

Supplementary Materials

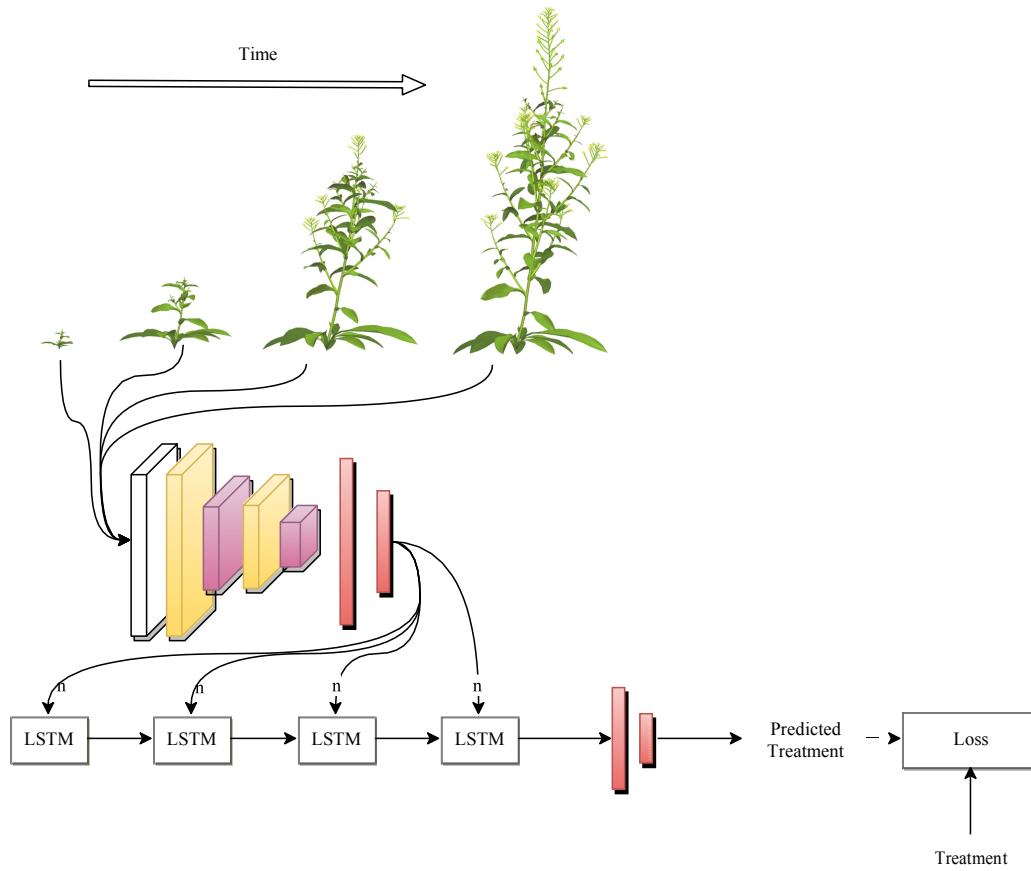


Figure S1: The deep network used in learning an embedding. A CNN takes a sequence of images at various timepoints and feeds outputs to an LSTM, which in turn is used to predict the treatment. The LSTM is removed and the CNN is retained to embed new samples.

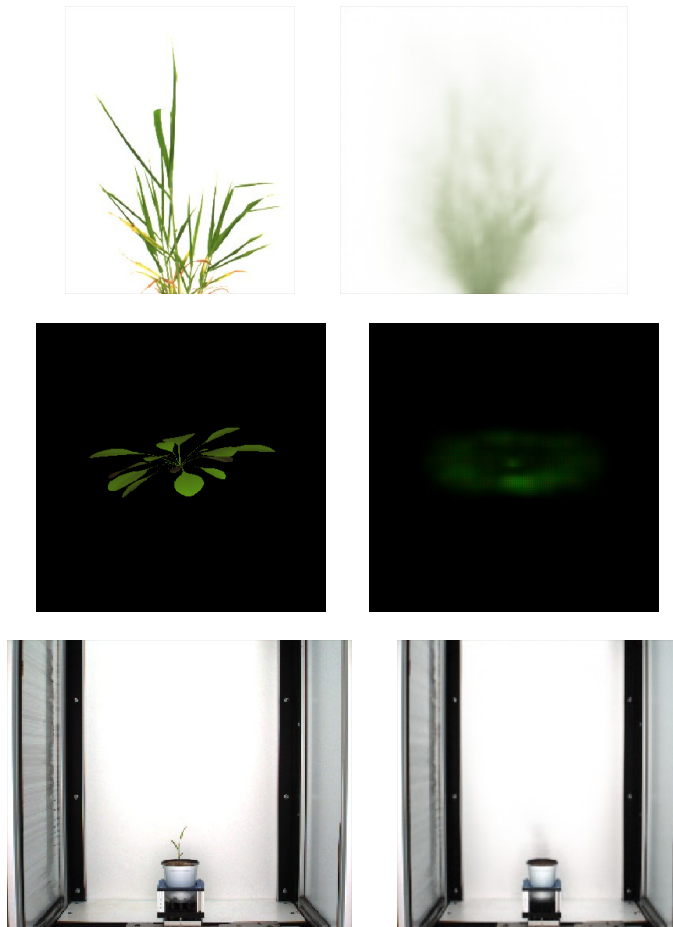


Figure S2: Real images from the *Setaria*, synthetic Arabidopsis, and sorghum datasets (left), and the same images predicted from their latent space encodings by a decoder network (right).

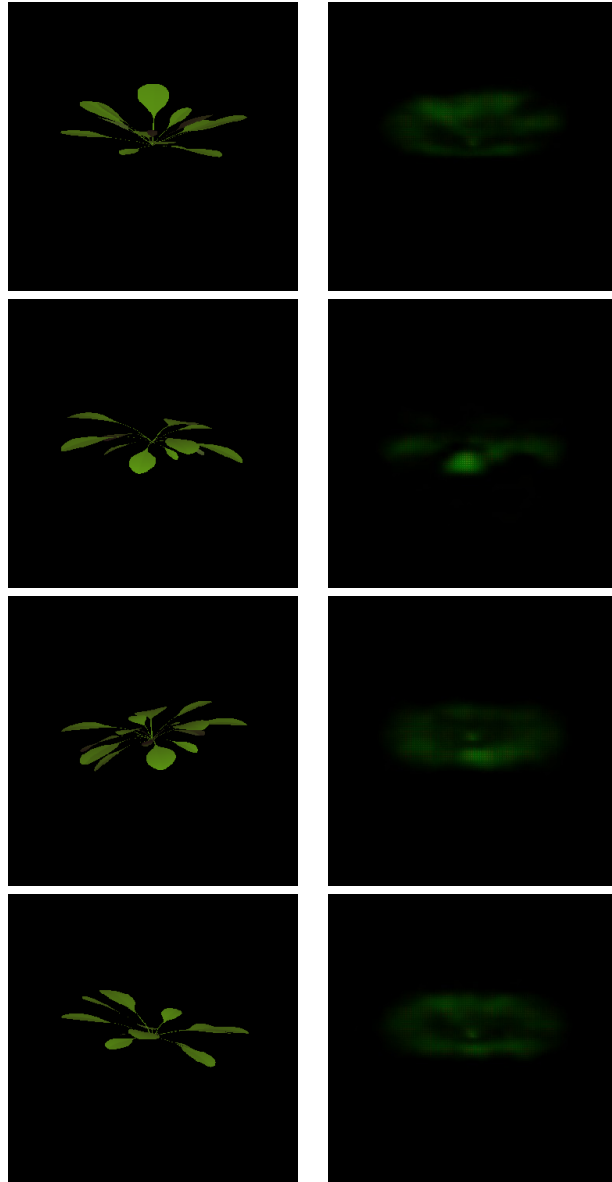


Figure S3: Additional examples of synthetic Arabidopsis rosettes (left) decoded from their latent space vectors (right).

	Size	Depth	Stride
conv + relu	3×3	1	1
max pool	3×3		3
conv + relu	3×3	16	1
max pool	3×3		3
conv + relu	3×3	32	1
max pool	3×3		3
conv + relu	3×3	32	1
max pool	3×3		2
fully connected + relu	64		
fully connected	16		

Table S1: Architecture details for the convolutional neural network used in the embedding. All pooling layers are followed by batch normalization.

	Size	Depth	Stride
conv + relu	3×3	16	1
upsample	3×3		2
conv + relu	3×3	32	1
conv + relu	3×3	32	1
conv + relu	3×3	32	1
upsample	3×3		2
conv + relu	3×3	32	1
conv + relu	3×3	32	1
conv + relu	3×3	32	1
upsample	3×3		2
conv + relu	3×3	64	1
conv + relu	3×3	64	1
conv + relu	3×3	64	1
upsample	3×3		2
conv + relu	3×3	64	1
conv + relu	3×3	64	1
conv + relu	3×3	64	1
upsample	3×3		2
conv + relu	3×3	32	1
conv + relu	3×3	32	1
conv + relu	3×3	16	1
upsample	3×3		2
conv + relu	3×3	16	1
upsample	3×3		2
conv	1×1	16	1

Table S2: Architecture details for the decoder network. The upsample blocks refer to transposed convolutions.

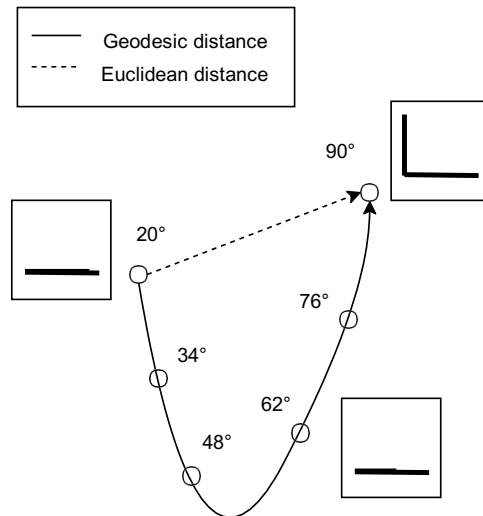


Figure S4: Distance between embeddings of images of lines at different angles in a hypothetical latent space. Here, the *semantic distance* is the difference in the interior angle. The euclidean distance between images in the image space is constant, and the euclidean distance (dotted arrow) between their encodings in the latent space is evidently not representative of the semantic distance. However, the geodesic path (solid arrow) between images represents the semantic distance well.



Figure S5: Untreated (left) and treated non-resistant (right) synthetic Arabidopsis plants at the final timepoint, showing differences in leaf elevation angle.

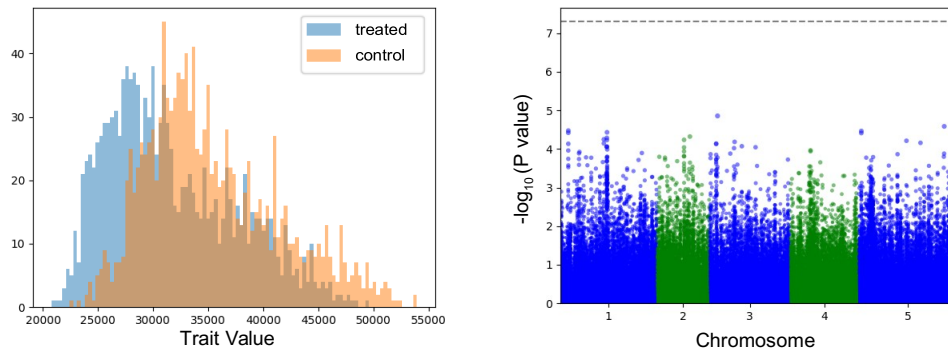


Figure S6: Ablation experiment using euclidean image distance between each pair of images in the sequence for the synthetic Arabidopsis dataset. The naive solution fails to recover the simulated tolerance QTL on chromosome 1.



Figure S7: Well-watered (left) and water-limited (right) examples of a particular line from the *Setaria* RIL population [12]

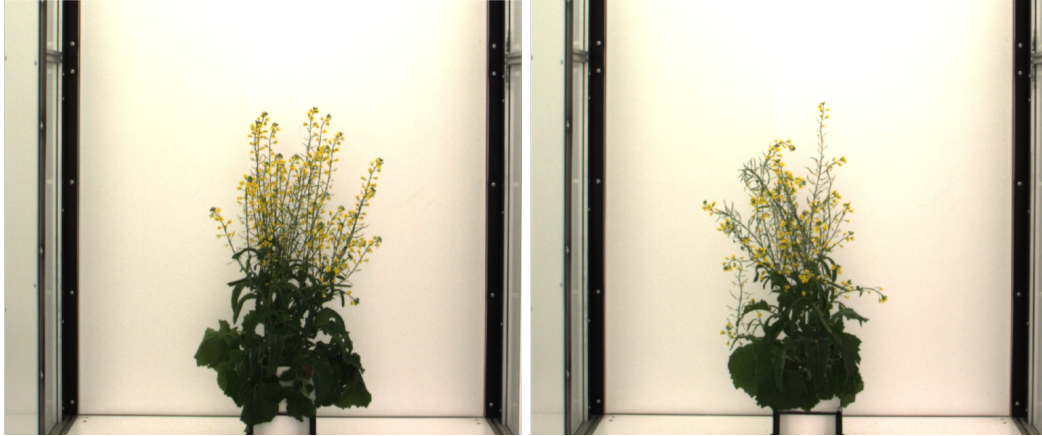


Figure S8: Well-watered (left) and water-limited (right) examples of a particular line from the *B. napus L.* NAM population

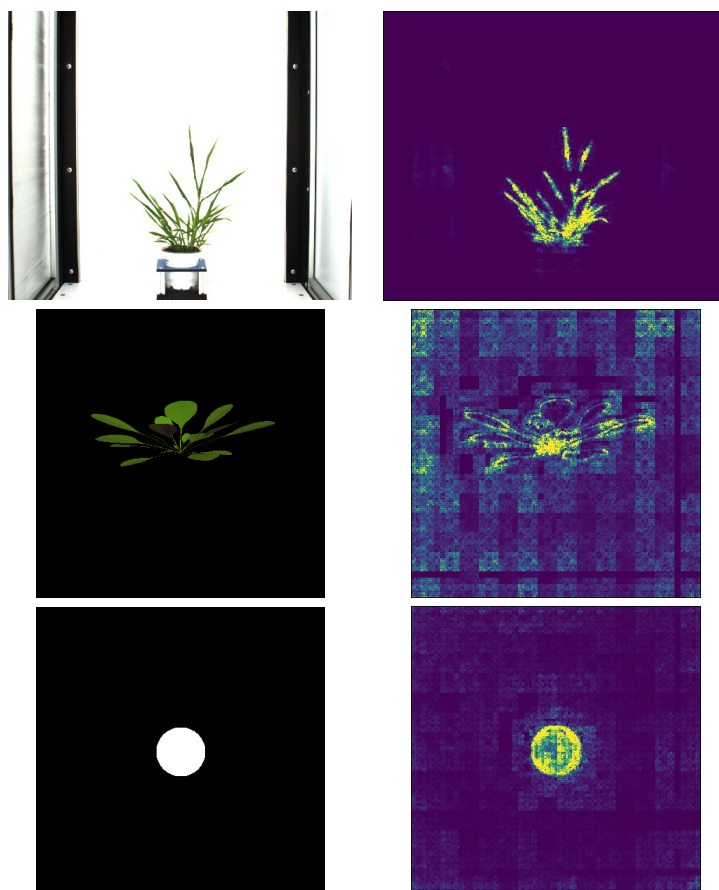


Figure S9: Example images from the *Setaria*, synthetic Arabidopsis, and synthetic circles datasets (left) and corresponding saliency maps generated using guided backpropagation (right). Intensity is higher for the pixels which have high saliency with respect to the latent space embedding of the image. The *Setaria* image is from an experiment carried out without cropping to include the background in the saliency demonstration.