with BlastX (http://blast.ncbi.nlm.nih.gov/Blast.cgi). For analysis of population structure, nucleotide sequences were aligned with MUSCLE (http://www.ebi.ac.uk/Tools/muscle/) and nucleotide sequences with BlastX (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Supporting Information

Figure S1 Bla-1/Sha incompatibility decreases seed set. (a) Normal appearing Col-0 plants that are either non-transgenic or carry only a single OAK transgene. The phenotype of F1 plants with both OAK transgenes is comparable to (b) Sha/Blal 1 F1 plants. (c) Total seed set after three months shown as box and whisker plots. Boxes cover the first and third quartile, and the whiskers represent values that are not more than 1.5 times the interquartile range. Two-tailed, unequal variance t-test showed statistical equivalence of seed set between wild-type plants and those with a single OAK transgene, and highly significant reduction of seed set in plants carrying both transgenes. (TIF)

Figure S2 High humidity suppresses outgrowth formation. Bla-1/Sha F1 plants were grown for 3 and a half weeks under either high humidity (covered with a dome and surrounded by water), normal humidity (controlled 65% humidity), or under drought stress conditions (65% humidity but minimal watering). Two representative leaves per treatment are shown. Outgrowths are indicated by arrows. (TIF)

Figure S3 Effect of auxin and cytokinin concentration on callus formation. Callus formation at 12 days for transverse sections of leaves and petioles of Bla-l, Bla-l/Sha F1 and Sha. Three representative tissue pieces are shown per accession and hormone concentration. (TIF)

Figure S4 Mapping interval for the Bla-l/Sha outgrowth causal gene. (a) Positional cloning markers used according to the cognate genes and position in Mb in reference accession Col-0. (b) The genes in reference accession Col-0 in the final mapping interval, with protein kinases marked in light grey and the RLKs highlighted in mid-grey. (TIF)

Figure S5 AmiRNA knockdown of OAK rescues the hybrid phenotype. AmiRNAs designed against each RLK in the OAK cluster from Bla-l (a) or Sha (b) were transformed into Bla-l/Sha F1 plants and plants heterozygous at the RLK locus identified in the next generation. One representative plant per line is shown. Scale bar = 1 cm. (TIF)

Figure S6 Potential LRR and malectin-like domains in OAK. (a) The consensus for plant-specific LRR domains is given below according to (Kobe, B. & Kajava, A.V. The leucine-rich repeat as a protein recognition motif. Curr. Opin. Struct. Biol. 11, 725-32; 2001), with residues conserved in over 50% of proteins shown in uppercase. Leucine residues from OAK at conserved positions are indicated in yellow, with other conserved residues highlighted in green. Less conserved residues or residues similar to those conserved are highlighted in light grey. (b) Predicted malectin-like domains (Schallus, T. et al. Malectin: a novel carbohydrate-binding protein of the endoplasmatic reticulum and a candidate player in the early steps of protein N-glycosylation. Mol. Biol. Cell 19, 3404-14; 2008) in OAKBla-1 and OAKSha. Although the amino acid sequence identity is low (11–15%), the secondary structure is more highly conserved, and the probability scores are very high. (DOC)

Figure S7 Divergence of RLK orthologs and paralogs. (a) Comparison of pairwise amino acid divergence between OAKBla-1 and OAKSha and between all RLKs in this cluster. (b) Comparison
of pairwise amino acid divergence between OAK and At5g39670a alleles from Bla-1 and Sha.

**Figure S8** Compatibility between OAK-containing accessions. Cytoscape (Shannon P, Markiel A, Ozier O, Baliga NS, Wang J, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498–2504) representation of crosses performed between OAK-containing accessions (names indicated in circles). Node color on the periphery indicates the haplotype group of the second malelten domain. Cvi-0, Cdm-0, ICE50, ICE226 and ICE228 alleles switch between haplotype groups within the second malelten domain, and are shown in intermediate colors. Absence of color indicates that the haplotype group is not known. Compatible hybrid combinations are indicated by grey edges, while incompatible interactions with outgrowths are represented by black (hybrid phenotype of intensity similar to Sha/Bla-1), red (phenotypic onset early as for Sha/Bla-1 but milder leaf twisting and loss of apical dominance) or blue (late onset of outgrowths with no other incompatible phenotypes) edges.

**Figure S9** Much of the OAK promoter is derived from a duplicated region of RLK coding sequence. Top 15 hits from LALIGN (http://www.ch.embnet.org/software/LALIGN_form.html) are shown according to position in the Bla-1 OAK promoter, linked to a color-matched box indicating position in the Col-0 RLK cluster.

**Figure S10** Alignment of the OAK proteins from Sha, Leo-1 and Bla-1. Amino acid differences between the three OAK proteins are indicated in purple (where Sha differs from Leo-1 and Bla, which are both incompatible with Sha), in cyan (where Bla-1 differs from Sha and Leo-1) and in red (where Leo-1 differs from Sha and Bla-1). Alignment was performed with CLUSTALW (Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, et al. (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31: 3497–3500).

**Figure S11** Expression of the OAK extracellular domain in hybrid plants can reduce the severity of aberrant phenotypes. The extracellular domains of OAKSha, OAKBla or At5g39670a under control of their native promoters or the 35S promoter were expressed in plants containing accessions (names indicated in circles). Node color on the periphery indicates the haplotype group of the secon

**Figure S12** Mis-expressed OAK couples to the salicylic acid signalling pathway. (a) Pro35S:nahG when introduced into Pro35S:OAKSha. Pro35S:nahG rescues the cell death phenotype. (b) Pro35S:nahG when introduced into Pro35S:OAKBla. Pro35S:nahG does not suppress the outgrowths, leaf twisting or loss of apical dominance.

**Table S1** Outgrowth formation in short-day grown Bla-1 and Bla-1/ShA F1 hybrids. Plants grown in 23°C short-day conditions were scored regularly for exotropic outgrowths on the petioles.

**Table S2** Outgrowth and lesioning phenotypes are correlated with reduced vegetative biomass. Average fresh weight of segregating sibling F2 plants grown at 16°C for 5 weeks is reported.

**Table S3** Overrepresented GO categories as determined by AmiGO among genes up- or down-regulated in Bla-1/ShA F1 hybrids.

**Table S4** Top ten up- and down-regulated genes in Bla-1/ShA F1 hybrids compared to parental genotypes. See Table S5 for more information.

**Table S5** Differentially regulated genes in Bla-1/ShA F1 hybrids compared to parental genotypes.

**Table S6** Similarity of OAK and related alleles. Nucleotide identity in percent is given on top, with amino acid identity given on bottom.

**Table S7** Survey of A. thaliana accessions for OAK duplication.

**Acknowledgments**

We thank Michael Hothorn (Salk Institute) for protein domain predictions; Eunyoung Chae for the identification of the Bla-1/Kond incompatibility; Eunyoung and Carmen Martin Pizarro for access to their many crosses; Yangol Guo for the A. lyrata sequences and population genetic analyses; Suna Kim for assistance with STRUCTURE analysis; Stephan Ossowski for assistance with WMD3; and Richard Clark, John Willis, and Dani Zamir for critical reading of the manuscript.

**Author Contributions**

Conceived and designed the experiments: LMS DW KB. Performed the experiments: LMS KB. Analyzed the data: LMS. Wrote the paper: LMS DW.
14. Meyers BC, Kosik A, Grego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis. Plant Cell 15: 839–854.

15. Shiu SH, Bleecker AB (2005) Expansion of the receptor-like kinase/Pelle gene family and receptor kinase proteins in Arabidopsis. Plant Physiol 139: 530–543.

16. Baumgarten A, Cannon S, Spangler R, May G (2003) Genome-level evolution of resistance genes in Arabidopsis thaliana. Genetics 165: 309–319.

17. Shiu SH, Karloowski WM, Pan R, Zheng YH, Mayer KFX, et al. (2004) Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. Plant Cell 16: 1220–1234.

18. Kuang H, Woo SS, Meyers BC, Nevo E, Michelmore RW (2004) Multiple genetic processes result in heterogeneous rates of evolution within the major Kuster lineage of resistance genes in rice. Plant Cell 16: 2070–2084.

19. Leister D (2004) Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends Genet 20: 116–122.

20. Rieseberg LH, Willis JH (2007) Plant speciation. Science 317: 910–914.

21. Bomblies K, Weigel D (2007) Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. Nat Rev Genet 8: 382–393.

22. Joshi MG (1972) Occurrence of genetic tumours in...
Supplementary Online Material for

Smith et al., Complex evolutionary events at a tandem cluster of *Arabidopsis thaliana* genes resulting in a single-locus genetic incompatibility
Supplementary Tables

**Supplementary Table 1.** Outgrowth formation in short-day grown Bla-1 and Bla-1/Sha F₁ hybrids.

| Genotype      | Experiment | n  | Plants with outgrowths (%) | First leaf with outgrowths |
|---------------|------------|----|----------------------------|-----------------------------|
| Bla-1         | 1          | 40 | 65                         | 24.1 ± 2.5                  |
| Bla-1         | 2          | 28 | 0                          | n/a                         |
| Bla-1/Sha F₁  | 1          | 39 | 100                        | 11.8 ± 1.8                  |

Plants grown in 23°C short-day conditions were scored regularly for extopic outgrowths on the petioles.
**Supplementary Table 2.** Outgrowth and lesioning phenotypes are correlated with reduced vegetative biomass

| Weight* (± standard deviation) | Without outgrowths (n) | With outgrowths (n) |
|-------------------------------|------------------------|---------------------|
| Not lesioned                  | 1.58 ± 0.53 g (27)     | 1.12 ± 0.44 g (39)  |
| Lesioned                      | 0.66 ± 0.26 g (16)     | 0.74 ± 0.29 g (32)  |

*average fresh weight of segregating sibling F$_2$ plants grown at 16°C for 5 weeks is reported.
**Supplementary Table 3.** Overrepresented GO categories as determined by AmiGO among genes up- or down-regulated in Bla-1/Shal F1 hybrids.

### Up-regulated genes

| GO category                          | Enrichment p-value |
|--------------------------------------|--------------------|
| response to other organism           | 9.35 x 10^{-5}     |
| response to stimulus                 | 4.74 x 10^{-5}     |
| response to biological stimulus      | 3.13 x 10^{-5}     |
| response to jasmonic acid stimulus   | 6.20 x 10^{-4}     |
| response to salicylic acid stimulus  | 4.70 x 10^{-3}     |
| multi-organism processes             | 1.88 x 10^{-4}     |
| catalytic activity                   | 1.29 x 10^{-5}     |

### Down-regulated genes

| GO category              | Enrichment p-value |
|--------------------------|--------------------|
| external encapsulating structure | 7.37 x 10^{-3} |
| cell part                | 9.93 x 10^{-3}     |
| catalytic activity       | 1.45 x 10^{-4}     |
**Supplementary Table 4.** Top ten up- and down-regulated genes in Bla-1/Sha F1 hybrids compared to parental genotypes. See Supplementary Table 7 for more information.

| Avg. fold change⁵ | Up-regulated genes                                                                                           | Down-regulated genes                                                                                           |
|-------------------|---------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|
| 6.3               | AT1G13470 Unknown protein                                                                                     | 42.9 AT1G72910/AT1G72930 putative disease resistance proteins (TIR-NBS class)                                  |
| 6.2*              | AT1G14870/AT1G14880 Uncharacterized protein                                                                   | 26.3* AT1G31580 ECS1                                                                                                |
| 4.0               | AT1G56140/AT1G56130/AT1G56120 Leucine-rich repeat protein kinases                                              | 19.3 AT4G02850 phenazine biosynthesis PhzC/PhzF family protein                                                   |
| 3.9               | AT3G28290/AT3G28300 AT14A’s, sequence similarity to integrins                                                | 15.8 AT4G05050 UBQ11 (UBIQUITIN 11); protein binding                                                           |
| 3.8               | AT3G48640 Unknown protein                                                                                     | 12.9 AT1G66690/AT1G66700 S-adenosyl-L-methionine:carboxyl methyltransferase family protein (AT1G66690); PXMT1; S-adenosylmethionine-dependent methyltransferase (AT1G66700) |
| 3.6               | AT2G18660 EXLB3 (EXPANSIN-LIKE B3 PRECURSOR)                                                                  | 12.7 AT4G29200 beta-galactosidase                                                                                  |
| 3.5               | AT4G23220 protein kinase family protein                                                                         | 11.6                                                                                                           |
| 3.3               | AT1G22590 AGL87; transcription factor                                                                          | 11.1 AT3G44430 Unknown protein                                                                                   |
| 3.2               | AT5G54610 ANK (ANKYRIN); protein binding                                                                        | 9.0 AT2G01090 ubiquinol-cytochrome C reductase complex 7.8 kDa protein, putative / mitochondrial hinge protein, putative |
| 3.2*              | AT5G55450 proteinase inhibitor/seed storage lipid transfer protein (LTP) family protein                         | 7.6 AT1G48598/AT1G48600 frame 31) (AT1G48598); phosphoethanolamine N-methyltransferase 2, putative (NMT2) (AT1G48600) |
The smaller ‘fold change’ between the parent and hybrid is reported when there was no significant difference between the parental lines. In the remaining cases, indicated with an asterisk, the change relative to the average of the parents is given.
**Supplementary Table 5.** Differentially regulated genes in Bla-1/Sha F₁ hybrids compared to parental genotypes.

| inverse FC | Average FC | Average ppf | Average P value | Array Element | Locus Identifier | Annotation |
|-----------|------------|-------------|----------------|---------------|-----------------|------------|
| 26.31578947 | 0.038 | 0 | 0 | 256497_at | AT1G31580 | ECS1 |
| 4.33557338 | 0.23065 | 0.00595 | 0.00005 | 257365_x_at | AT2G26020 | PDF1.2b (plant defensin 1.2b) |
| 3.74181478 | 0.26725 | 0.0057 | 0.00005 | 249052_at | AT5G44420 | PDF1.2 (Low-molecular-weight cysteine-rich 77) |
| 3.702332469 | 0.2701 | 0.0004 | 0 | 255852_at | AT1G66970 | glycerophosphoryl diester phosphodiesterase family protein |
| 3.220611916 | 0.3015 | 0.00095 | 0 | 249942_at | AT5G22300 | NIT4 (NITRILASE 4) |
| 3.19539626 | 0.31295 | 0.01865 | 0.00015 | 258277_at | AT3G26830 | PAD3 (PHYTOALEXIN DEFICIENT 3); oxygen binding |
| 2.988196623 | 0.33465 | 0.00175 | 0 | 263046_at | AT2G05380 | GRP3S (GLYCINE-RICH PROTEIN 3 SHORT ISOFORM) |
| 2.754062242 | 0.3631 | 0.0022 | 0 | 266275_at | AT2G05380 | tropinone reductase, putative / tropine dehydrogenase, putative |
| 2.595380223 | 0.3853 | 0.0028 | 0 | 252698_at | AT3G43670 | copper amine oxidase, putative |
| 2.565418163 | 0.3898 | 0.00245 | 0 | 248377_at | AT1G51720 | similar to Os07g0467200 [Oryza sativa (japonica cultivar-group)] (GB:NP_001059590.1); similar to hypothetical protein Os1_025030 [Oryza sativa (indica cultivar-group)] (GB:EAZ03798.1); contains domain PTHR13680 (PTHR13680); contains domain PTHR13680:SF1 (PTHR13680:SF1) |
| 2.551671345 | 0.3919 | 0.00335 | 0 | 246420_at | AT1G66970 | similar to unknown protein [Arabidopsis thaliana] (TAIR:AT3G03010.1); similar to unknown protein [Arabidopsis thaliana] (TAIR:AT3G03010.2); similar to hypothetical protein [Vitis vinifera] (GB:CAN83813.1); contains InterPro domain Peptidyl-tRNA hydrolase, PTH2 (InterPro:IPR002833) |
| 2.509725185 | 0.39845 | 0.0029 | 0 | 260151_at | AT1G52910 | similar to unknown protein [Arabidopsis thaliana] (TAIR:AT3G15480.1); similar to unknown protein [Populus trichocarpa] (GB:ABK94458.1); contains InterPro domain Peptidyl-tRNA hydrolase DUF1218 (InterPro:IPR009606) |
| 2.501876407 | 0.3997 | 0.00285 | 0 | 260693_at | AT4G12320; AT4G12310 | proton-dependent oligopeptide transport (POT) family protein |
| 2.501876407 | 0.3997 | 0.00325 | 0 | 257880_at | AT4G12320; AT4G12310 | AAET1/ACN1 (ACYL-ACTIVATING ENZYME 7); AMP binding / acetate-CoA ligase |
| 2.409058058 | 0.4151 | 0.00405 | 0 | 254835_s_at | AT4G14110 | basic helix-loop-helix (bHLH) family protein |
| 2.396357537 | 0.4173 | 0.00385 | 0 | 263883_at | AT2G21830 | DC1 domain-containing protein |
| 2.385496183 | 0.4192 | 0.0056 | 0 | 259331_at | AT3G03840 | auxin-responsive protein, putative |
| 2.362111728 | 0.42335 | 0.0043 | 0 | 245331_at | AT4G14110 | basic helix-loop-helix (bHLH) family protein |
| Gene ID          | Log2 FC | FDR  | P-value | Description                                                                 |
|-----------------|---------|------|---------|-----------------------------------------------------------------------------|
| AT1G55370       | 2.355   | 0.42455 | 0.0059 | carbohydrate binding / catalytic                                           |
| AT1G03220       | 2.3345  | 0.42835 | 0.0046 | lipase, putative                                                            |
| AT2G27360       | 2.3036  | 0.43415 | 0.0078 | GTP-binding family protein                                                  |
| AT4G02790       | 2.2596  | 0.4484  | 0.00495| cysteine proteinase, putative                                               |
| AT3G11400       | 2.2054  | 0.4596  | 0.00595| EIF3G1 (eukaryotic translation initiation factor 3G1); RNA binding / translation initiation factor |
| AT2G27420       | 2.184   | 0.4635  | 0.01375| phosphoglycerate dehydrogenase                                              |
| AT5G12110       | 2.1215  | 0.46835 | 0.01405| pentatricopeptide (PPR) repeat-containing protein                          |
| AT3G56200       | 2.0665  | 0.4839  | 0.00925| amino acid transporter family protein                                       |
| AT3G07800       | 2.0130  | 0.49675 | 0.009   | thymidine kinase, putative                                                  |
| AT1G17745       | 2.0112  | 0.4972  | 0.01355| elongation factor 1B alpha-subunit 1 (eEF1Balpha1)                         |
| AT4G19100       | 2.0061  | 0.4985  | 0.01155| disease resistance protein (TIR-NBS-LRR class), putative                    |
| AT4G17040       | 1.9946  | 0.50135 | 0.01015| ATP-dependent Clp protease proteolytic subunit, putative                    |
| AT2G42740       | 1.986   | 0.5035  | 0.0123  | RPL16A (ribosomal protein large subunit 16A); structural constituent of ribosome |
| AT2G43510       | 1.9770  | 0.5058  | 0.0116  | ATT11 (ARABIDOPSIS THALIANA TRYPsin INHIBITOR PROTEIN 1)                    |
| AT5G45420       | 1.9642  | 0.5091  | 0.0217  | myb family transcription factor                                            |
| Gene ID  | Fold Change | P-value  | E-value | Description and Function |
|---------|-------------|----------|---------|--------------------------|
| AT1G21790 | 1.960976566 | 0.50995  | 0.00005 | Similar to unnamed protein product [Vitis vinifera] (GB:CA061872.1); contains InterPro domain TRAM, LAG1 and CLN8 homology; (InterPro:IPR006634) |
| AT1G72030 | 1.935171746 | 0.51675  | 0.00001 | GCN5-related N-acetyltransferase (GNAT) family protein similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G30160.2); contains InterPro domain Protein of unknown function DUF295 (InterPro:IPR005174) |
| AT1G05540 | 1.9289203   | 0.52005  | 0.00001 | Similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G30160.2); contains InterPro domain Protein of unknown function DUF295 (InterPro:IPR005174) |
| AT1G01430 | 1.9223349   | 0.52015  | 0.00001 | Similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G30160.2); contains InterPro domain Protein of unknown function DUF295 (InterPro:IPR005174) |
| AT1G01430 | 1.921045049 | 0.52055  | 0.00001 | Similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G30160.2); contains InterPro domain Protein of unknown function DUF295 (InterPro:IPR005174) |
| AT1G01430 | 1.916075877 | 0.5219   | 0.00001 | Similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G30160.2); contains InterPro domain Protein of unknown function DUF295 (InterPro:IPR005174) |
| AT1G01430 | 1.911314985 | 0.5232   | 0.00001 | Similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G30160.2); contains InterPro domain Protein of unknown function DUF295 (InterPro:IPR005174) |
| AT2G02020 | 1.902578519 | 0.5256   | 0.00001 | Similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G30160.2); contains InterPro domain Protein of unknown function DUF295 (InterPro:IPR005174) |
| AT5G37740 | 1.905627497 | 0.52945  | 0.00001 | Similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G30160.2); contains InterPro domain Protein of unknown function DUF295 (InterPro:IPR005174) |
| AT5G61950 | 1.908634962 | 0.53515  | 0.00001 | Similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G30160.2); contains InterPro domain Protein of unknown function DUF295 (InterPro:IPR005174) |
| AT2G02020 | 1.905627497 | 0.53595  | 0.00001 | Similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G30160.2); contains InterPro domain Protein of unknown function DUF295 (InterPro:IPR005174) |
| AT3G46980 | 1.908634962 | 0.53655  | 0.00001 | Similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G30160.2); contains InterPro domain Protein of unknown function DUF295 (InterPro:IPR005174) |
| AT3G63330 | 1.839418744 | 0.54365  | 0.00001 | Similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G30160.2); contains InterPro domain Protein of unknown function DUF295 (InterPro:IPR005174) |
| AT5G02180 | 1.842638659 | 0.5489   | 0.00001 | Similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G30160.2); contains InterPro domain Protein of unknown function DUF295 (InterPro:IPR005174) |

**Function Annotations:**
- **AT1G21790:** Contains InterPro domain TRAM, LAG1 and CLN8 homology.
- **AT1G72030:** GCN5-related N-acetyltransferase (GNAT) family protein.
- **AT1G05540:** Similar to unnamed protein product [Vitis vinifera] (GB:CA061872.1).
- **AT1G01430:** Contains InterPro domain Protein of unknown function DUF295.
- **AT2G02020:** Proton-dependent oligopeptide transport (POT) family protein.
- **AT5G37740:** Contains InterPro domain Protein of unknown function DUF295.
- **AT5G61950:** Contains InterPro domain Protein of unknown function DUF295.
- **AT2G02020:** Contains InterPro domain Protein of unknown function DUF295.
- **AT3G46980:** Contains InterPro domain Protein of unknown function DUF295.
- **AT3G63330:** Contains InterPro domain Protein of unknown function DUF295.
- **AT5G02180:** Contains InterPro domain Protein of unknown function DUF295.

**Additional Information:**
- AT1G21790 is similar to unnamed protein product [Vitis vinifera] (GB:CA061872.1).
- AT1G05540 contains InterPro domain TRAM, LAG1 and CLN8 homology.
- AT1G01430 contains InterPro domain Protein of unknown function DUF295.
- AT2G02020 contains InterPro domain Protein of unknown function DUF295.
- AT5G37740 contains InterPro domain Protein of unknown function DUF295.
- AT5G61950 contains InterPro domain Protein of unknown function DUF295.
- AT2G02020 contains InterPro domain Protein of unknown function DUF295.
- AT3G46980 contains InterPro domain Protein of unknown function DUF295.
- AT3G63330 contains InterPro domain Protein of unknown function DUF295.
- AT5G02180 contains InterPro domain Protein of unknown function DUF295.
| Gene ID | Log2 Ratio | Fold Change | FDR | P-Value | Description |
|---------|------------|-------------|-----|---------|-------------|
| 260453_s_at | 1.820167455 | 0.5494 | 0.02115 | 0.0002 | AT1G72510; AT2G09970 |
| 248082_at | 1.819505095 | 0.5496 | 0.0199 | 0.0002 | AT5G55400 |
| 263275_at | 1.810458957 | 0.55125 | 0.0301 | 0.0002 | ALDH6B2 (Aldehyde dehydrogenase 6B2); 3-chloroallyl aldehyde dehydrogenase |
| 246966_at | 1.807337791 | 0.5533 | 0.02815 | 0.0002 | AT5G24850 |
| 265139_at | 1.804891255 | 0.55405 | 0.0217 | 0.0002 | AT1G51310 |
| 249521_at | 1.804402743 | 0.5542 | 0.02225 | 0.0002 | AT5G38690 |
| 266038_at | 1.79937022 | 0.55575 | 0.0194 | 0.0002 | AT3G02220 |
| 254431_at | 1.796299623 | 0.5567 | 0.02275 | 0.0002 | AT4G20840 |
| 266038_at | 1.791312136 | 0.55825 | 0.0252 | 0.0002 | AT5G04760 |
Identical to F-box/Kelch-repeat protein At5g49000 [Arabidopsis Thaliana] (GB:Q9FI70;GB:Q8GY04); similar to kelch repeat-containing F-box family protein [Arabidopsis thaliana] (TAIR:AT4G39550.1); similar to 117M18_27 [Brassica rapa] (GB:AAZ66946.1); contains InterPro domain Kelch repeat type 1 (InterPro:IPR006652); contains InterPro domain Kelch-type beta propeller (InterPro:IPR015915); contains InterPro domain Cyclin-like F-box (InterPro:IPR001810); contains InterPro domain Kelch related (InterPro:IPR013089); contains InterPro domain Galactose oxidase/kelch, beta-propeller (InterPro:IPR011043); contains InterPro domain Lipid-binding START (InterPro:IPR002913).

**Identified Genes:***

| Gene ID     | Description                                                                                                                                 |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| AT1G26560   | glycosyl hydrolase family 1 protein                                                                                                                                                                      |
| AT1G47920   | syntaxin-related family protein                                                                                                                                                                          |
| AT3G52940   | FK (FACKEL); delta14-sterol reductase                                                                                                                                                                      |
| AT5G08640   | FLS (FLAVONOL SYNTHASE)                                                                                                                                                                                  |
| AT5G28007   | nodulin MtN3 family protein                                                                                                                                                                              |
| AT3G19260   | LAG1 HOMOLOG 2 (LONGEVITY ASSURANCE GENE1 HOMOLOG 2)                                                                                                                                                     |
| AT2G36230   | APG10 (ALBINO AND PALE GREEN 10); 1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide isomerase                 |
| AT1G59960   | allinase family protein                                                                                                                                                                                  |
| AT4G24670   | glycosyl hydrolase family 1 protein                                                                                                                                                                        |
| AT1G26560   | syntaxin-related family protein                                                                                                                                                                          |
| AT5G17170   | ENH1 (ENHANCER OF SOS3-1); metal ion binding                                                                                                                                                             |
| AT3G52940   | FK (FACKEL); delta14-sterol reductase                                                                                                                                                                      |
| AT5G58310   | hydrolase, alpha/beta fold family protein                                                                                                                                                                  |
| AT3G19260   | LAG1 HOMOLOG 2 (LONGEVITY ASSURANCE GENE1 HOMOLOG 2)                                                                                                                                                     |
| AT1G24575   | unknown protein                                                                                                                                                                                           |
| AT5G08640   | FLS (FLAVONOL SYNTHASE)                                                                                                                                                                                  |
| AT1G02820   | late embryogenesis abundant 3 family protein / LEA3 family protein                                                                                                                                         |
| AT1G05430   | similar to unnamed protein product [Vitis vinifera] (GB:CAO41766.1); contains InterPro domain Lipid-binding START (InterPro:IPR002913)              |
| AT3G28007   | nodulin MtN3 family protein                                                                                                                                                                              |
| AT5G51500   | UBC30 (UBIQUITIN-CONJUGATING ENZYME 30); ubiquitin-protein ligase                                                                                                                                          |
| AT4G4830    | methionine sulfoxide reductase domain-containing protein / SeIR domain-containing protein                                                                                                                |
| AT3G43610   | tubulin binding                                                                                                                                                                                           |
| AT4G35350   | XCP1 (XYLEM CYSTEINE PEPTIDASE 1); cysteine-type peptidase                                                                                                                                             |
| Gene Symbol | Gene Name | Description |
|-------------|-----------|-------------|
| AT3G50430   | similar to Os07g0120700 [Oryza sativa (japonica cultivar-group)] (GB:NP_001058781.1); similar to unnamed protein product [Vitis vinifera] (GB:CAO17953.1) |
| AT5G45620   | 26S proteasome regulatory subunit, putative (RPN9) |
| AT1G79560   | EMB1047/FTSH12 (EMBRYO DEFECTIVE 1047); ATP-dependent peptidase/ATPase/metallopeptidase |
| AT3G05180   | GDP-motif lipase/hydrolase family protein |
| AT5G41990; AT5G41992 | cinnamoyl-CoA reductase family |
| AT2G33590   | EMB1047/FTSH12 (EMBRYO DEFECTIVE 1047); ATP dependent peptidase/ATPase/metallopeptidase |
| AT5G41990; AT5G41992 | DNA-binding protein-related |
| AT4G37450   | AGP18 (Arabinogalactan protein 18) |
| AT5G41990; AT5G41992 | DNA-binding protein-related |
| AT4G37450   | DIR1 (DEFECTIVE IN INDUCED RESISTANCE 1); lipid binding |
| AT4G24340; AT4G24350 | short-chain dehydrogenase/reductase (SDR) family protein |
| AT4G24340; AT4G24350 | phosphorlyase family protein |
| AT4G24340; AT4G24350 | phosphorlyase family protein |
| AT4G24340; AT4G24350 | phosphorlyase family protein |
| AT4G24340; AT4G24350 | phosphorlyase family protein |

**Note:** The table lists genes with significant expression changes, along with their descriptions and related protein families. Each gene is associated with specific biological functions and annotations, providing insights into developmental and regulatory processes in *A. thaliana*.
| Gene Accession | Fold Change | q-value | P-value | Description |
|----------------|-------------|---------|---------|-------------|
| AT4G10040      | 1.625223468 | 0.6153  | 0.0008  | CYTC-2 (CYTOCHROME C-2); electron carrier |
| ATG27460       | 1.624827362 | 0.6154  | 0.0008  | NPRG1 (NO POLLEN GERMINATION RELATED 1); calmodulin binding |
| AT5G08535      | 1.619039909 | 0.6176  | 0.0007  | D111/G-patch domain-containing protein |
| AT3G62150      | 1.612771636 | 0.6200  | 0.0010  | PGP21 (P-GLYCOPEPTIDE 21); ATPase, coupled to transmembrane movement of substances |
| AT4G14420      | 1.612253124 | 0.6202  | 0.0010  | lesion inducing protein-related |
| AT1G47813      | 1.61108426  | 0.6207  | 0.0015  | [AT1G47813, similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G47820.1); similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G47820.2);[AT1G47820, similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G47813.1); similar to unnamed protein product [Vitis vinifera] (GB:CAO40107.1)] |
| AT2G47510      | 1.605651895 | 0.6228  | 0.0010  | pectinacetylesterase, putative |
| AT2G47510      | 1.60307791  | 0.6235  | 0.0015  | [AT2G47510, FUM1 (FUMARASE 1); fumarate hydratase];[AT5G50950, fumarate hydratase, putative / fumarase, putative] |
| AT1G74950      | 1.60012801  | 0.6249  | 0.0010  | RHM1/ROL1 (RHAMNOSE BIOSYNTHESIS1); UDP-L-rhamnose synthase/ UDP-glucose 4,6-dehydratase/ catalytic |
| AT3G16780      | 1.603720632 | 0.6238  | 0.0015  | [AT3G16780, similar to unknown [Brassica rapa] (GB:ABL97948.1)] |
| AT3G58070      | 1.598465473 | 0.6256  | 0.0010  | GIS (GLABROUS INFLORESCENCE STEMS); nucleic acid binding / transcription factor/ zinc ion binding |
| AT4G14615      | 1.596169194 | 0.6256  | 0.0010  | similar to unknown protein [Arabidopsis thaliana] |
| AT5G23210      | 1.588562351 | 0.6295  | 0.0015  | calcium-binding EF hand family protein |
| AT4G14615      | 1.579030475 | 0.6333  | 0.0015  | [AT4G14615, similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G52825.1); similar to unknown protein product [Vitis vinifera] (GB:CAO71274.1)] |
| ATG23870       | 1.578531965 | 0.6335  | 0.0015  | SCPL34; serine carboxypeptidase |
| AT5G20685      | 1.57790927  | 0.6337  | 0.0015  | metal ion binding |
| ATG27000       | 1.576719395 | 0.6350  | 0.0015  | bZIP family transcription factor |
| AT5G27000      | 1.573687938 | 0.6354  | 0.0015  | similar to unknown [Brassica rapa] (GB:ABL97948.1) |
| AT5G23870      | 1.563159940 | 0.6395  | 0.0015  | pectinacetylesterase family protein |
| AT5G26930      | 1.561889887 | 0.6402  | 0.0015  | pectinacetylerase, putative |
| AT1G25990      | 1.55788915  | 0.6421  | 0.0015  | [AT1G53900, GTP binding / translation initiation factor];[AT1G53880, GTP binding / translation initiation factor] |
| AT5G26930      | 1.55557284  | 0.6425  | 0.0015  | HMGB2 (HIGH MOBILITY GROUP B 2); transcription factor |
| AT3G45770      | 1.552433439 | 0.6445  | 0.0015  | oxidoreductase, zinc-binding dehydrogenase family protein |
| p-value  | Log2FC | Adj p-value | Log2FC | Symbol   | Gene Name and Description                                                                 |
|---------|--------|-------------|--------|----------|------------------------------------------------------------------------------------------|
| 1.547987616 | 0.646  | 0.07165 | 0.00125 | AT4G37870 | PCK1/PEPCK (PHOSPHOENOLPYRUVATE CARBOXYKINASE 1); ATP binding / phosphoenolpyruvate carboxykinase (ATP) |
| 1.546072975 | 0.6468 | 0.06435 | 0.00125 | AT3G17780 | similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G48440.1); similar to unknown [Populus trichocarpa] (GB:ABK93075.1) |
| 1.537751807 | 0.6503 | 0.07335 | 0.00125 | AT5G19260 | similar to unknown protein [Arabidopsis thaliana] (TAIR:AT3G06020.1); similar to hypothetical protein [Vitis vinifera] (GB:CAN75990.1) |
| 1.53233221 | 0.6526 | 0.0765 | 0.00135 | AT5G04830 | similar to unknown [Populus trichocarpa x Populus deltoides] (GB:ABK96633.1); contains domain SSF54427 (SSF54427) |
| 1.529051988 | 0.654  | 0.07805 | 0.00145 | AT5G40890 | ATCLC-A (CHLORIDE CHANNEL A); anion channel/ voltage-gated chloride channel |
| 1.526484506 | 0.6551 | 0.0786 | 0.0014 | AT4G15700 | glutaredoxin family protein |
| 1.519295047 | 0.6582 | 0.08095 | 0.00145 | AT4G12420 | SKU5 (skewed 5); copper ion binding |
| 1.51779616 | 0.65885 | 0.07715 | 0.0014 | AT5G27620 | CYCH;1 (CYCLIN H;1); cyclin-dependent protein kinase/ protein binding / protein kinase |
| 1.516990291 | 0.6592 | 0.08055 | 0.00175 | AT3G23620 | brix domain-containing protein |
| 1.513775356 | 0.6606 | 0.08075 | 0.00155 | AT3G21250 | ATMRP6 (Arabidopsis thaliana multidrug resistance-associated protein 6) |
| 1.513202694 | 0.66085 | 0.07865 | 0.00145 | AT1G69523 | UbiE/COQ5 methyltransferase family protein |
| 1.511601542 | 0.66155 | 0.07745 | 0.0016 | AT4G34270 | TIP41-like family protein |
| 1.508523156 | 0.6629 | 0.08205 | 0.00155 | AT1G06210 | VHS domain-containing protein / GAT domain-containing protein |
| 1.491646778 | 0.6704 | 0.08915 | 0.0019 | AT1G52760 | esterase/lipase/thioesterase family protein |
| 1.491313101 | 0.67055 | 0.0934 | 0.00185 | AT5G43750 | similar to unnamed protein product [Vitis vinifera] (GB:CAO71280.1) |
Down-regulated genes where there was a significant expression level difference between parents. Lowest fold change is reported only. (For all genes in this case it was Sha).

| inverse FC | FC   | pfp | P.value | Array Element | Locus Identifier | Annotation |
|------------|------|-----|---------|---------------|-----------------|------------|
| 42.91845494 | 0.0233 | 0   | 0       | 262374_s_at   | AT1G72910; AT1G72930 | [AT1G72910, disease resistance protein (TIR-NBS class), putative]; [AT1G72930, TIR (TOLL/INTERLEUKIN-1 RECEPTOR-LIKE); transmembrane receptor] |
| 19.26782274 | 0.0519 | 0   | 0       | 255450_at     | AT4G02850      | phenazine biosynthesis PhzC/PhzF family protein |
| 15.82278481 | 0.0632 | 0   | 0       | 255257_at     | AT4G05050      | [AT1G66690, S-adenosyl-L-methionine:carboxyl methyltransferase family protein]; [AT1G66700, PXMT1; S-adenosylmethionine-dependent methyltransferase] |
| 12.88659794 | 0.0776 | 0   | 0       | 256376_s_at   | AT1G66690; AT1G66700 | ubiquinol-cytochrome C reductase complex 7.8 kDa protein, putative / mitochondrial hinge protein, putative |
| 12.65822785 | 0.079  | 0   | 0       | 253707_at     | AT4G29200      | beta-galactosidase |
| 11.61440186 | 0.0861 | 0   | 0       | 252659_at     | AT3G44430      | unknown protein |
| 11.0864745  | 0.0902 | 0   | 0       | 262206_at     | AT2G01090      | ubiquinol-cytochrome C reductase complex 7.8 kDa protein, putative / mitochondrial hinge protein, putative |
| 9.04159132  | 0.1106 | 0   | 0       | 261309_at     | AT1G48598; AT1G48600 | [AT1G48598, CPuORF31 (Conserved peptide upstream open reading frame 31)]; [AT1G48600, phosphoethanolamine N-methyltransferase 2, putative (NMT2)] |
| 7.604562738 | 0.1315 | 0   | 0       | 255065_s_at   | AT4G08870; AT4G08900 | [AT4G08870, arginase, putative]; [AT4G08900, arginase] |
| 7.490636704 | 0.1335 | 0   | 0       | 245729_at     | AT1G73490      | RNA recognition motif (RRM)-containing protein similar to unknown protein [Arabidopsis thaliana] |
| 7.132667618 | 0.1402 | 0   | 0       | 263023_at     | AT1G23960      | (TAIR:AT1G23970.1); contains InterPro domain Protein of unknown function DUF626, Arabidopsis thaliana (InterPro:IPR006462) |
| 6.447453256 | 0.1551 | 0   | 0       | 258027_at     | AT3G19155      | binding |
| 5.737234653 | 0.1743 | 0   | 0       | 246417_at     | AT5G16990      | NADP-dependent oxidoreductase, putative similar to unknown protein [Arabidopsis thaliana] (TAIR:AT5G45520.1); similar to unnamed protein product [Vitis vinifera] (GB:CAO43141.1); similar to Os01g0799000 [Oryza sativa (japonica cultivar-group)] (GB:NP_001044526.1); contains domain SSF52047 (SSF52047); contains domain G3DSA:3.80.10.10 (G3DSA:3.80.10.10) |
| 5.730659026 | 0.1745 | 0   | 0       | 248944_at     | AT5G45500      | transposable element gene |
| 5.420054201 | 0.1845 | 0   | 0       | 245032_at     | AT2G26630      | transposable element gene |
| Average FC | Average p | Average F | Array Element | Locus Identifier | Annotation |
|------------|-----------|-----------|---------------|-----------------|------------|
| 6.14745    | 0.00025   | 0         | 262832_s_at   | AT1G14870;AT1G14880 | [AT1G14870, Identical to Uncharacterized protein At1g14870 [Arabidopsis thaliana] (GB:Q9LQU4); similar to unknown protein [Arabidopsis thaliana] (TAIR:AT5G35525.1); similar to unnamed protein product [Vitis vinifera (GB:CAO42338.1); contains InterPro domain Aspartic acid and asparagine hydroxylation site (InterPro:IPR000152); contains InterPro domain Protein of unknown function Cys-rich (InterPro:IPR006461)];[AT1G14880, similar protein [Arabidopsis thaliana] (TAIR:AT1G14870.1); similar to unnamed protein product [Vitis vinifera] (GB:CAO42338.1); similar to unnamed protein product [Arabidopsis thaliana] (TAIR:AT1G14870.1); similar to unnamed protein product [Vitis vinifera] (GB:CAO42338.1); similar to unnamed protein product [Arabidopsis thaliana] (TAIR:AT1G14870.1); similar to unnamed protein product [Arabidopsis thaliana] (TAIR:AT1G14870.1)]; contains InterPro domain Protein of unknown function Cys-rich (InterPro:IPR006461)] |
| 3.5968     | 0.0011    | 0         | 266070_at     | AT2G18660        | EXLB3 (EXPSIN-LIKE B3 PRECURSOR) |
| 3.4526     | 0.00065   | 0         | 254255_at     | AT4G23220        | protein kinase family protein |
| 3.15565    | 0.0033    | 0         | 248062_at     | AT5G55450        | protease inhibitor/seed storage/lipid transfer protein (LTP) family protein |
| 3.06945    | 0.03985   | 0.00105   | 250445_at     | AT5G10760        | aspartyl protease family protein |
| 2.96       | 0.00365   | 0         | 249096_at     | AT5G43910        | pfkB-type carbohydrate kinase family protein |
| 2.9273     | 0.0038    | 0         | 245329_at     | AT4G14365        | zinc finger (C3HC4-type RING finger) family protein / ankyrin repeat far |
| 2.92105    | 0.00235   | 0         | 265228_s_at   | ATMG01190;ATMG01190 | mitochondrial, putative |
| 2.83215    | 0.00175   | 0         | 248810_at     | AT5G47280        | ADR1-L3 (ADR1-LIKE 3); ATP binding / nucleoside-triphosphatase/ nucleotide binding / protein binding |
| 2.82005    | 0.00255   | 0         | 245422_at     | AT4G17470        | palmitoyl protein thioesterase family protein |
| 2.7618     | 0.00265   | 0         | 247604_at     | AT5G60950        | COBL5 (COBRA-LIKE PROTEIN 5 PRECURSOR) |
| 2.7256     | 0.0068    | 0.00005   | 254521_at     | AT5G44820        | similar to unknown protein [Arabidopsis thaliana] (TAIR:AT4G19970.1); similar to unknown protein product [Vitis vinifera] (GB:CAO46707.1); contains domain PTHR10483:SF6 (PTHR10483:SF6); contains domain PTHR10483 (P1 binding domain) |
| 2.7202     | 0.0038    | 0         | 259561_at     | AT1G21250        | WAK1 (CELL WALL-ASSOCIATED KINASE); kinase |
| 2.60305    | 0.03915   | 0.00105   | 251673_at     | AT3G57240        | BG3 (BETA-1,3-GLUCANASE 3); hydrolase, hydrolyzing O-glycosyl co |
| 2.46155    | 0.0093    | 0.0001    | 253423_at     | AT4G32280        | IAA29 (indoleacetic acid-induced protein 29); transcription factor |
| 2.4512     | 0.0039    | 0         | 250277_at     | AT5G12940        | leucine-rich repeat family protein |
| 2.45005    | 0.0044    | 0         | 245265_at     | AT4G14400        | ACD6 (ACCELERATED CELL DEATH 6); protein binding |
| 2.4438     | 0.00675   | 0.00005   | 259272_at     | AT3G01290        | band 7 family protein |
| 2.4005     | 0.0159    | 0.00025   | 248327_at     | AT5G32750        | heavy-metal-associated domain-containing protein |
| 2.35755    | 0.0059    | 0.00005   | 258856_at     | AT3G02040        | SRG3 (SENESCENCE-RELATED GENE 3); glycerophosphodiester phosphodiesterase |
| 2.3575     | 0.00805   | 0.00005   | 265441_at     | AT2G20870        | cell wall protein precursor, putative |
| 2.3187      | 0.01085  | 0.0001    | 249813_at     | AT5G23940        | EMB3009 (EMBRYO DEFECTIVE 3009); transferase |
| Gene Name | Log2 Fold Change | P-Value | q-Value | Gene Symbol | Function |
|-----------|-----------------|---------|---------|-------------|----------|
| AT2G29730 | 2.2809          | 0.0097  | 0.001   | 266643_s_at | [AT2G29730, UDP-glucoronsyl/UDP-glucosyl transferase family protein]; [AT2G29710, UDP-glucoronsyl/UDP-glucosyl transferase family protein] |
| AT5G52810 | 2.3185          | 0.0056  | 0.0005  | 248330_at   | ornithine cyclodeaminase/mu-crystallin family protein |
| AT3G50660 | 2.25645         | 0.0141  | 0.0005  | 245052_at   | pectinesterase family protein |
| AT1G03850 | 2.1688          | 0.0344  | 0.0005  | 252414_at   | glutaredoxin family protein |
| AT1G49050 | 2.09265         | 0.0150  | 0.0005  | 252360_at   | homoserine dehydrogenase, putative |
| AT5G44820 | 2.03105         | 0.01475 | 0.0015  | 261969_at   | ABC1 family protein |
| AT3G47420 | 2.0117          | 0.01225 | 0.0015  | 266613_at   | gibberelin-regulated family protein |
| AT5G48550 | 2.0243          | 0.0213  | 0.0035  | 252976_s_at | InterPro domain Phospholipase-like, arabidopsis (InterPro:IPR007942) |
| AT4G23260 | 2.01655         | 0.0285  | 0.0005  | 256834_at   | 2-oxoglutarate-dependent dioxygenase, putative |
| AT1G12710 | 2.0279          | 0.03185 | 0.00035 | 261193_at   | similar to unknown protein [Arabidopsis thaliana] (TAIR:ATG32928.1) |
| AT3G43550 | 2.0105          | 0.01625 | 0.0015  | 254247_at   | 2-oxoglutarate-dependent dioxygenase, putative |
| AT3G43855 | 2.0243          | 0.0213  | 0.0035  | 252976_s_at | similar to unknown protein [Arabidopsis thaliana] (TAIR:ATG20950.1) |
| AT4G49050 | 2.0058          | 0.01795 | 0.0015  | 252652_at   | ADP2-A12 (PHLOEM PROTEIN 2-A12); carbohydrate binding |
| AT3G44720 | 2.00845         | 0.0188  | 0.002   | 26077a_at   | AHP5 (HISTIDINE-CONTAINING PHOSPHOTRANSFER FACTOR 5) |
| AT2G45490 | 2.00117         | 0.01755 | 0.0015  | 264838_at   | ADP2-A12 (PHLOEM PROTEIN 2-A12); carbohydrate binding |
| AT2G14900 | 2.00295         | 0.01225 | 0.0015  | 268613_at   | gibberelin-regulated family protein |
| AT1G12710 | 2.0058          | 0.01795 | 0.0015  | 252652_at   | similar to unknown protein [Arabidopsis thaliana] (TAIR:ATG20950.1) |
| AT1G03400 | 2.00117         | 0.01755 | 0.0015  | 264838_at   | ADP2-A12 (PHLOEM PROTEIN 2-A12); carbohydrate binding |
| AT2G45490 | 2.00845         | 0.0188  | 0.002   | 26077a_at   | AHP5 (HISTIDINE-CONTAINING PHOSPHOTRANSFER FACTOR 5) |
| AT1G12710 | 2.0058          | 0.01795 | 0.0015  | 252652_at   | similar to unknown protein [Arabidopsis thaliana] (TAIR:ATG20950.1) |
| AT1G03400 | 2.00117         | 0.01755 | 0.0015  | 264838_at   | ADP2-A12 (PHLOEM PROTEIN 2-A12); carbohydrate binding |
| AT2G45490 | 2.00845         | 0.0188  | 0.002   | 26077a_at   | AHP5 (HISTIDINE-CONTAINING PHOSPHOTRANSFER FACTOR 5) |
| AT1G12710 | 2.0058          | 0.01795 | 0.0015  | 252652_at   | similar to unknown protein [Arabidopsis thaliana] (TAIR:ATG20950.1) |
| AT1G03400 | 2.00117         | 0.01755 | 0.0015  | 264838_at   | ADP2-A12 (PHLOEM PROTEIN 2-A12); carbohydrate binding |
| AT2G45490 | 2.00845         | 0.0188  | 0.002   | 26077a_at   | AHP5 (HISTIDINE-CONTAINING PHOSPHOTRANSFER FACTOR 5) |
| Gene ID   | log2 Fold Change | P Value | q Value | Description                                                                 |
|----------|-----------------|---------|---------|-----------------------------------------------------------------------------|
| AT3G17790| 1.9806          | 0.0154  | 0.00015 | ATACP5 (acid phosphatase 5); acid phosphatase/ protein serine/threonine phosphatase |
| AT3G47800| 1.9697          | 0.01975 | 0.00015 | aldose 1-epimerase family protein                                           |
| AT2G24160| 1.96025         | 0.0487  | 0.00095 | AT5G58980; ATACP5 (acid phosphatase 5); acid phosphatase/ protein serine/threonine phosphatase |
| AT3G47800| 1.9569          | 0.02865 | 0.0003  | aldose 1-epimerase family protein                                           |
| AT2G04450| 1.9555          | 0.01855 | 0.00025 | AT5G58980; ATACP5 (acid phosphatase 5); acid phosphatase/ protein serine/threonine phosphatase |
| AT5G03200| 1.95245         | 0.02675 | 0.00025 | aldose 1-epimerase family protein                                           |
| AT1G07380| 1.94855         | 0.021   | 0.00025 | AT5G58980; ATACP5 (acid phosphatase 5); acid phosphatase/ protein serine/threonine phosphatase |
| AT1G70830| 1.9434          | 0.02625 | 0.00025 | aldose 1-epimerase family protein                                           |
| AT5G63080| 1.9386          | 0.0183  | 0.00025 | AT5G58980; ATACP5 (acid phosphatase 5); acid phosphatase/ protein serine/threonine phosphatase |
| AT3G51430| 1.93185         | 0.021   | 0.00025 | AT5G58980; ATACP5 (acid phosphatase 5); acid phosphatase/ protein serine/threonine phosphatase |
| AT5G19240| 1.92305         | 0.021   | 0.00025 | AT5G58980; ATACP5 (acid phosphatase 5); acid phosphatase/ protein serine/threonine phosphatase |
| AT1G17430| 1.91915         | 0.02275 | 0.00025 | AT5G58980; ATACP5 (acid phosphatase 5); acid phosphatase/ protein serine/threonine phosphatase |
| AT4G23180| 1.91375         | 0.02385 | 0.00025 | AT5G58980; ATACP5 (acid phosphatase 5); acid phosphatase/ protein serine/threonine phosphatase |
| AT1G07380| 1.9086          | 0.02775 | 0.0003  | AT5G58980; ATACP5 (acid phosphatase 5); acid phosphatase/ protein serine/threonine phosphatase |
| AT3G47800| 1.9055          | 0.03055 | 0.00035 | AT5G58980; ATACP5 (acid phosphatase 5); acid phosphatase/ protein serine/threonine phosphatase |
| AT2G35020| 1.8843          | 0.0435  | 0.0006  | AT5G58980; ATACP5 (acid phosphatase 5); acid phosphatase/ protein serine/threonine phosphatase |
| AT2G04450| 1.88265         | 0.026   | 0.0003  | AT5G58980; ATACP5 (acid phosphatase 5); acid phosphatase/ protein serine/threonine phosphatase |
| AT3G48080| 1.8822          | 0.0313  | 0.00035 | AT5G58980; ATACP5 (acid phosphatase 5); acid phosphatase/ protein serine/threonine phosphatase |

Note: The table above lists the genes and their corresponding fold changes, along with their significance levels and descriptions. The genes are related to various biological processes and functions, including phosphatase activity, serine/threonine phosphorylation, and disease resistance.
### Protein List

| Expression | FDR | Adjusted FDR | Gene ID | Description |
|-----------|-----|-------------|---------|-------------|
| 1.8812    | 0.05095 | 0.0008 | 251035\_at | AT5G02220 | similar to unknown [Picea sitchensis] (GB:ABK23883.1); similar to hypc protein [Vitis vinifera] (GB:CAN70860.1) |
| 1.88      | 0.0242  | 0.0004 | 245399\_at | AT4G17340 | DELTA-TIP2/TIP2:2 (tonoplast intrinsic protein 2;2); water channel |
| 1.87965   | 0.02975  | 0.0005 | 262888\_at | AT1G14970 | RDR1 (RNA-DEPENDENT RNA POLYMERASE 1); RNA-directed RNA polymerase/ nucleic acid binding |
| 1.8731    | 0.0327  | 0.0004 | 251668\_at | AT3G57010 | strictosidine synthase family protein |
| 1.8705    | 0.0466  | 0.00075 | 251705\_at | AT3G56400 | WRKY70 (WRKY DNA-binding protein 70); transcription factor |
| 1.8668    | 0.0297  | 0.00055 | 253238\_at | AT4G34480 | glycosyl hydrolase family 17 protein |
| 1.8657    | 0.02635 | 0.0004 | 251422\_at | AT3G60540 | sec61beta family protein |
| 1.86385   | 0.0264  | 0.00035 | 253377\_at | AT4G33300 | ADR1-L1 (ADR1-LIKE 1); ATP binding / protein binding |
| 1.8551    | 0.03015 | 0.00045 | 254283\_s\_at | AT4G22870;AT4G22880 | [AT4G22870, leucoanthocyanidin dioxygenase, putative / anthocyanidin synthase, putative][AT4G22880, LDOX (TANNIN DEFICIENT SEED 4)] |
| 1.85185   | 0.0253  | 0.00025 | 267096\_at | AT2G38180 | GDSL-motif lipase/hydrolase family protein |
| 1.85065   | 0.0309  | 0.00035 | 262910\_at | AT1G59710 | similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G27100.1); unknown [Populus trichocarpa] (GB:ABK94560.1); contains InterPro do |
| 1.84555   | 0.02735 | 0.0003 | 263953\_at | AT2G36050 | ATOPF15/OFP15 (Arabidopsis thaliana ovate family protein 15) |
| 1.83975   | 0.0263  | 0.0003 | 246071\_at | AT5G20150 | SPX (SYG1/Pho81/XPR1) domain-containing protein |
| 1.83575   | 0.0303  | 0.00055 | 250937\_at | AT5G03230 | similar to unknown protein [Arabidopsis thaliana] (TAIR:AT5G60680.1); unnamed protein product [Vitis vinifera] (GB:CAO21845.1); contains Int domain |
| 1.83555   | 0.03125 | 0.00035 | 258173\_at | AT3G21630 | CERK1 (CHITIN ELICITOR RECEPTOR KINASE 1); kinase/ receptor protein/ transmembrane receptor protein kinase |
| 1.82585   | 0.03365 | 0.0004 | 253401\_at | AT4G32870 | similar to unknown protein [Arabidopsis thaliana] (TAIR:AT2G25770.2); unknown protein [Arabidopsis thaliana] (TAIR:AT2G25770.1); similar to [Populus trichocarpa x Populus deltoides] (GB:ABK96434.1); contains SSF55961 (SSF55961) |
| 1.82375   | 0.0281  | 0.0005 | 250661\_at | AT5G07030 | pepsin A |
| 1.81815   | 0.03315 | 0.00035 | 249904\_at | AT5G22700 | F-box family protein |
| 1.81525   | 0.04665 | 0.0012 | 261240\_at | AT1G32940 | ATSBT3.5; subtilase |
| 1.8127    | 0.04035 | 0.0005 | 253722\_at | AT4G29190 | zinc finger (CCCH-type) family protein |
| 1.80975   | 0.02725 | 0.00035 | 245602\_at | AT4G14270 | Protein containing PAM2 motif which mediates interaction with the PAB of polyadenyl binding proteins. |
| 1.8091    | 0.02905 | 0.00035 | 248248\_at | AT5G53120 | SPDS3 (SPERMIDINE SYNTHASE 3) |
| 1.80425   | 0.0295  | 0.0004 | 258786\_at | AT3G11820 | SYP121 (syntaxin 121); SNAP receptor |
| 1.8041    | 0.0536  | 0.00085 | 255294\_at | AT4G04750 | carbohydrate transmembrane transporter/ sugar:hydrogen ion symporter |
| 1.80165   | 0.0488  | 0.0008 | 267246\_at | AT2G30250 | WRKY25 (WRKY DNA-binding protein 25); transcription factor |
| 1.8007    | 0.0287  | 0.0004 | 267425\_at | AT2G35060 | KUP11 (K+ uptake permease 11); potassium ion transmembrane transp |
| Log2 fold change | P-value | E-value | Gene ID | Description | Log2 fold change | P-value | E-value | Gene ID | Description |
|-----------------|---------|---------|---------|------------|-----------------|---------|---------|---------|------------|
| 1.7999          | 0.0433  | 0.0006  | 250891_at | AT5G04530 | beta-ketoacyl-CoA synthase family protein |
| 1.7995          | 0.051   | 0.00075 | 263914_at | AT2G36400 | AIGRF3 (GROWTH-REGULATING FACTOR 3) |
| 1.79825         | 0.0298  | 0.0005  | 247632_at | AT5G60460 | sec61beta family protein |
| 1.79375         | 0.04    | 0.0005  | 258351_at | AT3G17700 | CNBT1 (CYCLIC NUCLEOTIDE-BINDING TRANSPORTER 1); calmod / cyclic nucleotide binding / ion channel |
| 1.78845         | 0.0444  | 0.0006  | 251010_at | AT5G02550 | unknown protein |
| 1.78635         | 0.05095 | 0.0008  | 259009_at | AT3G09260 | PYK10 (phosphate starvation-response 3.1); hydrolase, hydrolyzing O-compounds |
| 1.7859          | 0.0359  | 0.00055 | 250083_at | AT5G17220 | ERD7 (EARLY-RESPONSIVE TO DEHYDRATION 7) |
| 1.78395         | 0.03345 | 0.00045 | 264787_at | AT2G17840 | GDSL-motif lipase, putative |
| 1.7803          | 0.0395  | 0.0005  | 245074_at | AT2G23200 | protein kinase family protein |
| 1.7794          | 0.03065 | 0.00045 | 245302_at | AT4G17695 | KAN3 (KANADI 3); DNA binding / transcription factor |
| 1.7741          | 0.0312  | 0.0004  | 267595_at | AT2G32990 | hydrolase, hydrolyzing O-glycosyl compounds |
| 1.7735          | 0.03605 | 0.0006  | 264223_s_at | AT3G16030 | CES101 (CALLUS EXPRESSION OF RBCS 101); carbohydrate binding similar to unnamed protein product [Vitis vinifera] (GB:CAO44135.1) |
| 1.7703          | 0.04765 | 0.00065 | 246905_at | AT5G25570 | LYM2 (LYSM DOMAIN GPI-ANCHORED PROTEIN 2 PRECURSOR) |
| 1.7681          | 0.0434  | 0.0007  | 263582_at | AT2G17120 | peptidase M16 family protein / insulinate family protein |
| 1.7665          | 0.0462  | 0.0007  | 251641_at | AT3G57470 | MLO2 (MILDEW RESISTANCE LOCUS O 2); calmodulin binding |
| 1.7632          | 0.03435 | 0.00045 | 262022_at | AT2G02400 | cinnamoyl-CoA reductase family |
| 1.76045         | 0.04095 | 0.00055 | 262455_at | AT1G11310 | ATERF-2/ATERF2/ERF2 (ETHYLENE RESPONSE FACTOR 2); DNA transcription activator/ transcription factor |
| 1.7564          | 0.04825 | 0.0007  | 262736_at | AT1G28570 | GDSL-motif lipase, putative |
| 1.7461          | 0.0403  | 0.0006  | 248794_at | AT5G47220 | disease resistance-responsive family protein / dirigent family protein |
| 1.74405         | 0.04065 | 0.0006  | 254909_at | AT4G11210 | similar to unknown protein [Arabidopsis thaliana] (TAIR:AT2G31670.1); unknown [Populus trichocarpa] (GB:AKB53075.1); contains InterPro domain Dimeric alpha-beta barrel (InterPro:IPR011007) |
| 1.7439          | 0.0563  | 0.0009  | 265142_at | AT1G51360 | unknown [Arabidopsis thaliana] (TAIR:AT2G31670.1); unknown [Populus trichocarpa] (GB:AKB53075.1); contains InterPro domain Dimeric alpha-beta barrel (InterPro:IPR011007) |
| 1.74325         | 0.04375 | 0.00085 | 244951_s_at | AT2G07723;A* | Unigene6245 (oligopeptide transporter 4); oligopeptide transporter |
| 1.73875         | 0.04805 | 0.0007  | 247284_at | AT5G64410 | ATOPT4 (oligopeptide transporter 4); oligopeptide transporter |
| 1.73605         | 0.0416  | 0.0008  | 264854_at | AT2G17450 | RHA3A (RING-H2 finger A3A); protein binding / zinc ion binding |
| 1.73425         | 0.06015 | 0.001   | 248568_at | AT5G49760 | leucine-rich repeat family protein / protein kinase family protein |
### Table 1: Genes with Heterozygous Disadvantage in A. thaliana

| Gene ID          | Gene Symbol | Description                                      |
|------------------|-------------|--------------------------------------------------|
| AT5G23410        |             | [AT5G23410, similar to FKF1 (FLAVIN-BINDING KELCH DOMAIN F BOX PROTEIN)], ubiquitin-protein ligase [Arabidopsis thaliana] (TAIR:AT1G6 similar to unnamed protein product [Vitis vinifera] (GB:CAO42365.1); cc domain PTHR23244 (PTHR23244); contains domain PTHR23244:SF9 (PTHR23244:SF9);[AT1G68050, FKF1 (FLAVIN-BINDING KELCH DOMAIN F BOX PROTEIN)]; ubiquitin-protein ligase];[AT5G42730, pseudogene similar to ACT domain-containing protein, similar to F-box family protein] |
| AT5G28890        |             | PLL4 (POLTERGEIST LIKE 4); protein serine/threonine phosphatase |
| AT5G25560        |             | zinc finger (C3HC4-type RING finger) family protein |
| AT5G60900        |             | RKL1 (RECEPTOR-LIKE PROTEIN KINASE 1); carbohydrate binding / kinase |
| AT5G42730        |             | thymidylate kinase family protein |
| AT1G68050        |             | similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G54200.1); similar to hypothetical protein [Vitis vinifera] (GB:CAN96469.1) |
| AT4G29030        |             | glycine-rich protein |
| AT5G05090        |             | myb family transcription factor |
| AT3G19010        |             | oxidoreductase, 2OG-Fe(II) oxygenase family protein |
| AT1G11800        |             | endonuclease/exonuclease/phosphatase family protein |
| AT4G18970        |             | GDSL-motif lipase/hydrolase family protein |
| AT4G46664        |             | unknown protein |
| AT3G46030        |             | HTB11; DNA binding |
| AT1G73805        |             | calmodulin binding |
| AT2G17020        |             | F-box family protein (FBL10) |
| AT1G62660        |             | beta-fructosidase (BFRUCT3) / beta-fructofuranosidase / invertase, vac |
| AT5G45670        |             | GDSL-motif lipase/hydrolase family protein |
| AT5G64370        |             | BETA-UP (BETA-UREIDOPROPIONASE); beta-ureidopropionase |
| AT3G23050        |             | IAA7 (AUXIN RESISTANT 2); transcription factor |
| AT3G01472        |             | similar to unknown protein [Arabidopsis thaliana] (TAIR:AT5G6020.1); unnamed protein product [Vitis vinifera] (GB:CAO45187.1); contains Int domain SFT2-like (InterPro:IPR011691) |
| AT4G26550        |             | protein binding |
| AT4G12070        |             | APP (ARABIDOPSIS POLY(ADP-RIBOSE) POLYMERASE); NAD+ AC ribosyltransferase |
| Gene Symbol | Log2 Fold | P-value | Description |
|-------------|-----------|---------|-------------|
| AT5G13740   | 1.65475   | 0.001   | ZIF1 (ZINC INDUCED FACILITATOR 1); carbohydrate transmembrane transporter/ sugar:hydrogen ion symporter |
| AT5G23570   | 1.6524    | 0.001   | SGS3 (SUPPRESSOR OF GENE SILENCING 3) |
| AT3G26170; AT3G26180 | 1.65155 | 0.001 | [AT3G26170, CYP71B19 (cytochrome P450, family 71, subfamily B, pc 19); oxygen binding];[AT3G26180, CYP71B20 (cytochrome P450, famili subfamily B, polypeptide 20); oxygen binding] |
| AT1G13980   | 1.64475   | 0.001   | GN (GNOM) |
| AT4G20320   | 1.64245   | 0.001   | CTP synthase |
| AT1G09415   | 1.64185   | 0.001   | NIMIN-3 (NIM1-INTERACTING 3) |
| AT1G68840   | 1.63775   | 0.001   | RAV2 (REGULATOR OF THE ATPASE OF THE VACUOLAR MEMBR binding / transcription factor |
| AT3G06160   | 1.6367    | 0.001   | transcriptional factor B3 family protein |
| AT4G37260   | 1.6186    | 0.001   | AMYB73/MYB73 (myb domain protein 73); DNA binding / transcription |
| AT2G16700   | 1.61055   | 0.001   | ADF5 (ACTIN DEPOLYMERIZING FACTOR 5); actin binding |
| AT2G36360   | 1.59975   | 0.001   | kelch repeat-containing protein |
| AT1G31710   | 1.5969    | 0.001   | copper amine oxidase, putative |
| AT5G03870   | 1.59165   | 0.001   | glutaredoxin family protein |
| AT4G09500   | 1.5865    | 0.001   | glycosyltransferase family protein |
| AT3G62650   | 1.58395   | 0.001   | HAC1 (P300/CBP ACETYLTRANSFERASE-RELATED PROTEIN 2 G H3/H4 histone acetyltransferase/ transcription cofactor |
| AT4G09480   | 1.5827    | 0.001   | transposable element gene |
| AT4G11530   | 1.57945   | 0.001   | glycine-rich cell wall protein-related |
| AT2G18250   | 1.5571    | 0.001   | ATCOAD (4-PHOSPHOPANTETHEINE ADENYLTRANSFERASE); nucleotidyltransferase/ pantetheine-phosphate adenyltransferase |
| AT3G55260   | 1.5562    | 0.001   | ATHEX2/HEXO1 (BETA-HEXOSAMINIDASE 1); beta-N-acetylhexosaminidase/ hydrolase, hydrolyzing O-glycosyl compounds |
| AT1G78210   | 1.5495    | 0.001   | hydrolase, alpha/beta fold family protein |
| AT4G11530   | 1.54495   | 0.001   | protein kinase family protein |
Upregulated genes where there was a significant expression level difference between parents. Lowest fold change is reported only. (For all genes in this case it was Sha).

| FC       | pfp | P.value | Array Element | Locus Identifier | Annotation                                                                 |
|----------|-----|---------|---------------|------------------|-----------------------------------------------------------------------------|
| 6.2546   | 0   | 0       | 259385_at     | AT1G13470         | similar to unknown protein [Arabidopsis thaliana] (TAIR:AT1G13520.1); unnamed protein product [Vitis vinifera] (GB:CAO42040.1); contains Int domain Protein of unknown function DUF1262 (InterPro:IPR010683) |
| 2.7634   | 0   | 0       | 255895_at     | AT1G18020;A      | [AT1G18020, 12-oxophytodienoate reductase, putative];[AT1G17990, 1 oxophytodienoate reductase, putative] |
| 3.2523   | 0   | 0       | 261942_at     | AT1G22590         | AGL87; transcription factor                                                |
| 4.0176   | 0   | 0       | 262082_s_at   | AT1G56140;AT1G56120 | [AT1G56140, leucine-rich repeat family protein / protein kinase family protein];[AT1G56130, leucine-rich repeat family protein / protein kinase protein];[AT1G56120, leucine-rich repeat family protein / protein kinase protein] |
| 3.8582   | 0   | 0       | 256601_s_at   | AT3G28290;A      | [AT3G28290, AT14A];[AT3G28300, AT14A]                                        |
| 3.8139   | 0   | 0       | 252345_at     | AT3G48640         | similar to unknown protein [Arabidopsis thaliana] (TAIR:AT5G66670.2); unknown protein [Arabidopsis thaliana] (TAIR:AT5G66670.1) |
| 2.1228   | 0.0026 | 0   | 245456_at    | AT4G16950         | RPP5 (RECOGNITION OF PERONOSPORA PARASITICA 5)                               |
| 3.1606   | 0   | 0       | 248169_at     | AT5G54610         | ANK (ANKYRIN); protein binding                                              |
**Supplementary Table 6.** Similarity of OAK and related alleles.

|            | At5g59670 Col-0 | At5g59670a Bla-1 | OAK Bla-1 | At5g59670a Sha | OAK Sha |
|------------|-----------------|------------------|-----------|----------------|---------|
| At5g59670 Col-0       | –               | 84               | 89        | 89             | 87      |
| At5g59670a Bla-1     | 75              | –                | 83        | 87             | 83      |
| OAK Bla-1        | 81              | 71               | –         | 91             | 94      |
| At5g59670a Sha     | 83              | 78               | 85        | –              | 95      |
| OAK Sha          | 79              | 72               | 91        | 93             | –       |

Nucleotide identity in percent is given on top, with amino acid identity given on bottom.
**Supplementary Table 7.** Survey of 87 *A. thaliana* accessions for OAK duplication.

| Accession ID | Accession name | OAK duplication |
|--------------|----------------|-----------------|
| CS76409      | Agu-1          | Yes             |
| CS76392      | Bak-2          | Yes             |
| CS76393      | Bak-7          | Yes             |
| CS22591      | Bor-4          | Yes\(^a\)       |
| CS76410      | Cdm-0          | Yes             |
| CS22614      | Cvi-0          | Yes             |
| CS22683      | Est-1          | Yes             |
| CS76423      | ICE102/Galdo-1 | Yes             |
| CS76363      | ICE112         | Yes             |
| CS76425      | ICE120/Valsi-1 | Yes             |
| CS76426      | ICE138/Leb-3   | Yes             |
| CS76379      | ICE150         | Yes             |
| CS76380      | ICE152         | Yes             |
| CS76381      | ICE153         | Yes             |
| CS76354      | ICE181         | Yes             |
| CS76355      | ICE212         | Yes             |
| CS76356      | ICE213         | Yes             |
| CS76349      | ICE226         | Yes             |
| CS76350      | ICE228         | Yes             |
| CS76419      | ICE29/Slavi-1  | Yes             |
| CS76372      | ICE33          | Yes             |
| CS76369      | ICE36          | Yes             |
| CS76348      | ICE50          | Yes             |
| CS76352      | ICE79          | Yes             |
| CS76362      | ICE91          | Yes\(^a\)       |
| CS76366      | ICE92          | Yes             |
| CS22651      | Kondara        | Yes\(^b\)       |
| CS22607      | Kz-9           | Yes             |
| CS76390      | Lag2-2         | Yes             |
| CS76413      | Leo-1          | Yes             |
| CS22686      | Ler            | Yes             |
| CS76388      | Lerik          | Yes             |
| CS76414      | Mer-6          | Yes             |
| CS76400      | Star-8         | Yes             |
| CS76403      | TüSB30-2       | Yes             |
| CS76391 | Vash | Yes |
|---------|------|-----|
| CS76408 | Wal-HäsB-4 | Yes |
| CS22679 | Bur-0 | No |
| CS22681 | Col-0 | No |
| CS76397 | Del-10 | No |
| CS76386 | Dog-4 | No |
| CS76411 | Don-0 | No |
| CS76399 | Ey 1.5-2 | No |
| CS76412 | Fei-0 | No |
| CS76404 | HKT2-4 | No |
| CS76373 | ICE1 | No |
| CS76367 | ICE104 | No |
| CS76365 | ICE106 | No |
| CS76364 | ICE107 | No |
| CS76361 | ICE111 | No |
| CS76424 | ICE119 | No |
| CS76385 | ICE127 | No |
| CS76384 | ICE130 | No |
| CS76383 | ICE134 | No |
| CS76353 | ICE163 | No |
| CS76357 | ICE169 | No |
| CS76358 | ICE173 | No |
| CS76370 | ICE21 | No |
| CS76351 | ICE216 | No |
| CS76347 | ICE49 | No |
| CS76377 | ICE60 | No |
| CS76378 | ICE61 | No |
| CS76420 | ICE63 | No |
| CS76371 | ICE7 | No |
| CS76421 | ICE70 | No |
| CS76375 | ICE71 | No |
| CS76374 | ICE72 | No |
| CS76376 | ICE73 | No |
| CS76422 | ICE75 | No |
| CS76368 | ICE93 | No |
| CS76359 | ICE97 | No |
| CS76360 | ICE98 | No |
| Accession | Name     | Present |
|-----------|----------|---------|
| CS76389   | Istisu-1 | No      |
| CS76395   | Kastel   | No      |
| CS76396   | Koch     | No      |
| CS76398   | Nemrut   | No      |
| CS76402   | Nie1.2   | No      |
| CS76415   | Ped-0    | No      |
| CS76416   | Pre-6    | No      |
| CS76417   | Qui-0    | No      |
| CS76406   | Rü3.1-27 | No      |
| CS22646   | Se-0     | No      |
| CS22647   | Ts-1     | No      |
| CS76401   | Tü-Sha-9 | No      |
| CS76407   | Tü-V-12  | No      |
| CS76405   | TüWa1-2  | No      |
| CS76418   | Vie-0    | No      |
| CS76387   | Xan-1    | No      |
| CS76394   | Yeg-1    | No      |

*These accessions also contain the At5g59670 Col-0 like promoter. Kondara has a similar incompatibility phenotype to Sha when crossed to Bla-1. It differs by two intergenic nucleotides in the 17.5 kb RLK cluster, so was excluded from population structure analyses.*
Supplementary Figures

Supplementary Figure 1. Bla-1/Sha incompatibility decreases seed set.

(a) Normal appearing Col-0 plants that are either non-transgenic or carry only a single OAK transgene. The phenotype of F₁ plants with both OAK transgenes is comparable to (b) Sha/Bla-1 F₁ plants. (c) Total seed set after three months shown as box and whisker plots. Boxes in box plot cover the first and third quartile, and the whiskers represent values that are not more than 1.5 times the interquartile range. A two-tailed, unequal variance t-test showed statistical equivalence of seed set between wild-type plants and those with a single OAK transgene, and highly significant reduction of seed set in plants carrying both transgenes.

Supplementary Figure 2. High humidity suppresses outgrowth formation.

Bla-1/Sha F₁ plants were grown for 3 and a half weeks under either high humidity (covered with a dome and surrounded by water), normal humidity (controlled 65% humidity), or under drought stress conditions (65% humidity but minimal watering). Two representative leaves per treatment are shown. Outgrowths are indicated by arrows.

Supplementary Figure 3. Effect of auxin and cytokinin concentration on callus formation.

Callus formation at 12 days for transverse sections of leaves and petioles of Bla-1, Bla-1/Sha F1 and Sha. Three representative tissue pieces are shown per accession and hormone concentration.

Supplementary Figure 4. Mapping interval for the Bla-1/Sha outgrowth causal gene.

(a) Positional cloning markers used according to the cognate genes and position in Mbp in reference accession Col-0. (b) The genes in reference accession Col-0 in the final mapping interval, with protein kinases marked in light grey and the RLKs highlighted in mid-grey.
Supplementary Figure 5. AmiRNA knockdown of OAK rescues the hybrid phenotype.
AmiRNAs designed against each RLK in the OAK cluster from Bla-1 (a) or Sha (b) were transformed into Bla-1/Sh a F1 plants and plants heterozygous at the RLK locus identified in the next generation. One representative plant per line is shown. Scale bar = 1 cm.

Supplementary Figure 6. Potential LRR and malectin-like domains in OAK.
(a) The consensus for plant-specific LRR domains is given below according to (Kobe, B. & Kajava, A.V. The leucine-rich repeat as a protein recognition motif. Curr. Opin. Struct. Biol. 11, 725-32; 2001), with residues conserved in over 50% of proteins shown in uppercase. Leucine resides from OAK at conserved positions are indicated in yellow, with other conserved residues highlighted in green. Less conserved residues or residues similar to those conserved are highlighted in light grey. (b) Predicted malectin-like domains (Schallus, T. et al. Malectin: a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein N-glycosylation. Mol. Biol. Cell 19, 3404-14; 2008) in OAKBl a-1 and OAKSha. Although the amino acid sequence identity is low (11-15%), the secondary structure is more highly conserved, and the probability scores are very high.

Supplementary Figure 7. Divergence of RLK orthologs and paralogs.
(a) Comparison of pairwise amino acid divergence between OAKBla-1 and OAKSha and between all RLKs in this cluster. (b) Comparison of pairwise amino acid divergence between OAK and At5g59670a alleles from Bla-1 and Sha.

Supplementary Figure 8. Compatibility between OAK-containing accessions.
Cytoscape (Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498-2504) representation of crosses performed between OAK-containing accessions (names indicated in circles). Node color on the periphery indicates the haplotype group of the second malectin domain. Cvi-0, Cdm-0, ICE50, ICE226 and ICE228 alleles.
switch between haplotype groups, and are shown in intermediate colours. Absence of color indicates that the haplotype group is not known. Compatible hybrid combinations are indicated by grey edges, and incompatible ones with outgrowths with black edges.

**Supplementary Figure 9.** Much of the OAK promoter is derived from a duplicated region of RLK coding sequence.

Top 15 hits from LALIGN (http://www.ch.embnet.org/software/LALIGN_form.html) are shown according to position in the Bla-1 OAK promoter, linked to a colour-matched box indicating position in the Col-0 RLK cluster.

**Supplementary Figure 10.** Alignment of the OAK proteins from Sha, Leo-1 and Bla-1.

Amino acid differences between the three OAK proteins are indicated in purple (where Sha differs from Leo-1 and Bla, which are both incompatible with Sha), in cyan (where Bla-1 differs from Sha and Leo-1) and in red (where Leo-1 differs from Sha and Bla-1). Alignment was performed with CLUSTALW (Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, et al. (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31: 3497-3500).

**Supplementary Figure 11.** Expression of the OAK extracellular domain in hybrid plants can reduce the severity of aberrant phenotypes.

The extracellular domains of OAK$_{Sha}$, OAK$_{Bla}$ or At5g59670$_{Col-0}$ under control of their native promoters or the 35S promoter were transformed into a segregating hybrid background and scored for the hybrid phenotype. Transformants were genotyped for allelic status at the endogenous OAK locus to identify heterozygous individuals. Plants with a mild phenotype where only a few outgrowths were observed on the petioles but that were otherwise phenotypically wild-type were combined with the "wild-type" category.
Supplementary Figure 12. Mis-expressed OAK couples to the salicylic acid signalling pathway.

(a) Pro$_{35S}$:nahG when introduced into $P_{At5g59670}$:OAK$_{Bla-1}$ $P_{At5g59670}$:OAK$_{Sha}$ rescues of the cell death phenotype. (b) Pro$_{35S}$:nahG when introduced into $P_{OAK}$:OAK$_{Bla-1}$ $P_{OAK}$:OAK$_{Sha}$ does not suppress the outgrowths, leaf twisting or loss of apical dominance.
Figure S1  Smith et al., 2011

A

normal

$\text{Pro}_{\text{OAK}}:\text{OAK}_{\text{Bla-1}}$

$\text{Pro}_{\text{OAK}}:\text{OAK}_{\text{Sha}}$

$F_1$ Sha/Bla-1

(3 plants)

B

C

$P << 0.001$

$\begin{array}{c}
\text{None} \\
\text{Pro}_{\text{OAK}}:\text{OAK}_{\text{Bla-1}} \\
\text{Pro}_{\text{OAK}}:\text{OAK}_{\text{Sha}} \\
\text{Pro}_{\text{OAK}}:\text{OAK}_{\text{Sha}} \\
\text{Pro}_{\text{OAK}}:\text{OAK}_{\text{Sha}}
\end{array}$

$n=22$

$n=20$

$n=23$

$n=23$

$\begin{array}{c}
\text{Total seeds/plant} \\
\text{Transgene}
\end{array}$
Supplementary Figure 2  

Smith et al., 2010
Supplementary Figure 3  

**Auxin (2,4-D)**

- **2.22 μM**
- **20 nM**
- **22 nM**

**Cytokinin (kinetin)**

- **0.2 nM**
- **2 nM**
- **20 nM**
- **200 nM**
### A

| Locus     | Protein                                                                 |
|-----------|-------------------------------------------------------------------------|
| At5g58780 | Sensitivity to red light reduced protein                               |
| At5g59460 | Myb family transcription factor                                         |
| At5g59560 | UDP-glucosyl transferase family protein                                 |
| At5g59590 | UDP-glucosyl transferase family protein                                 |
| At5g59600 | Pentatricopeptide (PPR) repeat-containing protein                       |
| At5g59610 | DNAJ heat shock N-terminal domain-containing protein                    |
| At5g59813 | Similar to unknown protein                                             |
| At5g59616 | Protein kinase-related                                                 |
| At5g59620 | CACTA-like transposase family                                          |
| At5g59630 | Pseudogene                                                             |
| At5g59640 | CACTA-like transposase family                                          |
| At5g59650 | LRR protein kinase                                                     |
| At5g59660 | LRR protein kinase                                                     |
| At5g59670 | LRR protein kinase                                                     |
| At5g59680 | LRR protein kinase                                                     |
| At5g59690 | Histone H4                                                             |
| At5g59700 | Putative protein kinase                                                |

### B

```
   11/384
   23.74
   11/384
   24.25 Mb
   23.96
   24.00
   24.05
   24.10
```

Supplementary Figure 4  Smith et al., 2010
A  LRR domains

Bla-1 OAK (amino acids 409-502)
PPRITSLNLSSSR LNGTIATAIQSLTQLETLDLSNNN LTGGVPEFLGK

Sha OAK (amino acids 409-502)
PPRITSLNLSSSR LNGTIATAIQSLTQLETLDLSNNN LTGGVPEFLGK

Where 1 = t/s 2 = g/-

B  Malectin-like domains

Bla-1

No 1

>2jwp_A Malectin, MGC80075; sugar binding, sugar binding protein; NMR (Xenopus laevis) PDB: 2j45 A*
Probval=99.79  E-value=5.1e-19  Score=164.79  Aligned_cols=143  Identities=14%  Similarity=0.123  Sum_probs=0.0

No 2

>2jwp_A Malectin, MGC80075; sugar binding, sugar binding protein; NMR (Xenopus laevis) PDB: 2j46 A*
Probval=99.61  E-value=2.7e-16  Score=146.27  Aligned_cols=152  Identities=11%  Similarity=0.043  Sum_probs=0.0

Supplementary Figure 6 (page 1)    Smith et al., 2010
Supplementary Figure 6 (page 2)    Smith et al., 2010
Figure S8  Smith et al., 2011
Leo_1  MESSFGLLLVLTVLTLTVTVQDOQDGQDGSLDCGLPPNETSLYKENRTGGLFSSDATTIQ 60
Bla_1  MESSFGLLLVLTVLTLTVTVQDOQDGQDGSLDCGLPPNETSLYKENRTGGLFSSDATTIQ 60
Sha  MESSFGLLLVLTVLTLTVTVQDOQDGQDGSLDCGLPPNETSLYKENRTGGLFSSDATTIQ 60

Leo_1  SGKTGVRQVQCQFSLKPYRTLRFPEGRVCYLSVFKERYLITASFLGYNYDGHNIA 120
Bla_1  SGKTGVRQVQCQFSLKPYRTLRFPEGRVCYLSVFKERYLITASFLGYNYDGHNIA 120
Sha  SGKTGVRQVQCQFSLKPYRTLRFPEGRVCYLSVFKERYLITASFLGYNYDGHNIA 120

Leo_1  PVFDLYGLPNMLANIDLEDVNGKWEEILHIPTSNSLQICLVKTGMATPLISSLELRPMRT 180
Bla_1  PVFDLYGLPNMLANIDLEDVNGKWEEILHIPTSNSLQICLVKTGMATPLISSLELRPMRT 180
Sha  PVFDLYGLPNMLANIDLEDVNGKWEEILHIPTSNSLQICLVKTGMATPLISSLELRPMRT 180

Leo_1  RSYTIESGSLKTFRRLYFNKSGSELRYSKDVYDRIWMPHFEDEWTQISTALRVNKNNDYE 240
Bla_1  RSYTIESGSLKTFRRLYFNKSGSELRYSKDVYDRIWMPHFEDEWTQISTALRVNKNNDYE 240
Sha  RSYTIESGSLKTFRRLYFNKSGSELRYSKDVYDRIWMPHFEDEWTQISTALRVNKNNDYE 240

Leo_1  LETDESDVVAMKNISASYGLSRINWQGDPCFPEQLRWDALDCSNTHISTPPRITSLNLSS 419
Bla_1  LETDESDVVAMKNISASYGLSRINWQGDPCFPEQLRWDALDCSNTHISTPPRITSLNLSS 419
Sha  LETDESDVVAMKNISASYGLSRINWQGDPCFPEQLRWDALDCSNTHISTPPRITSLNLSS 419

Leo_1  KLYVPSTEVPEKLSLTTFQSPSPTSCNGWECYFQLIRTKRSTLPPLLNEVYTVIQFPQ 359
Bla_1  KLYVPSTEVPEKLSLTTFQSPSPTSCNGWECYFQLIRTKRSTLPPLLNEVYTVIQFPQ 359
Sha  KLYVPSTEVPEKLSLTTFQSPSPTSCNGWECYFQLIRTKRSTLPPLLNEVYTVIQFPQ 359

Leo_1  QLSTQGYKQFKAEVDLLLRAHHTNLVSLVGYCHEGNHLALIYEFLPNGDLKQHLSGKGGK 659
Bla_1  QLSTQGYKQFKAEVDLLLRAHHTNLVSLVGYCHEGNHLALIYEFLPNGDLKQHLSGKGGK 659
Sha  QLSTQGYKQFKAEVDLLLRAHHTNLVSLVGYCHEGNHLALIYEFLPNGDLKQHLSGKGGK 659

Leo_1  SIINWSTRLIAEAALGELYLHICTPPMVHRDVKTANILLDENFKAKLDALPFDLSRFQ 719
Bla_1  SIINWSTRLIAEAALGELYLHICTPPMVHRDVKTANILLDENFKAKLDALPFDLSRFQ 719
Sha  SIINWSTRLIAEAALGELYLHICTPPMVHRDVKTANILLDENFKAKLDALPFDLSRFQ 719

Leo_1  VKGEFYDSTLVAAPYLDPEYPYRGLKRSKREVSKYSFGIVLLEMTNQPVIQTSNANI 779
Bla_1  VKGEFYDSTLVAAPYLDPEYPYRGLKRSKREVSKYSFGIVLLEMTNQPVIQTSNANI 779
Sha  VKGEFYDSTLVAAPYLDPEYPYRGLKRSKREVSKYSFGIVLLEMTNQPVIQTSNANI 779

Leo_1  TQRVGEIANGNILEMPDKLCSDKYDIKASRADLHAMCDSSSKRPSVEIQVQLK 839
Bla_1  TQRVGEIANGNILEMPDKLCSDKYDIKASRADLHAMCDSSSKRPSVEIQVQLK 839
Sha  TQRVGEIANGNILEMPDKLCSDKYDIKASRADLHAMCDSSSKRPSVEIQVQLK 839

Leo_1  ECILCENSRINNGLESEMVVDLSSETLMAR- 873
Bla_1  ECILCENSRINNGLESEMVVDLSSETLMAR- 873
Sha  ECILCENSRINNGLESEMVVDLSSETLMAR- 873

Figure S10  Smith et al., 2011
Normal or near-normal phenotype

Pro35S:EDBla-1
Pro35S:EDCol-0
ProAt5g59670:EDCol-0

Percentage of T1 plants

Control
ProOAK:EDBla-1
ProOAK:EDSha
Pro35S:EDSha
ProAt5g59670:EDSha
ProAt5g59670:EDCol-0
Pro35S:EDCol-0

n = 156 50 44 72 54 38 65

Standard hybrid phenotype

Normal or near-normal phenotype
