

A single haplotype hyposensitive to light and requiring strong vernalization dominates *Arabidopsis thaliana* populations in Patagonia, Argentina

Luciana Kasulin^{1§}, Beth A. Rowan^{2§}, Rolando J.C. León^{1}, Verena J. Schuenemann^{3,4}, Detlef Weigel²
and Javier F. Botto^{1*}**

¹Universidad de Buenos Aires, Facultad de Agronomía, IFEVA-CONICET, Av. San Martín 4453, C1417DSE, Ciudad de Buenos Aires, Argentina. ²Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany. ³Institute for Archaeological Sciences, University of Tübingen, 72070 Tübingen, Germany. ⁴Senckenberg Center for Human Evolution and Palaeoecology, University of Tübingen, Tübingen, Germany. [§]Both authors contributed equally. *Correspondence should be addressed to JFB (botto@agro.uba.ar). ** Deceased on November 11, 2015

Supporting Information

Table of Contents

- 1. Figure S1.** Single nucleotide polymorphisms (SNPs) in Pat.
- 2. Figure S2.** Distribution of shared and segregating SNPs among four Pat accessions across sequence annotation classes.
- 3. Figure S3.** Library characteristics and sequence comparison for the Pat herbarium specimen.
- 4. Figure S4.** Additional principal components analyses for Pat compared with 1135 worldwide accessions.
- 5. Figure S5.** Physiological responses for Pat plants collected in the four different sites.
- 6. Figure S6.** FRc and Bc light responses for Pat and Col-0 during de-etiolation.
- 7. Figure S7.** Hypocotyl length, leaf angle, petiole length, lamina width, lamina length, and flowering time for Col-0 and Pat plants cultivated in WL and simulated shade (EOD-FR).

- 8. Figure S8.** Photoperiodic flowering response without vernalization for Col-0 and Pat plants cultivated in long-day (LD) and short-day (SD).
- 9. Figure S9.** Flowering response at different cultivation temperatures.
- 10. Figure S10.** Diagrams of SNPs at *BBX4* and *PHYB* genes.
- 11. Figure S11.** Hypocotyl length for accessions with the Pat allele conferring the V979I change in *PHYB*.
- 12. Figure S12.** Flowering time (leaf number) for genetic groups (determined by ADMIXTURE) of worldwide *A. thaliana* accessions.
- 13. Table S1.** Primers used in this study.
- 14. Table S2.** Candidate genes used in this study.
- 15. Table S3.** Geographical data for collection sites.
- 16. Table S4.** Vegetation cover and type for collection sites.
- 17. Table S5.** Texture, ionic and general soil composition for collection sites.
- 18. Table S6.** Clusters of Pat private SNPs on chromosomes 4 and 5. * indicates non-synonymous changes leading to a premature stop codon.
- 19. Table S7.** Candidate genes with polymorphisms in the coding region and low SNP frequency (<0.10) among 1135 accessions from the 1001 Genomes database.

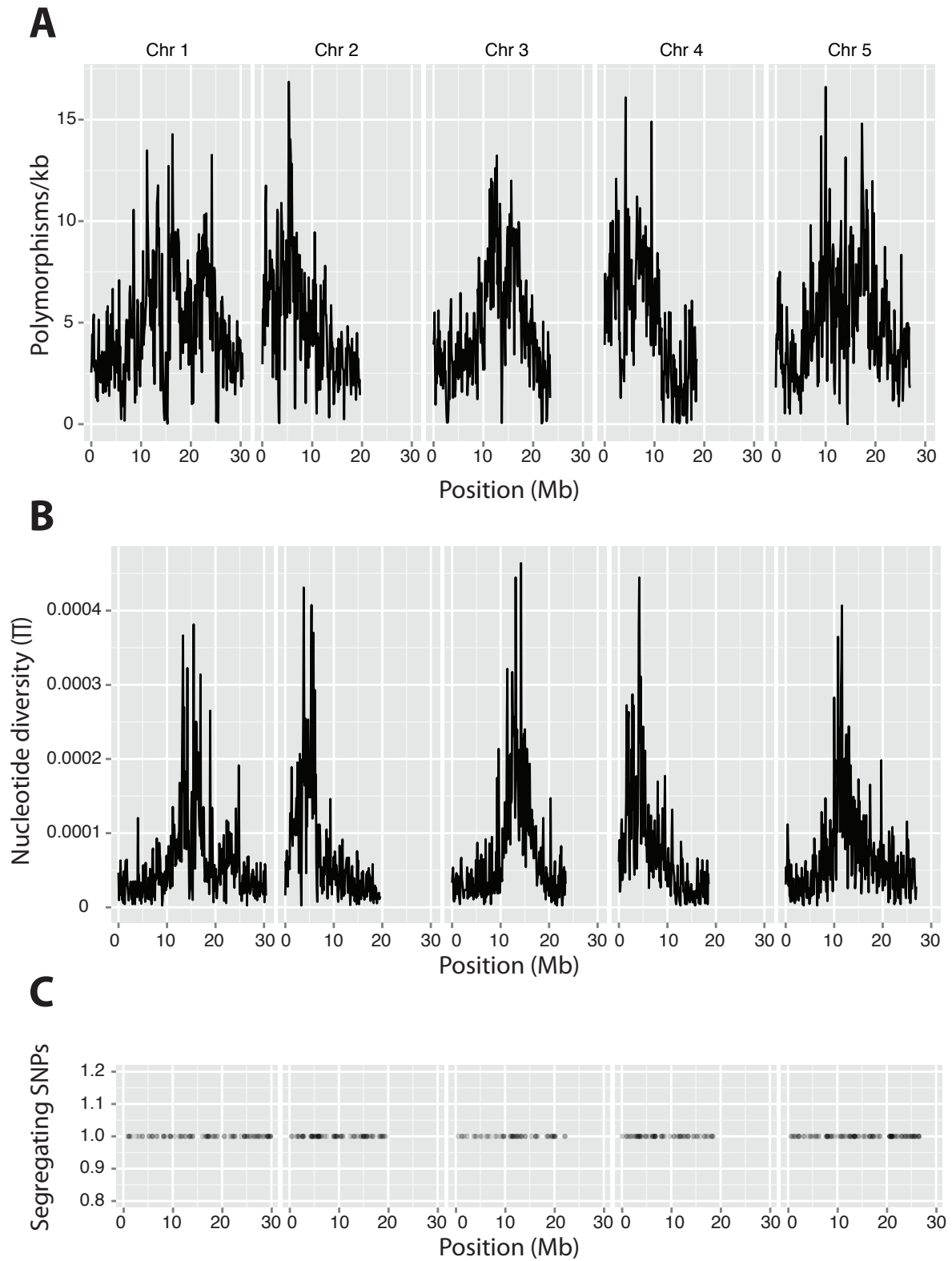


Figure S1. Single nucleotide polymorphisms (SNPs) in Pat. A. SNP density in 100-kb windows throughout the genome. **B.** Nucleotide diversity in 100-kb windows throughout the genome. **C.** Distribution of 406 SNPs segregating among the four Pat accessions throughout the genome plotted as points with an alpha of 0.2 to make over-plotted points more visible. The chromosome labels at the top of panel **A** apply to all panels.

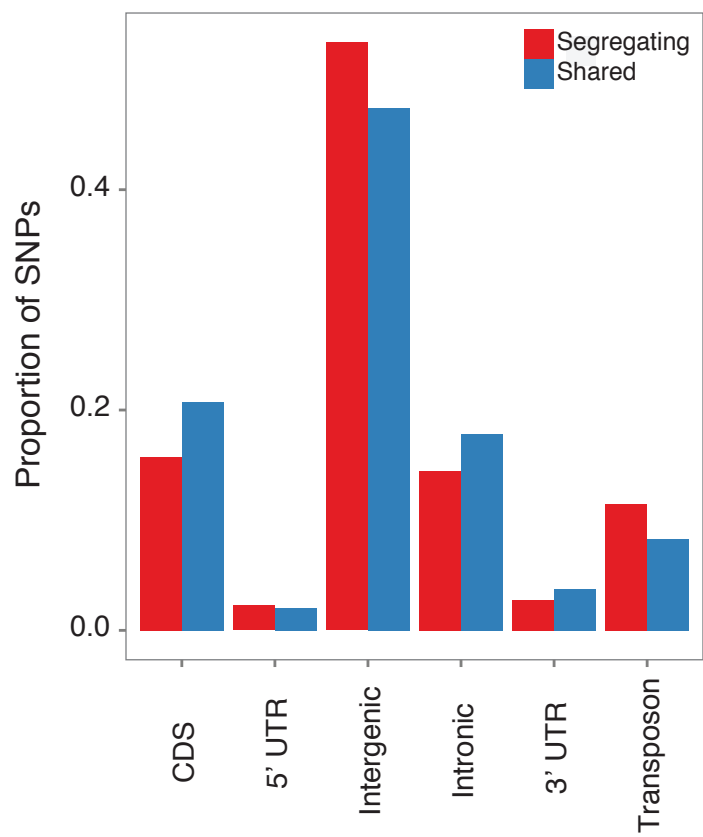


Figure S2. Distribution of shared and segregating SNPs among four Pat accessions across sequence annotation classes.

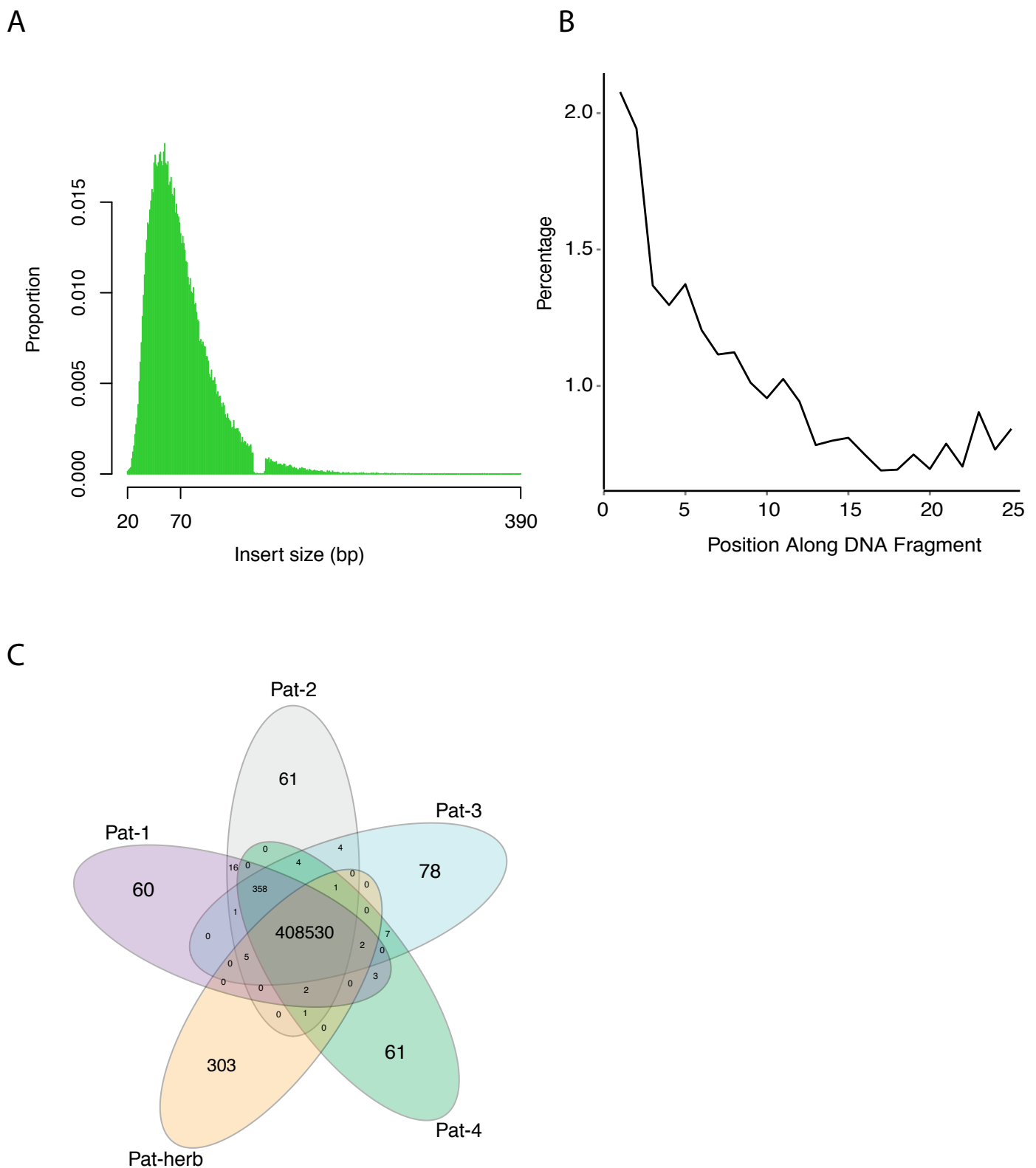


Figure S3. Library characteristics and sequence comparison for the Pat herbarium specimen. A. Insert size distribution. **B.** 5' C to T mutation rates by read position. **C.** Venn Diagram showing overlap of SNPs among all four 2010 Pat accessions and the 1967 herbarium specimen.

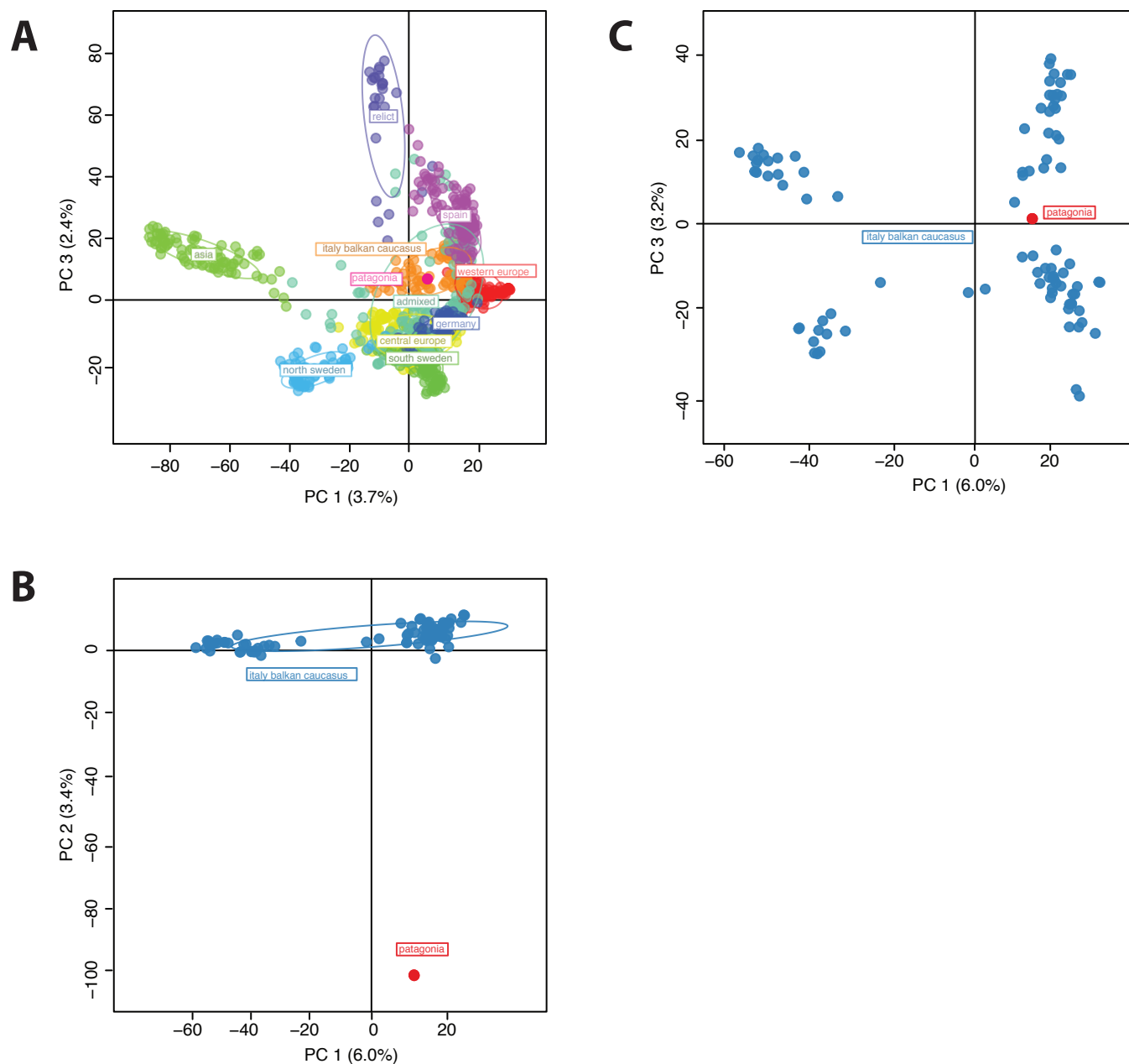


Figure S4. Additional principal components analyses for Pat compared with 1135 worldwide accessions. A. Component 1 and Component 3 for all accessions. **B.** Component 1 and Component 2 for Pat and the Italy-Balkan-Caucasus ADMIXTURE cluster only. **C.** Component 1 and Component 3 for Pat and the Italy-Balkan-Caucasus ADMIXTURE cluster only.

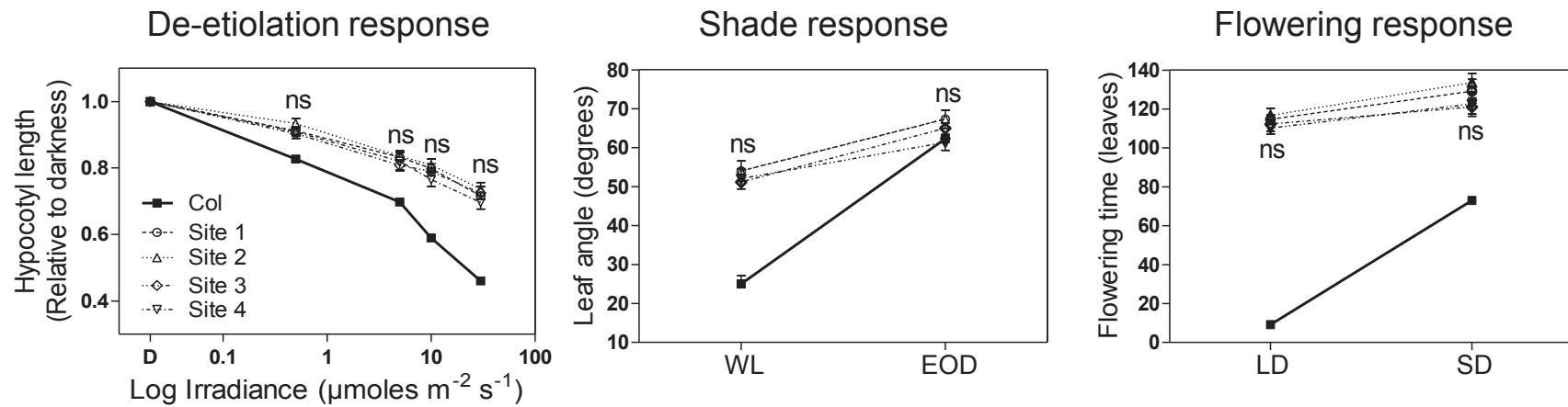


Figure S5. Physiological responses for Pat plants collected in the four different sites. Each point represents mean \pm s.e.m ($n \geq 6$). ns= not significant between Pat populations by ANOVA test.

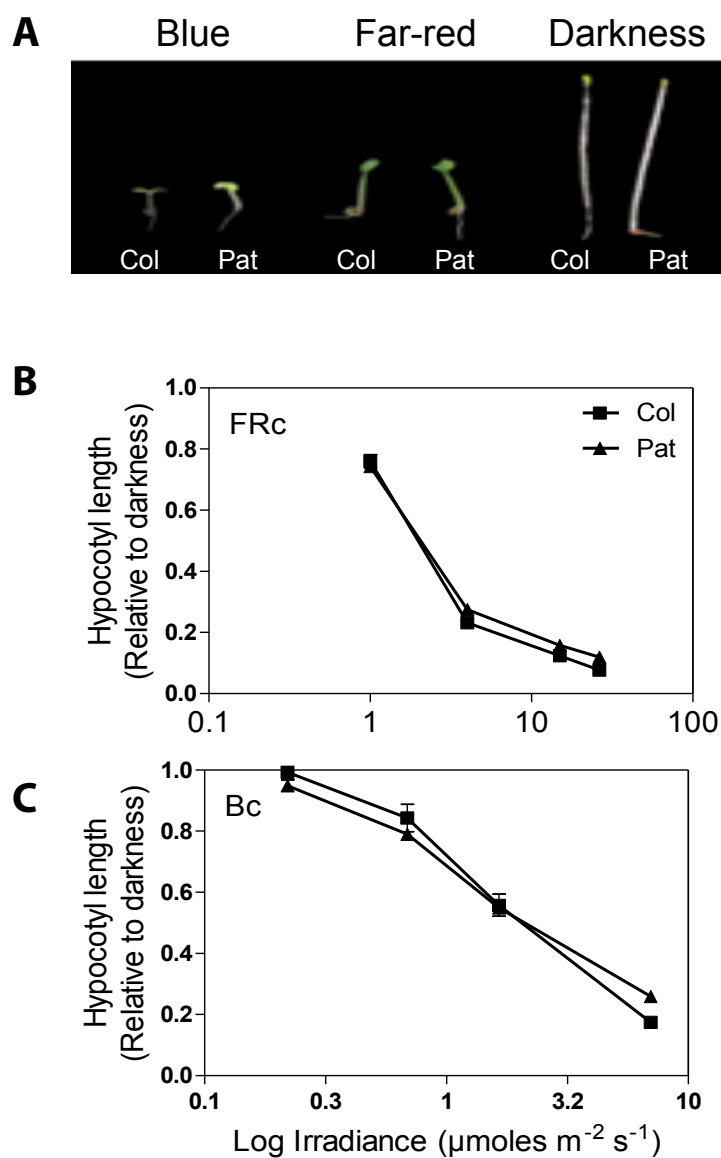


Figure S6. FRc and Bc light responses for Pat and Col-0 during de-etiolation. **A.** Representative photos for seedlings grown in Bc and FRc light, and dark control. **B.** Fluence-response curves for hypocotyl length in seedlings grown for 4 days in FRc light. **C.** Fluence-response curves for hypocotyl length in seedlings grown for 4 days in Bc light. Each point represents mean \pm s.e.m ($n \geq 6$).

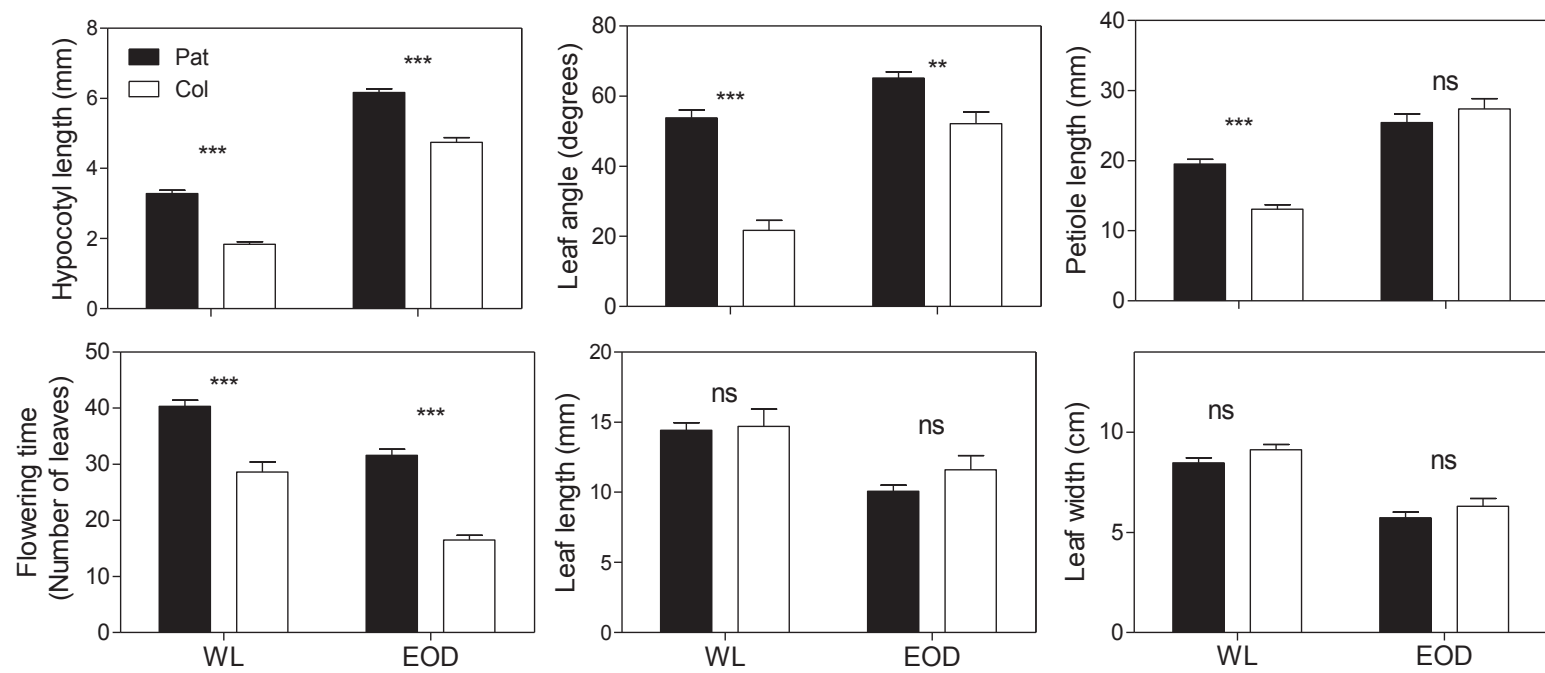


Figure S7. Hypocotyl length, leaf angle, petiole length, lamina width, lamina length, and flowering time for Col-0 and Pat plants cultivated in WL and simulated shade (EOD-FR). Each bar represents mean \pm s.e.m ($n \geq 6$). *** $P < 0.0001$, ns=not significant by ANOVA test.

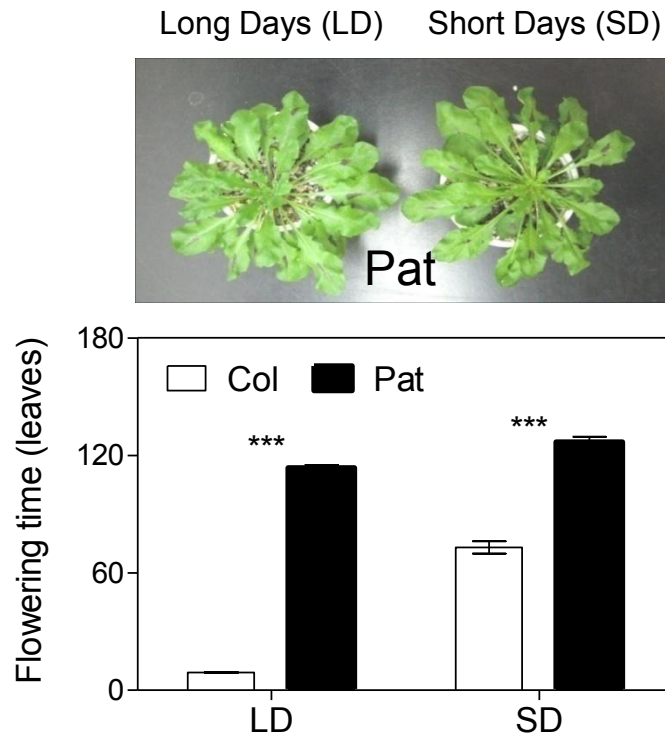


Figure S8. Photoperiodic flowering response without vernalization for Col-0 and Pat plants cultivated in long-day (LD) and short-day (SD). Each bar represents mean \pm s.e.m (n= 15). *** P< 0.0001 by ANOVA test. Representative photographs of 45-d-old Pat plants cultivated in LD or SD.

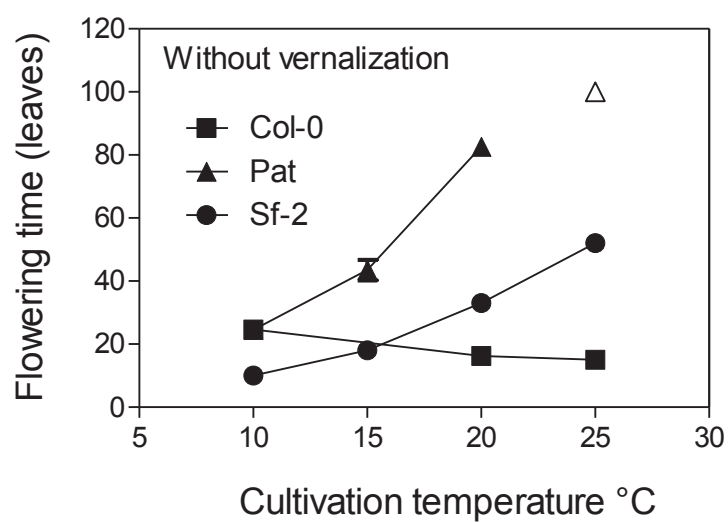


Figure S9. Flowering response at different cultivation temperatures. Plants were grown under continuous white light without vernalization in SD. Each point represents mean \pm s.e.m. ($n \geq 6$). Open symbol indicates that Pat plants had not flowered by the end of the experiment after 130 days.

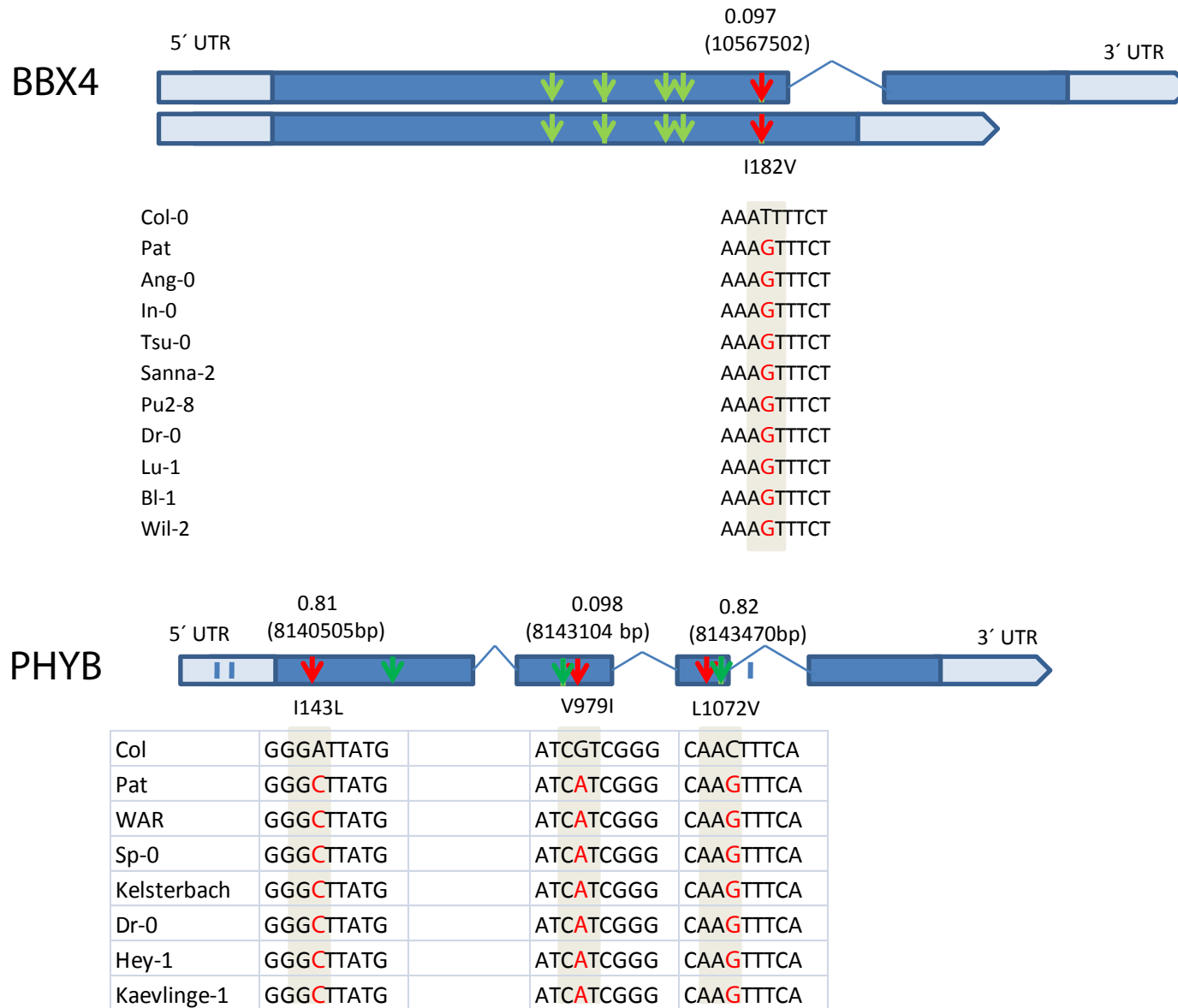


Figure S10. Diagrams of SNPs at BBX4 and PHYB genes. Arrows indicate synonymous (green) or non-synonymous (red) SNPs. The 5' and 3' untranslated regions (UTRs) are shown as light blue boxes, exons are indicated by dark blue boxes, and introns are drawn as arched lines. The vertical blue lines indicate SNPs in UTRs. The frequency and position of non-synonymous SNPs across 1135 *A. thaliana* accessions are indicated above the red arrows. The amino acid position with the reference and alternate amino acids encoded by the SNPs are shown below the red arrows.

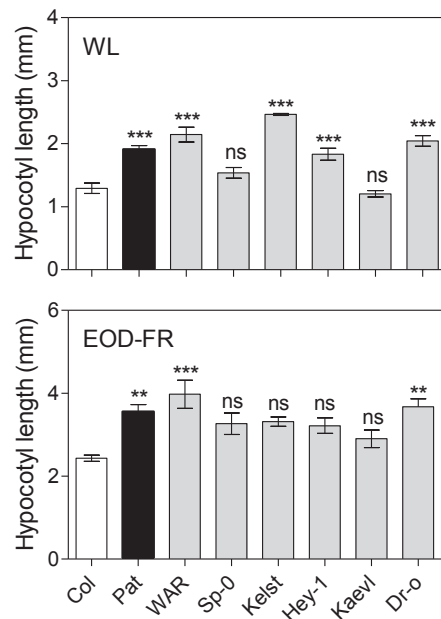


Figure S11. Hypocotyl length for accessions with the Pat allele conferring the V979I change in PHYB. Seedlings were grown in WL and simulated shade (EOD-FR). Each bar represents mean \pm s.e.m. ($n \geq 6$). *** $P < 0.0001$, ** $P < 0.001$, ns= not significant by ANOVA test.

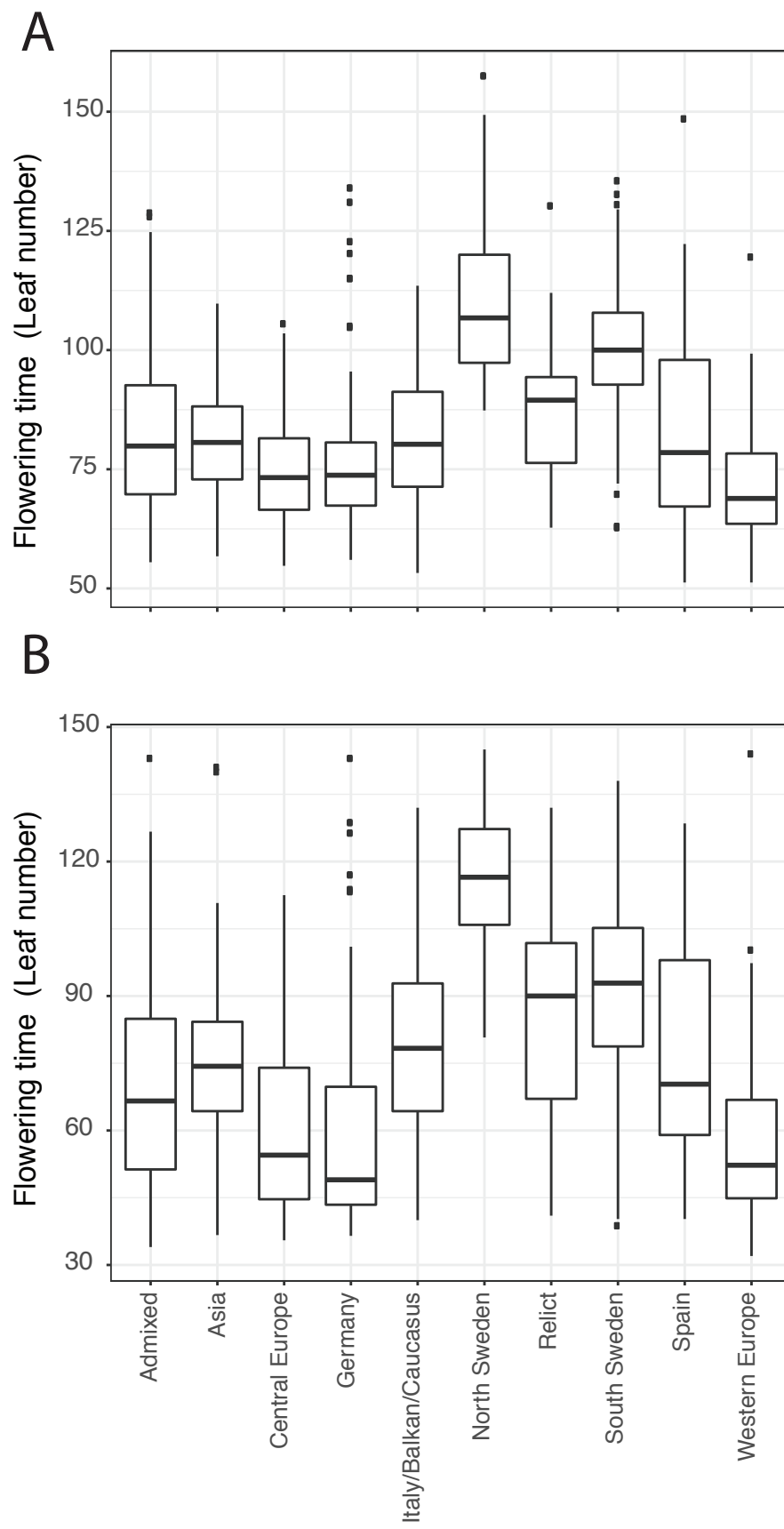


Figure S12. Flowering time (leaf number) at 10°C (A) and 16°C (B) for genetic groups (determined by ADMIXTURE) of worldwide *A. thaliana* accessions. Data were obtained from <https://arapheno.1001genomes.org/>.

Table S1: Primers used in this study

Primers for qRT-PCR		FW	RV
PP2A	AT1G13320	TAACGTGGCCAAAATGATGC	GTTCTCCACAACCGCTTGGT
BBX4	AT2G24790	CCGTTGGTGCCTGAAAGTG	GGAGATGACGAACACGGAGA
HY5	AT5G11260	CCATCAAGCAGCGAGAGGTCATCA	CGCCGATCCAGATTCTCTACCGAA
PIF3	AT1G09530	GTTTACCGCCATATCACTCTCTTGC	TGATGTGTCGTCTCCTGTTACTGTC
PIF5	AT3G59060	GAACCTGGCAAGAAGAACCTTAGC	AGTTGAAACAGAACAGTAACACAGAGAG
BBX31	AT3G21890	CGCGTGAGTTTCTGACAACA	GACTCGGCGTTCCTCTGTAG
ATHB2	AT4G16780	ATGACCCAAAACATACTTGAC	CAGGACCCACGCATTGACC
PIL1	AT2G46970	AAGACCACCTACGATGTTGCCAATG	GTAGCCTGAGTCTGAGAAACACACG
HFR1	AT1G02340	TGCCATCGCCGCTAATTCCG	ACCAAACCGTGAAGAGACTGAGG
FLC	AT5G10140	AAGCCTTGGATCATCAGTCAAAAGC	TCTAGTCACGGAGAGGGCAGTC
FRI	AT4G00650	AGG GCG TAG AGC ATT TAC	TAA TCC AAC TCT CAA TCT TCA
YLS8	AT5G08290	CACCACAATAAGTCAGTGGAGTAACTCCTTAC	GGATGAGACCTGTATGCAG

Table S2. Candidate genes

Associated with de-etiolation process (14)

PIF3	AT1G09530	
PIF1	AT2G20180	
HY5	AT5G11260	
HYH	AT3G17609	
PHYA	AT1G09570	
LAF1	AT4G25560	
BBX22	AT1G78600	
BBX4	AT2G24790	
GATA2	AT2G45050	
SPA	AT2G46340	
IAA7/AXR2	AT3G23050	
PHY3	AT3G22170	
CRY1	AT4G08920	
BBX25	AT2G31380	

Associated with shade avoidance response (10)

PHYB	AT2G18790	
ATHB2	AT4G16780	
PIL1	AT2G46970	
HFR1	AT1G02340	
PIF4	AT2G43010	
PIF7	AT5G61270	
BBX21	AT1G75540	
BBX24	AT1G06040	
COP1	AT2G32950	
PIF5	AT3G59060	

Associated with vernalization (11)

SVP	AT2G22540	
CLF	AT2G23380	
FES1	AT2G33835	
VRN1	AT3G18990	
VRN5	AT3G24440	
FRI	AT4G00650	
SWN	AT4G02020	
VRN2	AT4G16845	
FLC	AT5G10140	
LHP1	AT5G17690	
VIN3	AT5G57380	

Table S3. Geographical parameters at the collection sites of four Patagonia populations.

Site number	Vegetation type	Altitude (msm)	Latitude	Longitude	Slope	Exposure
1	Degraded grassland	859	46° 51' S	71° 52' W	30%	NE
2	Degraded grassland	807	46° 50,62' S	71° 51,97' W	30%	NE
3	Burned shrub steppe	530	46° 14,27' S	71° 38,68' W	10%	S
4	Shrub steppe	530	46° 14,27' S	71° 38,68' W	10%	S

Table S4. Vegetation cover and type for collection sites

Site number	Vegetation type	Vegetation cover (%)	Grasslands (%)	Shrubs (%)	Companion species
1	Degraded grassland	80	70	10	<i>Festuca pallences</i> , <i>Senecio filaienoides</i> , <i>Taraxacum officinale</i> , <i>Erodium cicutarium</i> , <i>Capsella bursa-pastoris</i> , <i>Cerastium arvense</i> , <i>Tripticum achilleae</i> , <i>Myosotis stricta</i> , <i>Cirsium</i> sp., <i>Loasa</i> sp., <i>Nardophyllum obtusifolium</i> , <i>Rumex acetosella</i> , <i>Acaena ovalifolia</i> and <i>Trifolium repens</i>
3	Burned shrub steppe	5	2.5	2.5	<i>Festuca pallences</i> , <i>Senecio filaienoides</i> , <i>Taraxacum officinale</i> , <i>Erodium cicutarium</i> , <i>Capsella bursa-pastoris</i> , <i>Cerastium arvense</i> , <i>Tripticum achilleae</i> , <i>Myosotis stricta</i> , <i>Cirsium</i> sp., <i>Loasa</i> sp., <i>Nardophyllum obtusifolium</i> , <i>Rumex acetosella</i> , <i>Acaena ovalifolia</i> and <i>Trifolium repens</i>
4	Shrub steppe	80	20	60	Shrubs: <i>Schinus polygamus</i> , <i>Berberis heterophylla</i> , <i>Colliguaya integerrima</i> , <i>Adesmia boronioides</i> , sprouting <i>Sisymbrium altissimum</i> . Grasses: <i>Erodium cicutarium</i> , <i>Descurania</i> sp., <i>Oenothera contorta</i> , <i>Solanum</i> sp., <i>Acaena ovalifolia</i> , <i>Microsteris gracilis</i> , <i>Chenopodium</i> sp., <i>Stipa speciosa</i> , <i>Festuca pallences</i> , <i>Schinus polygamus</i> , <i>Berberis heterophylla</i> , <i>Adesmia boronioides</i> , <i>Senecio filaginoides</i> , <i>Loasa</i> sp and <i>Acaena ovalifolia</i> .

Table S5. Soil texture and ionic composition of collection sites

Site number	Vegetation type	Soil texture	Soil texture			Ionic composition								
			Sand (%)	Clay (%)	Silt (%)	pH	C t (%)	Organic matter (%)	C.E.	Ca	Mg	K	Na	CIC
									mmhos/cm	meq/100g	meq/100g	meq/100g	meq/100g	meq/100g
1	Degraded grassland	loam	42.5	18.0	39.5	7.0	4.0	6.9	0.5	22.7	3.1	3.1	0.1	27.1
3	Burned shrub steppe	Gravelly sandy loam	76.8	7.5	15.7	7.2	2.7	4.7	0.9	14.7	4.3	2.9	0.1	19.3
4	Shrub steppe	Gravelly sandy loam	77.9	5.8	16.4	7.0	2.5	4.4	0.4	11.2	2.3	1.5	0.1	13.6

Table S6. Clusters of Pat private SNPs on chromosomes 4 and 5

* indicates non-synonymous changes leading to a premature stop codon

Gene ID	Gene Name	Annotation	Total Private SNPs	Synonymous Private SNPs	Nonsynonymous Private SNPs
AT4G21326	ATSBT3.12	subtilase	16	2	5*
AT4G21330	DYT1	bHLH transcription factor	6	2	3
AT4G21340	B70	transcription factor	9	0	2
AT4G21350	B80	U-box protein	4	3	1
AT4G21380	ARK3	receptor-like kinase	37	14	10
AT4G21390	B120	carbohydrate binding kinase	1	1	0
AT4G21400	CRK28	cysteine rich receptor-like kinase	25	10	9
AT4G21410	CRK29	cysteine rich receptor-like kinase	7	2	4
AT4G21420	n/a	gypsy transposon	8	0	0
AT4G21430	B160	transcription factor	10	3	4
AT4G21440	MYB102	MYB transcription factor	4	1	2
AT4G21450	n/a	PapD-like superfamily protein	1	0	0
AT4G21460	n/a	ribosomal protein S24/S35	24	6	6
AT5G63000	n/a	Tim family protein	20	8	0
AT5G63020	n/a	CC-NBS-LRR protein	22	12	10
AT5G63030	GRXC1	glutaredoxin family protein	2	2	0
AT5G63040	n/a	unknown protein	12	3	6
AT5G63050	EMB2759	embryo development	29	0	0
AT5G63060	n/a	Sec14p-like transporter	15	2	0
AT5G63063	n/a	plant thionin family	2	0	2
AT5G63065	n/a	plant thionin family	3	1	2
AT5G63070	n/a	ribosomal protein S19	5	1	4

Table S7. Polymorphisms in candidate genes related with seedling de-etiolation, shade avoidance and vernalization responses.

Only candidate genes with low SNP frequency in the 1001 Genomes dataset are shown (freq<0.10).

Only polymorphisms in the coding region of genes are shown.

Functional group	Gene	ATG ID	Number of Amino Acids	Starting Position	Ending Position	SNP	Amino Acid Position	Amino Acid Change	SNP frequency
De-etiolation	BBX4	AT2G24790	295	10566898	10568145	10567502	182	LxV	0.10*
De-etiolation/Shade	PHYB	AT2G18790	1173	8139881	8144430	8143104	980	Vxl	0.09*
Shade	PIL1	AT2G46970	417	19295490	19297826	19295970	299	MxV	0.02*
Shade	ATHB2	AT4G16780	284	9449114	9450743	9450492	247	QxP	0.05*
Vernalization	FRI	AT4G00650	314	269026	271503	269162	46	TxM	0.03*
Vernalization	VIN3	AT5G57380	644	23246395	23249504	23248316	78	MxL	0.03*