Supplementary information

The giant diploid faba genome unlocks variation in a global protein crop

In the format provided by the authors and unedited



Supplementary Figure 1. Collinearity of physical and genetic maps and ChIP-seq localisation of centromeres. **a**, Collinearity of physical and genetic maps. The antidiagonal alignments in chr1, chr2, chr4, and chr6 were a result of the arbitrary orientation of linkage groups in prior genetic maps. **b**, chromatin immunoprecipitation sequencing (ChIP-seq) localization of centromeres.

Vigna radiata Vigna angularis Vicia sativa Vicia faba Trifolium subterraneum Trifolium pratense Prunus persica Pisum sativum Phaseolus vulgaris Medicago truncatula Lupinus angustifolius Lupinus albus Lotus japonicus Lens culinaris Glycine max Cicer arietinum Cajanus cajan Arachis ipaensis Arachis duranensis Ó 10000 20000 30000 40000 50000 Number of genes

Supplementary Figure 2. A plot of homologous gene number in each of the 19 species. Singlecopy orthologues, single-copy homologous genes in the gene families shared among species; multiple-copy orthologues, multicopy homologous genes in gene families shared among species; unique paralogues, genes of the strain unique to the family; other orthologues, all other genes; unclustered genes, genes not clustered into any family. The horizontal bars represent the number of protein-coding genes.



Supplementary Figure 3. Distribution of the synonymous substitution rate (Ks) between paralogous gene pairs of faba bean and other legumes.



Supplementary Figure 4. **a**, Proportion of genes tandemly duplicated across species. **b**, Estimated age of tandem duplications.



Supplementary Figure 5. Full-length representation of *Copia* and *Gypsy* super family. *Ogre* (n=2,157,340) and *SIRE* (n=918,214) are the largest elements in *Gypsy* (n=2,564,151) and *Copia* (n=1,578,228) respectively. centre line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range; points, outliers



Supplementary Figure 6. **a**, Methylation pattern based on gene features including first exon, first intron, last exon, last intron, and the 5-kb flanking regions in the Hedin/2 genome. **b**, Expression comparison between all annotated genes and gene-body methylated genes (CG context) in young leaf tissue. All: n=34,221, CG: n=12,453. centre line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range; points, outliers



Supplementary Figure 7. Pretzel (pulses.plantinformatics.io) visualisations integrating the consensus genetic map from Webb et al. (2016) together with 3 QTLs for frost tolerance from Sallam et al. (2016) (left axis), aligned to the Hedin/2 genome (middle and right axes). The gene annotation for Hedin/2 is shown together with the three most abundant repeat types (*Ogre, SIRE* and rnd-1). Pretzel enables rapid, interactive interrogation of multiple data types and integration of legacy knowledge with the new assemblies.



Supplementary Figure 8. Quantile-Quantile plots of seed area GWAS. Analysis was performed using three methods (BLINK, farmCPU, EMMAX) and phenotypes from six trials (11, 22, 23, 25, 26, 30). BLUE - best linear unbiased estimator across trials.



Supplementary Figure 9. Quantile-Quantile plots of seed length GWAS. Analysis was performed using three methods (BLINK, farmCPU, EMMAX) and phenotypes from six trials (11, 22, 23, 25, 26, 30). BLUE - best linear unbiased estimator across trials.



Supplementary Figure 10. Quantile-Quantile plots of seed width GWAS. Analysis was performed using three methods (BLINK, farmCPU, EMMAX) and phenotypes from six trials (11, 22, 23, 25, 26, 30). BLUE - best linear unbiased estimator across trials.



Supplementary Figure 11. Quantile-Quantile plots of thousand grain weight (TGW) GWAS. Analysis was performed using three methods (BLINK, farmCPU, EMMAX) and phenotypes from six trials (11, 22, 23, 25, 26, 30). BLUE - best linear unbiased estimator across trials.

chr4 Region:3.31 (Mb)



Supplementary Figure 12. LD patterns in the genomic region surrounding SNP chr4_299823118 highly associated with seed size and located in gene Vfaba.Hedin2.R1.4g051440 (indicated in bold) orthologous to AtCYP78A genes. Green lines indicate SNP positions. Green star indicates significant SNP.



Supplementary Figure 13. PCA of libraries showing tissue and developmental stage differentiation. **a**, Hedin/2 libraries. **b**, Tiffany libraries.



Supplementary Figure 14. A dot plot alignment of PPO locus of Hedin/2 and Tiffany. The red and blue dots represent genes coloured according to strand.



Supplementary Figure 15. Principle behind individual annotations and 'transfer and gap fill' strategy. Hedin/2 annotation was transferred onto Tiffany using Liftoff. To prevent formation of chimeric gene models, caused for example by SVs, transferred models with in-frame stop codons were removed and replaced by Tiffany genes. Gene models unique to Tiffany were also added to the annotation. Overall, we observed that the 'transfer and gap fill' approach resulted in more syntenic genes and more genes with the same CDS length in both accessions.



Supplementary Figure 16. Dissected parts of a pod of cv Tiffany: pod wall, testa, cotyledon, funiculus and embryo axis.



Supplementary Figure 17. Composition and phenotypic differentiation of testa and cotyledon libraries from a developmental pod-fill gradient. centre line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range; points, outliers

Supplementary Table 3. Structural annotation statistics for Hedin/2 and Tiffany assemblies

Annotation	Gene total length	Gene average length	Gene number	Exon total length	Exon average length	CDS average length
Hedin/2 all (coding + IncRNA)	104,751,785	2,357	44,439	42,500,136	224	1,163
Hedin/2 complete (coding with complete CDS only)	95,042,832	2,777	34,221	39,546,400	231	1,164
Tiffany complete (coding with complete CDS only)	95,719,908	2,812	34,043	38,674,818	226	1,148

Supplementary Table 4. Annotation BUSCO completeness. Complete: gene length within two standard deviations of the BUSCO group mean length. Duplicated: Complete genes found with more than one copy. Fragmented: genes only partially recovered. Missing: genes not recovered. Groups checked: resolution

Annotation	Complete [S:Single	Fragmented	Missing	Groups	BUSCO database
	copy, D:Duplicated]			cnecked	
	96.8%[S:78.0%,D:18	2.0%	1.2%	255	embryophyta_odb10
Hadia/2 all	.8%]				
neuin/2 ali	94.7%[S:90.6%,D:4.	0.9%	4.4%	5366	fabales_odb10
	1%]				
	96.8%[S:78.0%,D:18	2.0%	1.2%	255	embryophyta_odb10
Hedin/2	.8%]				
complete	94.5%[S:90.4%,D:4.	0.9%	4.6%	5366	fabales_odb10
	1%]				
Tiffony	94.9%[S:77.3%,D:17	2.7%	2.4%	255	embryophyta_odb10
annality	.6%]				
complete	90.2%[S:85.7%,D:4.	1.0%	8.8%	5366	fabales_odb10
	5%]				

Species	Gene number	Genes in families	Un- clustered genes	Family number	Unique families	Average genes per family
Arachis duranensis	32,899	27,703	5,196	17,302	341	1.6
Arachis ipaensis	35,785	28,950	6,835	17,438	573	1.66
Cajanus cajan	30,842	26,425	4,417	16,507	256	1.6
Cicer arietinum	28,260	25,996	2,264	15,181	322	1.71
Glycine max	51,431	42,082	9,349	17,633	396	2.39
Lens culinaris	38,624	33,149	5,475	16,672	929	1.99
Lotus japonicus	25,460	23,104	2,356	14,613	324	1.58
Lupinus albus	37,837	30,588	7,249	17,318	289	1.77
Lupinus angustifolius	33,663	31,447	2,216	16,644	200	1.89
Medicago truncatula	31,571	28,978	2,593	16,857	368	1.72
Phaseolus vulgaris	27,263	25,366	1,897	16,890	112	1.5
Pisum sativum	43,677	32,990	10,687	18,684	1,476	1.77
Prunus persica*	26,631	22,127	4,504	14,490	604	1.53
Trifolium pratense	33,342	31,284	2,058	17,108	388	1.83
Trifolium subterraneum	39,578	29,312	10,266	17,066	740	1.72
Vicia faba	34,221	28,970	5,211	17,721	700	1.63
Vicia sativa	43,831	34,301	9,530	18,275	1,246	1.88
Vigna angularis	26,445	24,753	1,692	16,506	111	1.5
Vigna radiata	26,715	24,917	1,798	16,511	144	1.51

Supplementary Table 6. Summary statistics of gene families

*outgroup.

Supplementary Table 13. Markers which show stable associations with seed traits

Trait	Reference	Alternative	Seed enlarging	Chromosome	Position
			allele		
Seed Width/Length/Area	Т	С	ALT	1S	154,878,079
Seed Width/Area	G	С	ALT	1S	658,098,019
Seed Length/TGW	A	Т	ALT	1L	1,535,386,197
Seed Width/Area	A	С	ALT	1L	562,543,737
Seed Area/Width	G	A	ALT	2	1,335,507,680
Seed Width/TGW	A	Т	REF	3	88,655,178
Seed Length/Area	С	G	ALT	3	165,271,925
Seed Area/Width	С	Т	ALT	4	299,823,118
Seed Length/Area	G	С	ALT	4	620,271,647
Seed Width/Length	G	С	ALT	4	586,613,628
Seed Area/TGW	Т	С	ALT	5	497,436,071
Seed Area/Width	Т	С	ALT	5	615,661,639
TGW	G	A	ALT	5	968,403,536
Seed Length	G	Т	ALT	6	1,019,941,571
Seed Length/Area	Т	С	ALT	6	1,232,515,946

Supplementary Note: Pretzel instructions

The below steps demonstrate how to compare the seed weight QTL on chromosome 4 reported in Khazaei et al. 2014 with the GWAS result for Seed Area/Width reported in the present paper.

1. Log in

Access Pretzel Pulses URL: https://pulses.plantinformatics.io/

https://pulses.plantinformatics.io	x +	v - 0
→ C ☆ @ pulses,	plantinformatics io	۵.6 🛧 🗯 🖬 🖬 🕯
*		v2.15.2 released 2022-11-03 View on GitHub Sign Up 🗳 Log In
	Pretzel Pulses An interactive, web-based environment for navigating multi-dimensional data bean, including genetic maps, OTLs and chromosome-scale physical assem	asets in chickpea, lupin, field pea, lentil and faba biles.
	Pulses Pretzel is hosted by Agriculture Victoria. Dataset curation is funded by GRDC. Pretzel is developed by Agriculture Victoria.	
	Users can explore the pre-loaded datasets as well as upload their own content.	20 30 20 40 40 40 40 40 40 40 40 40 40 40 40 40
	A user guide focused on Pulses Pretzel, including animated examples, is available here.	
	A general Pretzel user guide is available here.	14
	Please contact info@plantinformatics.io with any questions or queries.	
	© LATEST RELEASE V2.15.2 © WEEKLY DOWNTIME FRIDAY 9-10AM EAST	
	Release Notes • 2022/09/03 /2(.14.2 PR #314 Group ownership of Datasets #283 • 2022/09/02 /2(.14.1 PR #313 Group ownership of Datasets #283 #216 comment 20 Aug 2021	

Click Log In in top right

https://pulses.plantinformatics.i: × +			× .	- 0	×
← → C △ @ pulses.plantinformatics.io/login		ං මු ල් 1/2 15 2 minored 2022 11 02) (හා හා Cithiub	☆ ★ E		1
	Skovbjerg_2022	← €, (d) V2.15.2 released 2022-11-03 V/ew on GilHub	x ★ ■ Sign Up 🕻	Log It	

Log in with the following details: Username: FabaBeanGenome Password: ViciaFaba99 2. Search for markers of interest

Click the Search tab at the top of the left panel, and cut/paste the markers from Khazaei et al. 2014, Figure 3, around the seed weight QTLs reported on chromosome 4 (Vf_Mt8g020800_001, CNGC4, Vf_Mt7g038120_001, Vf_Mt8g040150_001) and click Search.



3. Add dataset to the view

Click the + button next to Faba bean integrated map Chr-IV



4. Add physical genome position with marker positions

Click on the Explorer tab on the left side panel.

https://pulses.plantinformatics.ic × +							~	-		×
← → C ☆ 🔒 pulses.plantinformatics.io/mapvie	w?mapsToView[]=61dfd11	18e784104e19f0f6	ac#left-panel-exp	olorer			• 순 ☆	*) :
Map Viewer Change Password & Gro	oups 👗					v2.15.2 released	2022-11-03 View on	GitHub	Logo	iut 🕞
Explorer View QSearch Opload	<									
O Datasources	*				Fabe been integrated map : Chr-IV					
https://pulses.plantinformatics.io					20					
FabaBeanGenome					40 -					
78015ab2j1vlc8DZO5ybEsP7nDBQs9wed2pO9I					e0_					
Connect new datasource					80 -					
					100 -					
🗁 Datasets 🖸 📃 💽					120 -					
All O Recent O Favourites O Viewed					1					
Group Viewed Chromosomes										
					194-					
Search All Datasets					200 -					
Case insensitve Match all					220 =					
pulses.plantinformatics.io										
Trait Ontology Genetic Map QTL	Q Dataset *	Features (0)	Paths 0	III Genotypes					4	×
All Datasets					Column Stretch Horizontal	Auto Column Width	Editable	Save	To Dat	abase
33										
Upload new data										
+ 20210909_Fababean_SNP_SSR_Seq_Info										

Add the following dataset:

- Vfaba_hedin_v1.Vfaba_v2_markers - chr4



The genetic map is now projected into the physical chromosome assembly.

5. Reverse the orientation of the inverted genetic consensus map

Click the text (Faba bean integrated map : Chr-IV) at the top of the left axis to open the axis menu. Click the button with up/down arrows to invert the axis.

https://pulses.plantinformatics.ic × +			v -	o ×
← → C ☆ 🌢 pulses.plantinformatics.io/mapview?mapsToView[]=	61dfd118e784104e19f0f6ac		아 년 ☆ 🗍 🛛	. :
Map Viewer Change Password & Groups 👗			v2.15.2 released 2022-11-03 View on GitHub	Logout 🕞
Explorer View QSearch Upload Your View QSearch View View	Faba been integrated map : Chr-IV		Vfaba_hadin_v1:chr4	
Vitaba. Tiffany.v1 Hedin_alignment Vitaba. Tiffany.v1 Annotation Vitaba. Tiffany.v1 annotation Vitaba. Dedin_v1 Vitaba. Dedin_v1 Vitaba. Dedin_v1 Carelio-Perdomo_SNPs Vitaba. Dedin_v1 LGC_Genomics_KASP Vitaba. Dedin_v1 LGC_Genomics_KASP Vitaba. Dedin_v1 Repeats SRE Vitaba. Dedin_v1 Repeats SRE Vitaba. Dedin_v1 Repeats SRE Vitaba. Dedin_v1 SPET_NORFAB_SNP_po Vitaba. Dedin_v1 Seed_QTLS Vitaba. Dedin_v1 Steet_QTLS	Image: Second	2021 250	- 50M - 60M - 60M - 60M - 100 - 120 - 120 - 120 - 120 - 120	
Vfaba_hedin_v1.Vfaba_v2.physical Vfaba_hedin_v1.Vfaba_v2.physical Vfaba_hedin_v1.Vfaba_v2.physical Dataset	* Features Paths Eatures Ea			4 ×
viaua_nouni_vi.viaua_vz_malkels Faba bes	an integrated map	apiHost	httpspulses_plantinformatics_io	*
		Meta	View	- 1
+ chr3 chr4 + chr5				
+ chr6		♥ Metadata {2}		
Vfaba_hedin_v1.annotation Created Last Update	1/13/2022, 6:13:28 PM	type:Genetic Map citation:Sudheesh S, Kimb Construction of	er RBE, Braich S, Paull JG, Forster JW, Kaur S (20) integrated linkage map and detection of quantitation	19) ve

Click the "x" in the top right of the axis menu to close it.

6. Add Jayakodi et al. 2022 GWAS results to the view Select the following dataset from the list in the left panel:

- Vfaba_hedin_v1.Jayakodi2022 - chr4

https://pulses.plantinformatics.io x +			∨ – □ ×
← → C ☆ 🔒 pulses.plantinformatics.io/mapview?	mapsToView[]=61dfd118e784104e19f0f6ac#left-panel-explorer		아 🖻 🏠 🌲 🖬 😩 🗄
Map Viewer Change Password 1 Group	۶Ă	v2.15.2 released 2	2022-11-03 View on GitHub Logout 🗗
Explorer View QSearch & Upload			
+ Nura x Faran	Faba bean Integrated map : C	Chr-IV Vfaba_hedin_v	y1 : ohr4
+ Nura x Farah QTL	220 =		
+ Quasar x Hivema	200 -		
	1		A
+ Sallam_2016_QTL		- 4	ADOM
+ Silian x Hiverna 🔒	10		
+ Vfaba.Tiffany.v1	140-	- 61	10
	120 -		
Viaba. Illiany.vit. Hedin_alignment		- 8	00M
+ Vfaba. Tiffany.v1. annotation	100 -	-1	1.03
+ Vfaba_consensus_Webb_2016	80 -		
➡ Vfaba bedin v1 🚔	60 -	- 1.	126
	40		
+ Vfaba_hedin_v1.Carrillo-Perdomo_SNPs		1	.40
- Vfaba_hedin_v1.Jayakodi2022	20 -	1	149
+ chr1L + chr1S + chr2	L 0.0	L**	
Cabra abri			
+ chr6		Canaturar	1. 4
+ Vfaba_hedin_v1.LGC_Genomics_KASP		Genotypes	7 *
+ Vfaba hedin v1.Repeats.Ogre	Vfaba_hedin_v1.Jayakodi2022	apiHost httpspulses_plantinfo	irmatics_io
	Namespace	View View View	1
VTaba_nedin_v1:Repeats.SIRE	Parent Vfaba_hedin_v1	w Metadata (4)	
+ Vfaba_hedin_v1.Repeats.rnd-1	Created 12/1/2022, 5:10:36 PM	shortName : Jayakodi2022	
+ Vfaba_hedin_v1.SPET_NORFAB_SNP_po sitions_v1	Last Updated 12/1/2022, 5:10:36 PM	citation : Jayakodi et al. 2022, The giant diploid a global protein crop.	faba genome unlocks variation in 🗸

7. Colour GWAS peak markers by trait

In the View tab of the left panel, select Trait.

https://pulses.plantinformatics.ic x +			V	- 🗆 ×
← → C ☆ 🔒 pulses.plantinformatics.io/mapview	mapsToView[]=61dfd118e784104e19f0f6ac#tab-view-Trait		• E \$	* 🗉 🛎 🗄
Map Viewer Change Password & Group	is 👗		v2.15.2 released 2022-11-03 View on Gi	tHub Logout 🕞
Explorer View Qsearch > Upload	Faba been integrated map : Chr-IV		Vfaba hedin v1:chr4	
	220 -			
Blocks Trait Ontology	200 -		- 200M	
Make this tab sticky	197-		- 400M	
Show / Hide QTLs of all Traits	1		- co	
Seed Area/Width	120 -		- 800M	
Seed Width/Length	80 -		- 1.03	
Seed Length/Area	60 -		- 128	
	40 -		1.43	
F Chart Options	0.0		1,66	
E Selected Axis Options				
ズ Flip Region ✔ Reset Zooms	O Dataset Features O Paths O III Genoty	pes		4 ×
	Vfaba_hedin_v1.Jayakodi2022	apiHost	httpspulses_plantinformatics_io	
Patns - Density - Sample 400	Namespace	Meta	View C Edit	_
	Parent Vfaba_hedin_v1	▼ Metadata {4}		
Path Stream Limit : 1000	Created 12/1/2022, 5:10:36 PM	shortName : Jaya	skodi2022	variation in
Binned Feature Counts	Last Updated 12/1/2022, 5:10:36 PM	a glo	obal protein crop.	vertecton in

8. Annotate the view with marker names

Click the triangles on the left axis to add the marker name to the visualisation.

https://pulses.plantinformatics.ic. x +					~ -		×
ightarrow $ ightarrow$ $ ightarrow$ $ ightarrow$ pulses.plantinformatics.io/mapview	/?mapsToView[]=61dfd118e784104e19f0f6ac#tab-	view-Trait		07 (É	☆ 🛊) (
Map Viewer Change Password 💄 Group	ips 👗			v2.15.2 released 2022-11-03 Viet	v on GitHub	Logo	ut C+
Explorer View QSearch & Upload	<						
🛛 Displayed Data 🛛 🔅 💽	A Faba bean I	integrated map : Chr-IV		Vfaba_hedin_v1 : chr4			
Blocks Trait Ontology		200 -		- 2004			
Make this tab sticky	Vf_Mt8g040150_001	180-		- 400M			
	Vt_Mt7g038120_001 CNGC4	10					
Show / Hide QTLs of all Traits	Vf_Mt8g020800_001	1-		- 60			
Seed Area/Width		120 -		- 800M			
* 1		100 -		105			
Seed Width/Length		80 -		- 1.00			
Seed Length/Area		60 -		- 1.23			
*1		40 -		- 1.43			
		20 -					
🗲 Chart Options		0.0		1.03			
: Selected Axis Options							
✓ Flip Region ✓ Reset Zooms	② Dataset * Features ① ② Pat	hs 🚺 🛛 🛄 Genotypes				4	×
	Vfaba_hedin_v1.Jayakodi2022		apiHost	httpspulses_plantinformatics_io			
Paths - Density - Sample 400	Namespace		Meta	View C Z Edit			
	Parent Vfaba_hedin_v1		♥ Metadata {4}				
Path Stream	Created 12/1/2022, 5:10:36 PM		shortName : Jaya	kodi2022			
			addated and 1 December 2	odi at al 2022 The giant diploid faba genome	mlocks vari	iation i	4n

9. Zoom in to region of interest

Click and drag on the region of interest to create a brush.



Click the Zoom button to zoom the axis to the region defined by the brush.



Zoom in to the right axis in the same way. Click the diamond on the side of the right axis to add a triangle at that position, then click the triangle to add the name of the QTL.

thttps://pulses.plantinformatics.io: x +			v	- 🗆 ×	:
\leftrightarrow \rightarrow \mathcal{C} \bigtriangleup \triangleq pulses.plantinformatics.io/mapview	v?mapsToView[]=61dfd118e784104e19f0f6ac#tab-view-Trait		• E \$	* 🗆 🛎 🗄	
Map Viewer Change Password 1 Grou	ips 🛦		v2.15.2 released 2022-11-03 View on Gif	Hub Logout 🕞	•
Explorer View Q.Search & Upload Olspłayed Data	Faba been integrated map : Chr-IV		Vfaba_hedin_v1 : chr4		
Blocks Trait Ontology 3 O Make this tab sticky	VT_M8g040150_001		- 2004		
Show / Hide QTLs of all Traits	170 - VL_M79038120_001 ¹ №-		54_209823118		
Seed Width/Length	CNGCI ►		- 3554		
Seed Length/Area	VE_M88g020600_001_{20-		- 4500		
	Zoom Reset		Zoom Reset		
Selected Axis Options					
≻⊄Flip Region	Dataset Features Paths Ill Gen	otypes		19 ×	:
	Vfaba_hedin_v1	apiHost	httpspulses_plantinformatics_io		*
Paths - Density - Sample 400	Namespace	Meta	View C Zedit		
	Parent	♥ Metadata {2}			
Path Stream Limit : 1000	Created 7/21/2022, 8:52:37 PM	source : Faba	Bean Genome Consortium		
Binned Feature Counts	Last Updated 7/21/2022, 8:52:37 PM	url : <u>https:/</u>	<u>/projects.au.dk/fabagenome/genomics-data</u>		÷

The resulting visualisation shows the S4_299823118 GWAS peak marker for Seed Area/Width as reported in this paper co-locates with the seed weight QTL reported in Khazaei et al. 2014.