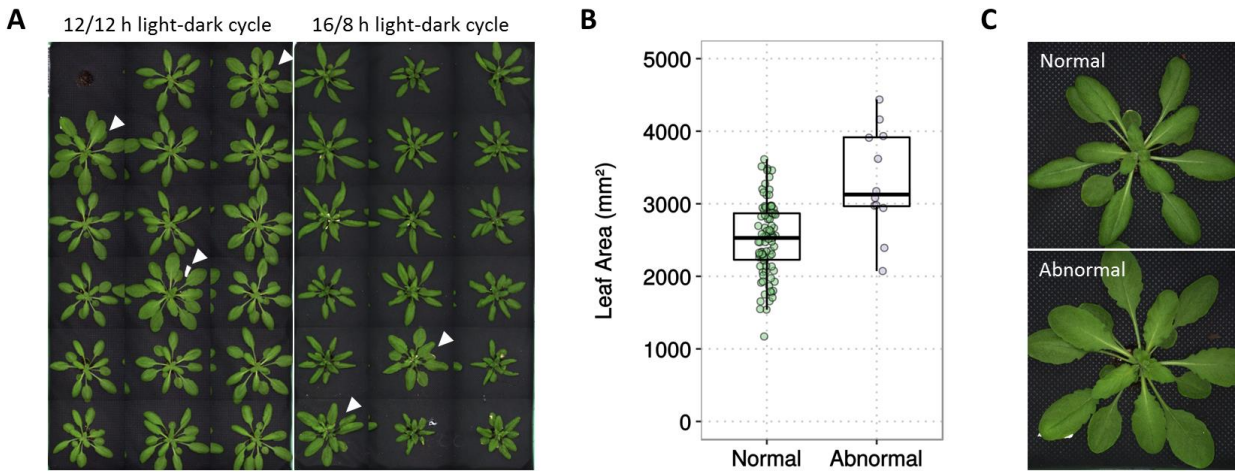
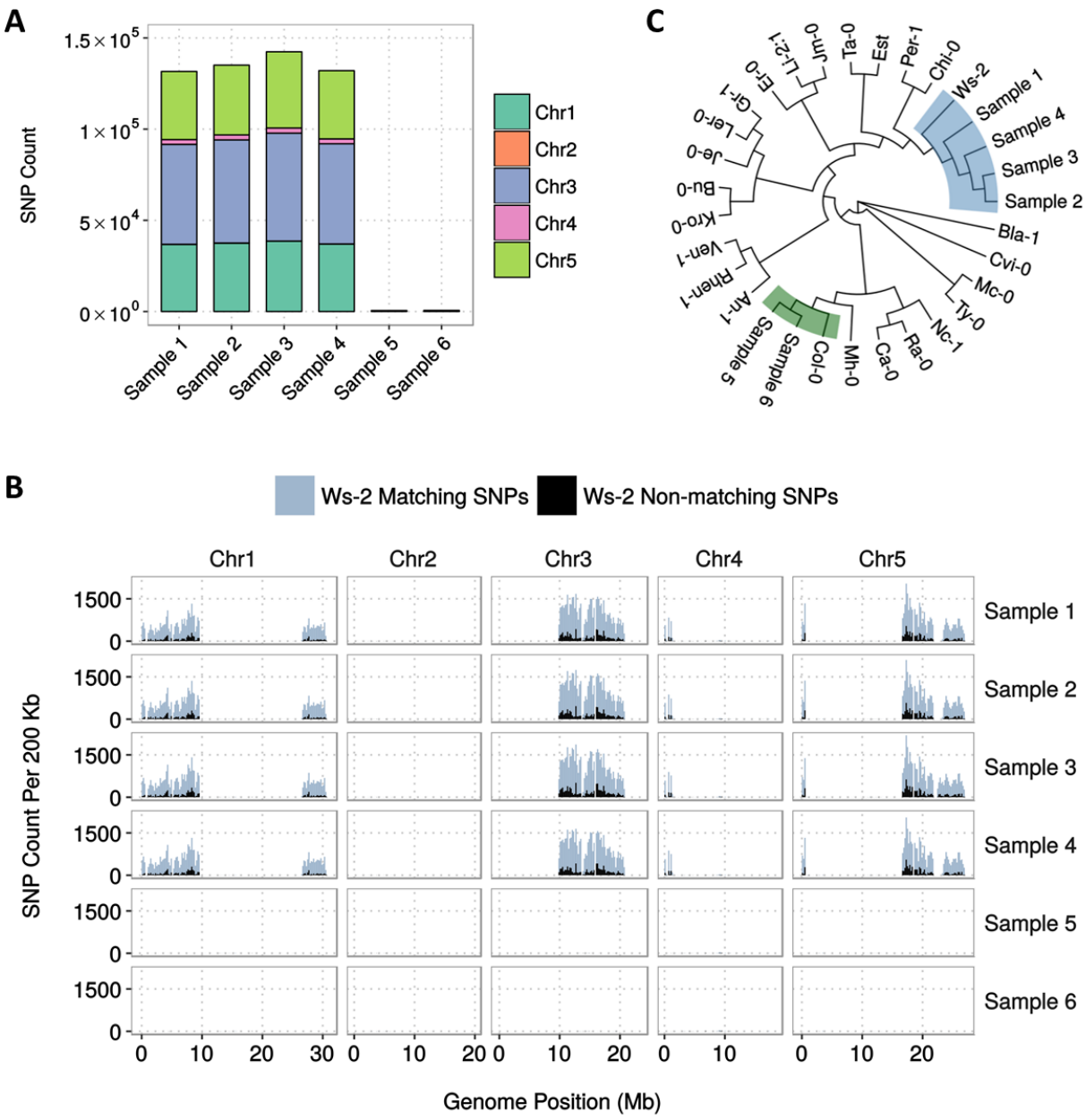


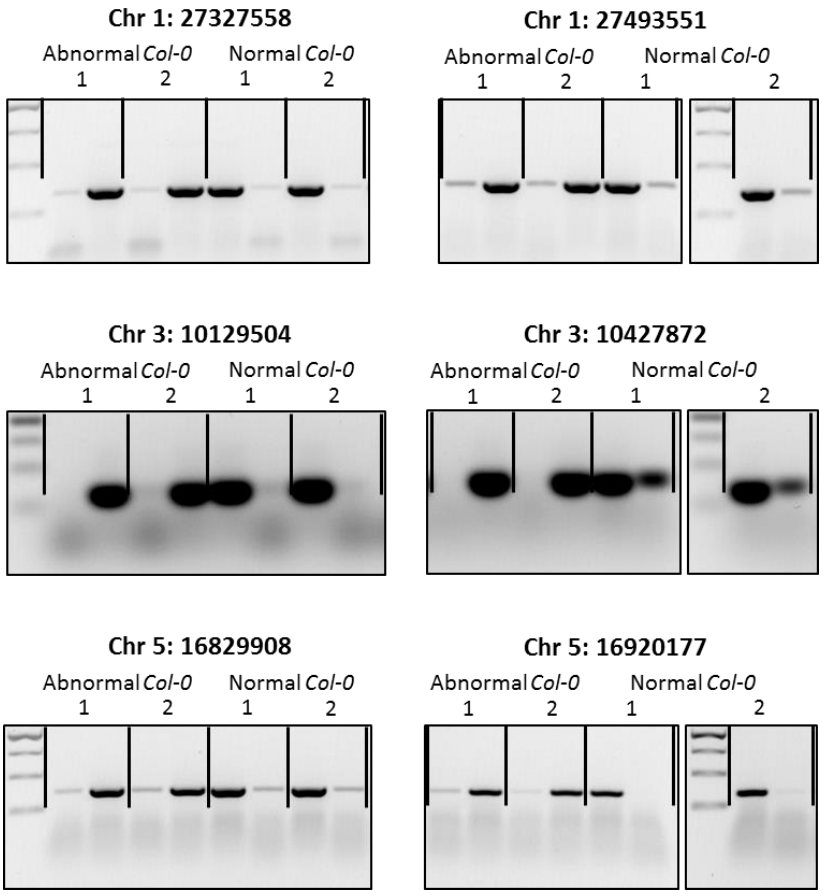
Supplemental Figure 1. Phenotype of abnormal plants in LEHLE SEEDS Company *Col-0* lot #214-509.

(A) Randomly planted seeds grown under short day and long day conditions. White arrows indicate phenotypically abnormal individuals.
(B) Leaf area of normal versus abnormal plants at 28 days old (n = 106).
(C) Representative image under short day conditions showing the flatter, more serrated leaves in the abnormal individual when compared to a normal *Col-0* plant.



Supplemental Figure 2. DNA sequencing of abnormal plants reveals large *Ws-2* introgressions.

(A) SNP counts in abnormal (Samples 1-4) and normal (Samples 5-6) *Col-0* plants.
(B) Chromosomal distribution of SNPs within each sample.
(C) Cladogram of sequenced samples plus 25 ecotypes, using SNP data from genomic regions showing high SNP density in the abnormal samples.



Supplemental Figure 3. Allele-specific PCR assays confirm presence of *Ws-2* introgression in abnormal plants.

Two abnormal and two normal sample results are shown for the indicated SNP genomic positions on Chromosomes 1, 3, and 5. For each PCR assay, the *Ws-2* specific primer result is in the right lane and the *Col-0* specific primer result is in the left lane. Primer sequences are listed in Table 1.

Supplemental Table 1. SNAP PCR primer sequences used for genotyping shown in Supplemental Figure 3.

Position	Primer name	Primer sequence
Chr1: 27327558	AT1G72560-Col0-F	TGTGACTGCCATCTTCAGTTCCCGT
	AT1G72560-Ws2-F	AAAATGTGACTGCCATCTTCAGTTCCACA
	AT1G72560-Rev	CTTTCCTTGATGATAGGGGTTTGCATCAC
Chr1: 27493551	AT1G73100-Col0-F	TGGTTCTTTCATATGTGAATATGCTGGTGAAGTTA
	AT1G73100-Ws2-F	TGGTTCTTTCATATGTGAATATGCTGGTGAAGTAG
	AT1G73100-Rev	GTCAATTCTGCCATAGGAGGGATATGACG
Chr3: 10129504	AT3G27360-Col0-F	CCTCTAATCCTCCTCGCAAGTTGAATATCCATA
	AT3G27360-Ws2-F	CTAATCCTCCTCGCAAGTTGAATATCCCTC
	AT3G27360-Rev	TGGCGACGAAAGCAGCAAGGA
Chr3: 10427872	AT3G28030-Col0-F	TCAGAAAGGGATAACCAAACAATTTGGTACAGA
	AT3G28030-Ws2-F	GTCAGAAAGGGATAACCAAACAATTTGGTACATT
	AT3G28030-Rev	CCCTTGATTCAGTGACAAGACTCTGGATTTT
Chr5: 16829908	AT5G42100-Col0-F	CTTAACAGTAACGACCTTCTAGTCCTAAATGTCAAC
	AT5G42100-Ws2-F	CTTAACAGTAACGACCTTCTAGTCCTAAATGTGTAA
	AT5G42100-Rev	TCACCATCTTCGTCTTTGCTCTCTTTAACG
Chr5: 16920177	AT5G42320-Col0-F	CAAATATGGTTACCATTTCTTCGTTCACAAAGAACT
	AT5G42320-Ws2-F	AAATATGGTTACCATTTCTTCGTTCACAAAGATGA
	AT5G42320-Rev	TTTTGACAGATGATTCCTTTTCTTCCAGACTTTT

SUPPLEMENTAL METHODS

Plant Growth Conditions and Phenotyping

Arabidopsis thaliana seeds were sown directly onto soil and stratified at 4°C for 2 days, then grown at 22°C under 12/12 hour light-dark cycles (short day) or 16/8 hour light-dark cycles (long day) as indicated. For leaf area, short day -grown plants were imaged and measured using a LemnaTec Scanalyzer HTS system.

Whole Genome DNA Sequencing and Analysis

DNA for abnormal and typical samples was extracted using the DNeasy Plant Mini Kit (Qiagen) according to manufacturer protocol. Sequencing was performed by the University of Minnesota Genomics Center using a HiSeq 2500 to generate 100 bp paired-end reads. Samples were prepared using TruSeq Nano DNA libraries with an average insert size of approximately 350 bp. All six libraries were pooled and sequenced in a single lane. Each library had an average Q-score above 35, and yielded over 16 million read pairs per sample. De-multiplexed sequencing data is deposited at SRA accession SRP071704.

Trim Galore (Babraham Bioinformatics) was used to remove remaining Illumina adapters and bases from the 5’ and 3’ ends with a PHRED score below 20. Reads were

aligned against the *Col-0* reference sequence version TAIR10 using GenomeMapper, with a maximum edit distance of 10% and a maximum gap of 7%. SNP calling was then performed using SHORE with the developer-included homozygous scoring matrix. Sample SNPs were further filtered to those within the nuclear genome with a quality score ≥ 35, depth ≥ 10, mapping-quality-zero ≤ 0.05*depth, alternative allele frequency ≥ 0.95, and unambiguous reference and alternative allele identities (A/T/G/C only). Filtering and calculation of SNP frequencies and genomic distributions were performed in R.

Variant files for *Arabidopsis* accessions sequenced by the Salk Institute, the Max Planck Institute for Developmental Biology (MPI), the George Mendel Institute of Molecular Plant Biology (GMI), and the Wellcome Trust Centre for Human Genetics (WTCHG) were obtained from the 1001 Genomes Project (http://1001genomes.org). Sample SNPs were considered to be matching those of another accession's if they shared the same position and alternative allele identity. To include the relatively common *C24* ecotype as a potential genetic donor, a variant file updated to reflect the TAIR10 reference was generated from publicly available sequencing data for *C24* (Schneeberger et al., 2011), using the same GenomeMapper and SHORE parameters as described above.

Phylogenetic analysis was based on introgression coordinates in the abnormal samples, and included regions Chr1: 2-8 Mb, Chr3: 10-20 Mb, and Chr5: 18-21 Mb. Twenty-five ecotypes, including *Col-0* and *Ws-2*, along with the abnormal samples were analyzed with SNPhylo, using a maximum-likelihood based method. For SNP pruning, parameters were set at minor allele frequency ≥ 0.05 and a linkage disequilibrium threshold of 0.7, resulting in 30101 SNP sites kept for analysis. The cladogram was then visualized using the R/Bioconductor package *ggtree* (http://bioconductor.org/packages/ggtree).

Allele-Specific PCR Assay

SNAP PCR primers were designed for the introgressed regions on Chromosomes 1, 3 and 5, and introgression-free region on Chromosome 2 as recommended by the SNAP protocol/SNAP program (Drenkard et al., 2000) for *Col-0* and *Ws-2* specific SNPs. SNP sites within genes were preferentially chosen (indicated within primer name), and the genomic positions given are based on the *Col-0* reference sequence version TAIR10. PCR cycling was done as 5 min at 94°C, (30 sec at 94°C, 1 min at 62°C) 35 times, and finally 10 min at 72°C.

SUPPLEMENTAL REFERENCES

Drenkard, E., Richter, B.G., Rozen, S., Stutius, L.M., Angell, N.A., Mindrinos, M., Cho, R.J., Oefner, P.J., Davis, R.W., and Ausubel, F.M. (2000). A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in Arabidopsis. Plant Physiol. **124:** 1483-92.

Schneeberger, K., Ossowski, S., Ott, F., Klein, J.D., Wang, X., Lanz, C., Smith, L.M., Cao, J., Fitz, J., Warthmann, N., Henz, S.R., Huson, D.H., and Weigel, D. (2011). Reference-guided assembly of four diverse Arabidopsis thaliana genomes. Proc. Natl. Acad. Sci. USA **108:** 10249-54.