Graduate School Course on Post-genomics and bio-informatics

A practical approach to phenomics

Darren M Wells

darren.wells@nottingham.ac.uk

Overview

- Definitions
- Technologies
- Examples
- Case study & tour – the Hounsfield Facility

Phenomics – some definitions

Phenotype: observable characteristics of an organism – result of genotype x environment interactions

Phenome: “the full set of phenotypes of an individual” (Houle *et al*., 2010)

Phenomics:

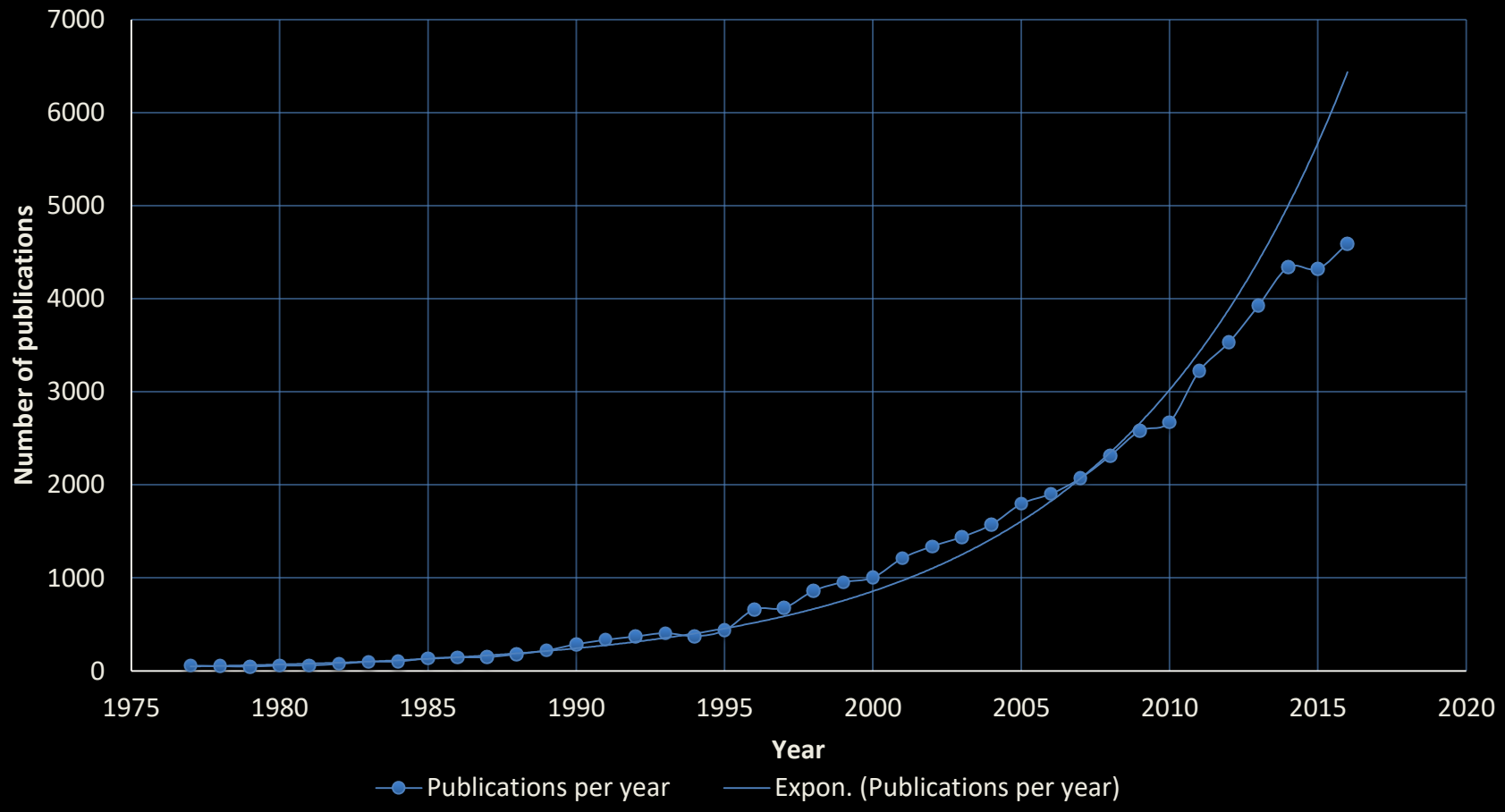
“…an emerging transdiscipline dedicated to the systematic study of phenotypes on a genome-wide scale” (Bilder, 2009)

Phenomics – characteristics

Phenomics is (usually):

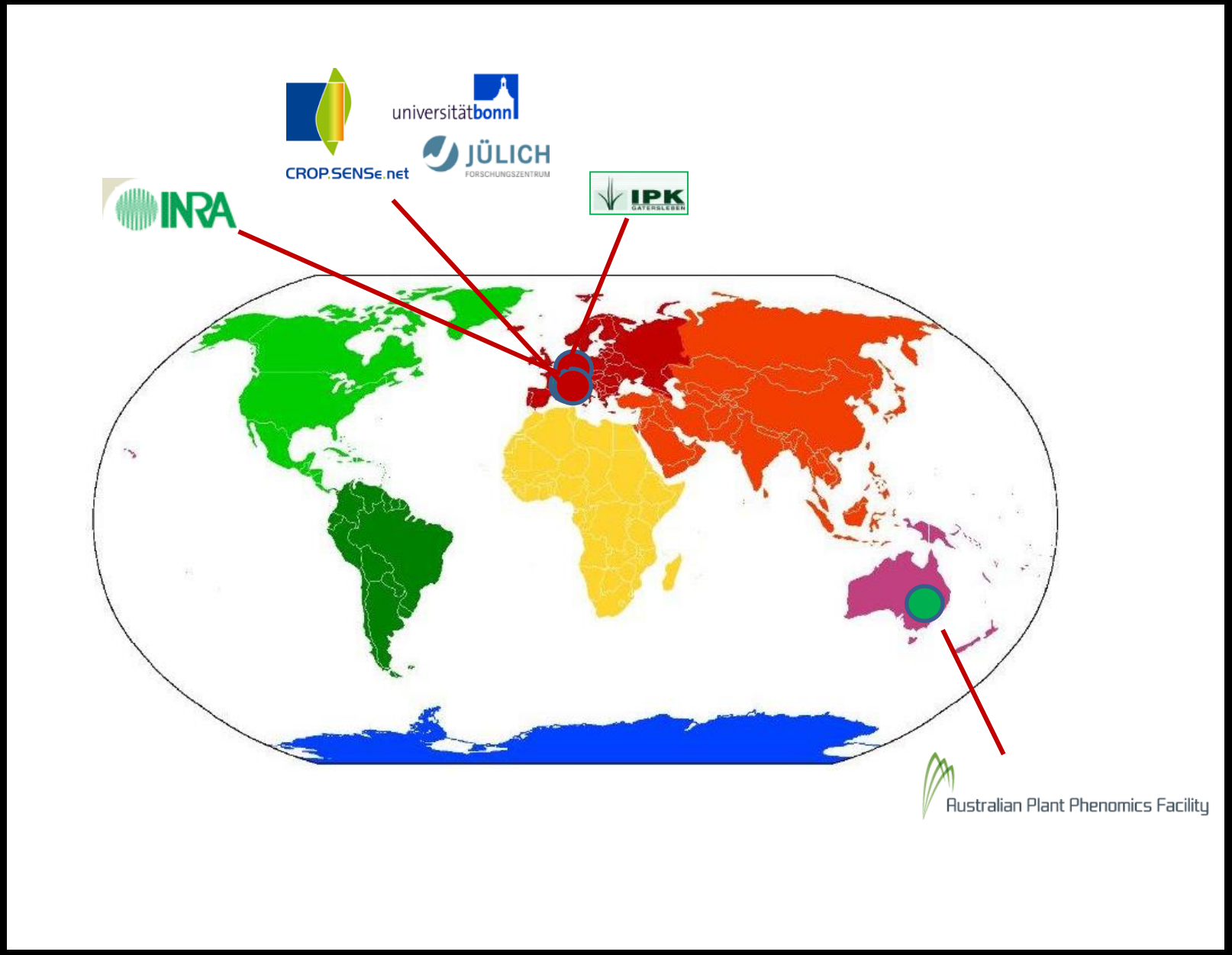
- Multi-disciplinary
- Multi-scale (cellular, organ, organism, population)
- High-throughput phenotyping

A RAPIDLY INCREASING ACTIVITY

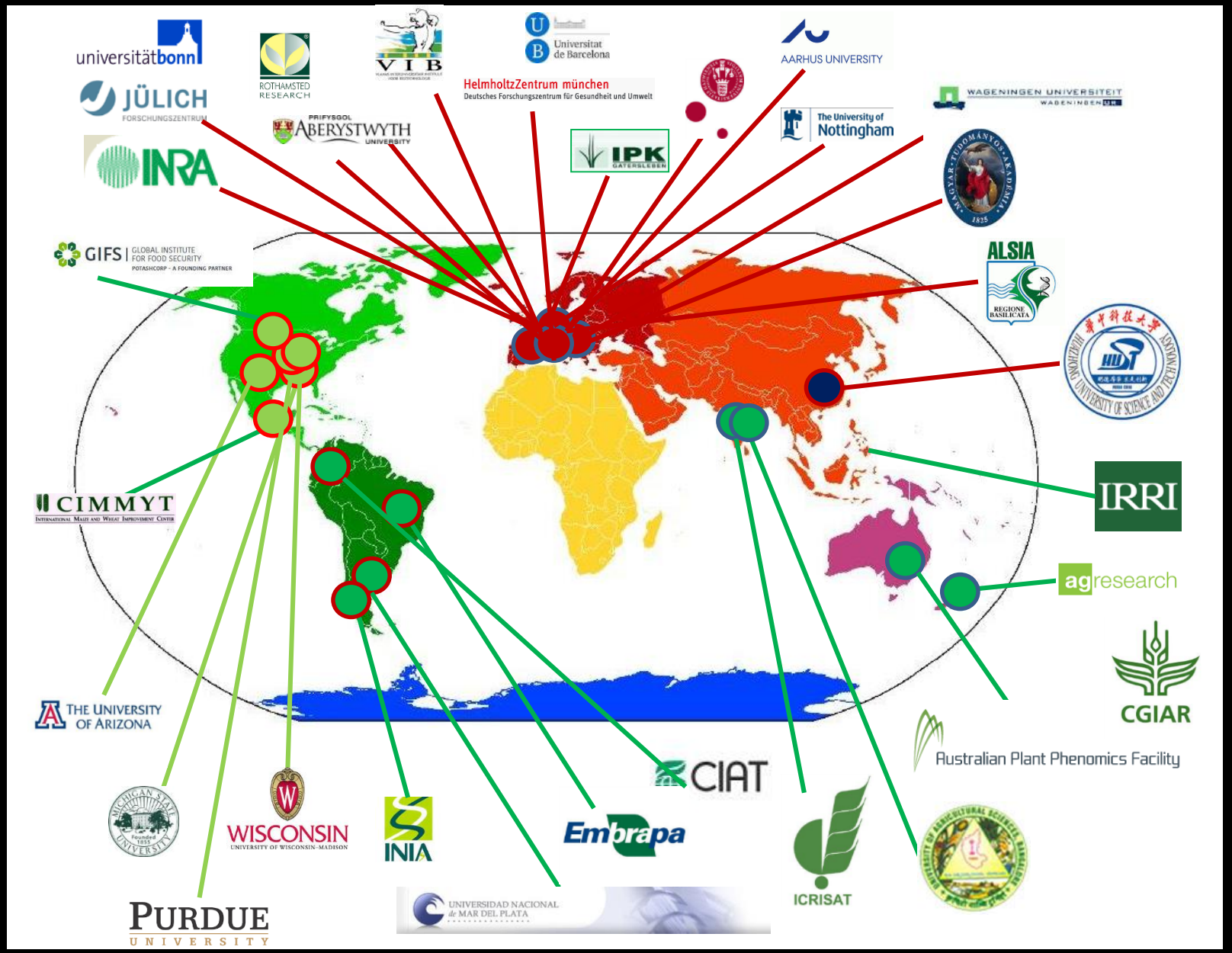


- Publications per year matching the query "plant & phenotyp*"

PHENOTYPING FACILITIES: 2008



PHENOTYPING FACILITIES: 2018



Phenotyping technologies

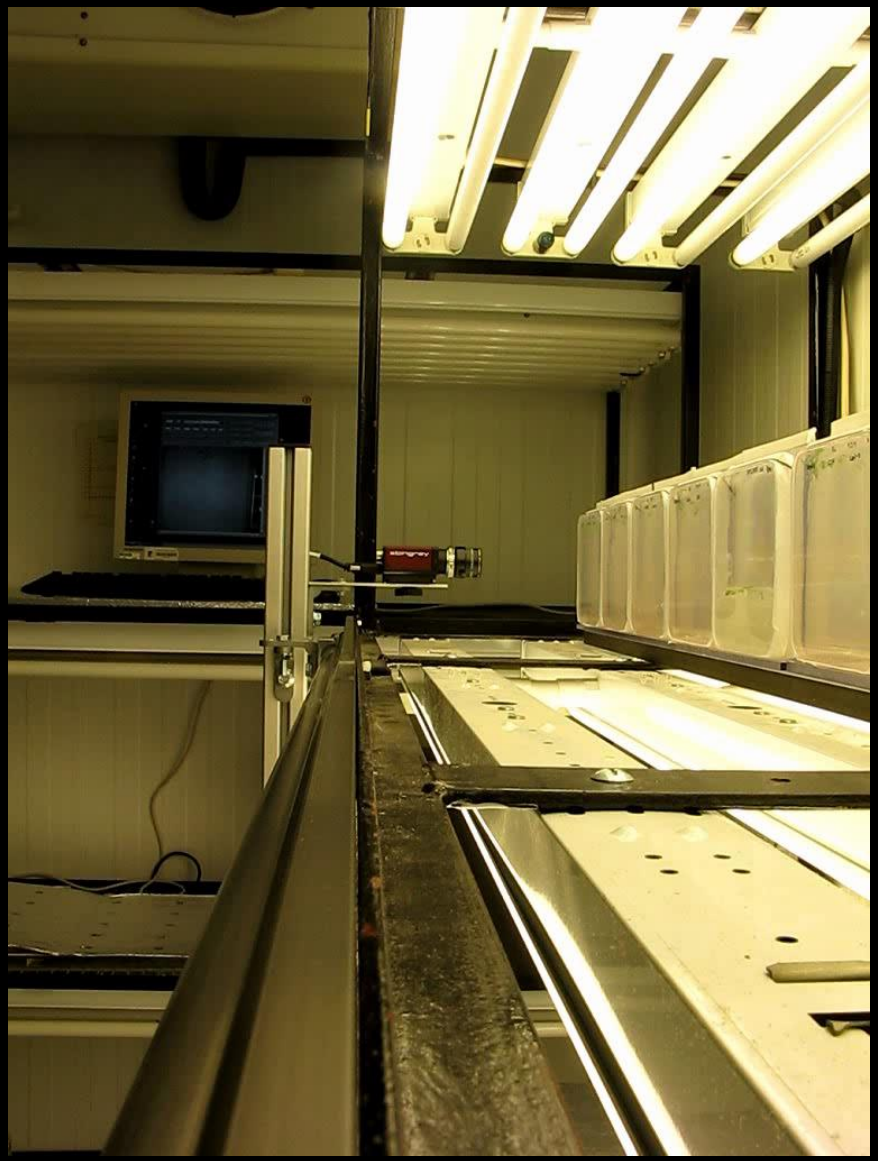
- Many and varied
- Selected according to availability and amenability to automation
- Often based on imaging techniques:
  - Visible/NIR/hyperspectral
  - Tomography (OPT, MRI, CT, PET etc)
  - Microscopy (light, EM)

Automated image acquisition and analysis

RootTrace I

Video 8x speed

French et al. (2009)*; Wells* et al. *(2012)*



RootTrace II

Naeem et al. *(2011); Wells* et al. *(2012)*

Examples

- Cell-scale: hormone dynamics in *Arabidopsis*
- Organ-scale: zebrafish larvae
- Organism level: wheat seedling root systems
- Population level: field crops
- Case study: developing a phenotyping facility at UoN

Cell-scale phenotyping - example

Development of fluorescence-based hormone sensors in *Arabidopsis* allows cellular-level modelling of distribution and response

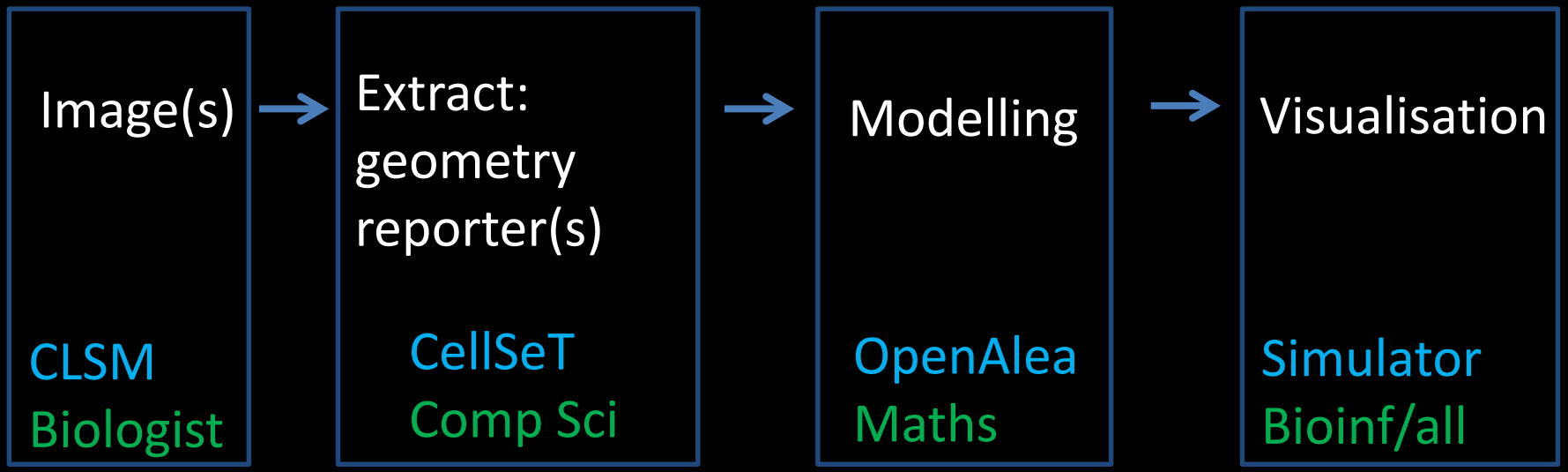
Required input from biologists, biophysicists, bioinformaticians, computer scientists (image analysts), mathematical modellers

Cellular resolution modelling

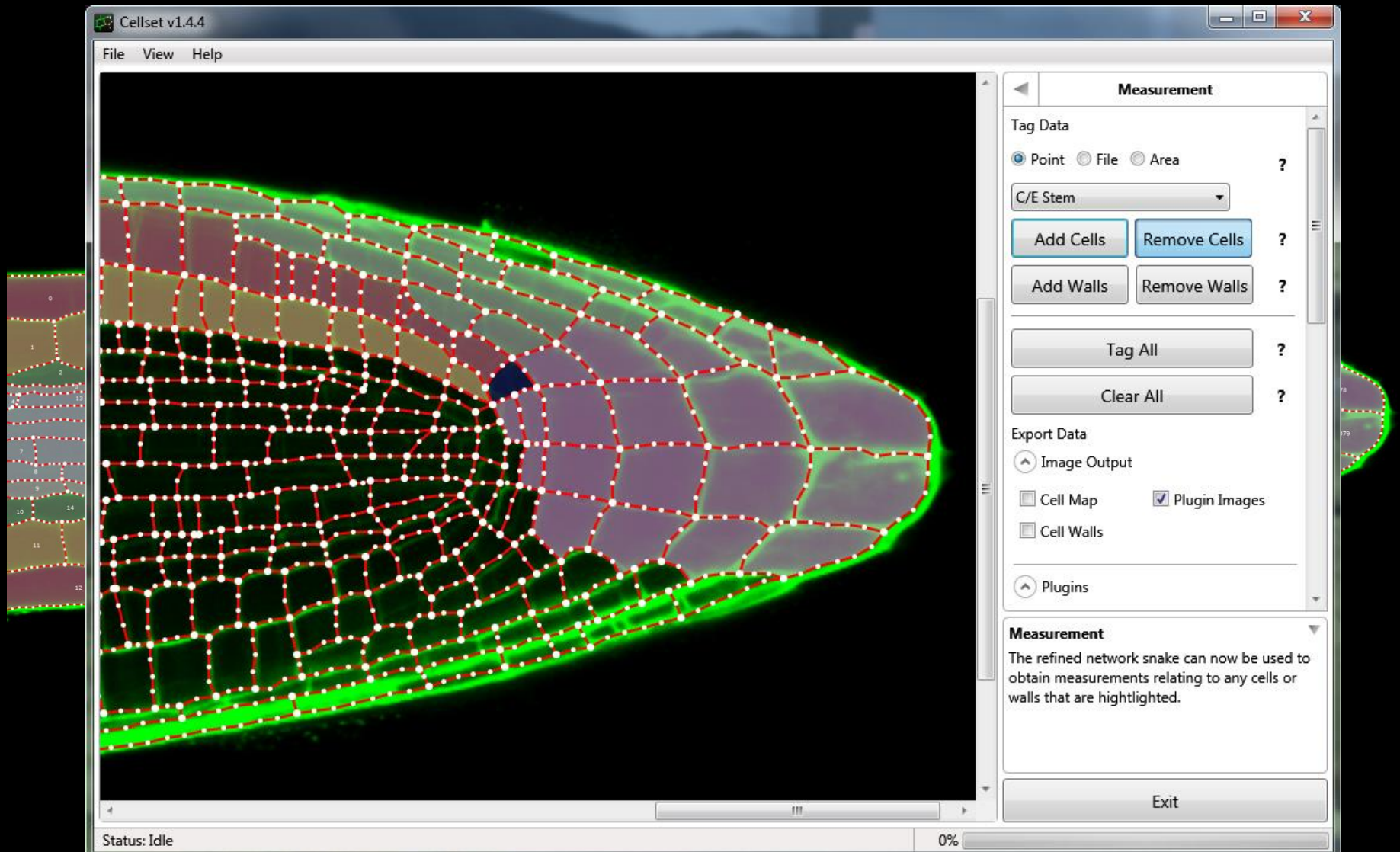
Requirements:

- extraction of cell-level fluorescence data:
  - to map root geometry
  - to quantify sensor fluorescence
- Biological parameter estimation to populate model with carriers etc.

Phenotyping pipeline



CellSeT - Tagging and measurement

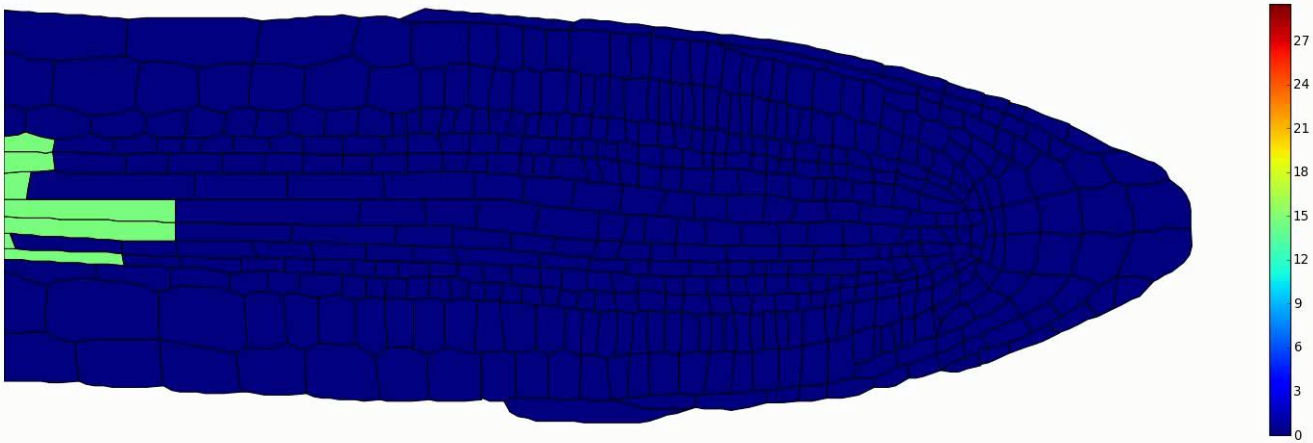


Modelling

- Extracted geometries read into a vertex-based model, based on the OpenAlea modelling framework.
- carrier distributions prescribed
- ODEs for:
  - diffusion of protonated auxin across cell membranes
  - carrier-mediated auxin transport
  - passive diffusion of auxin within the cell wall
  - degradation of biosensor fluorescence via a parameterised network model

http://openalea.gforge.inria.fr

In silico simulation of the hormone fluxes through the root tissue, using segmented cell geometries:



Visualisation/interaction via simulator website

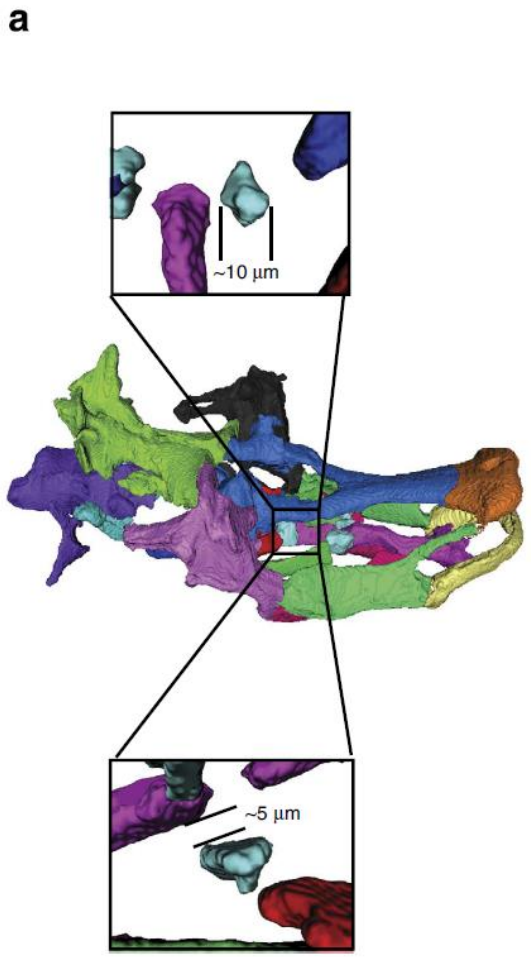
http://www.simuplant.org/

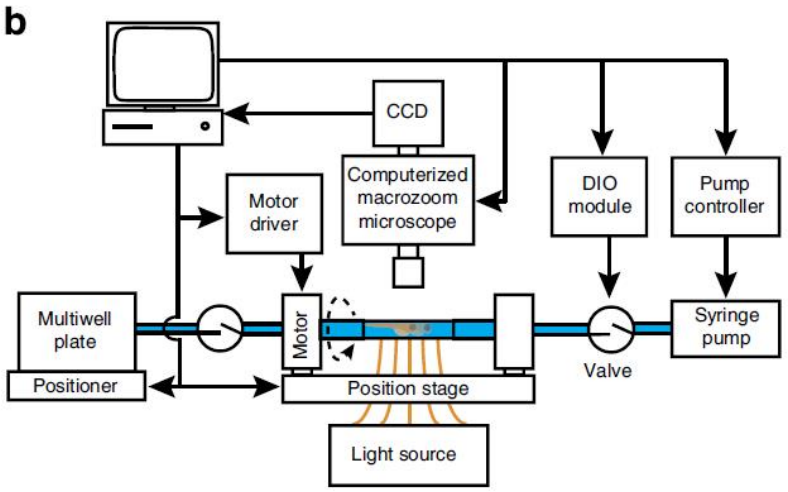
Band et al. *(2014)*

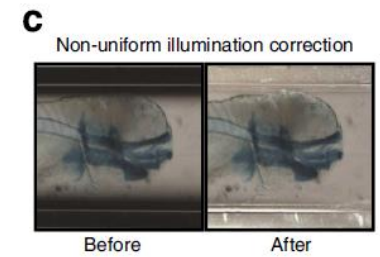
Organ-scale phenotyping

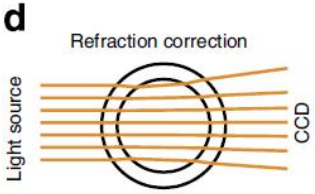
- Zebrafish anatomy
  - High throughput (18 seconds/larva)
  - Optical projection tomography (320 images)
  - 3D reconstruction of entire larvae
  - Craniofacial cartilage (dyed)
    - 200 independent morphological measurements
    - Used in studies of teratogen action

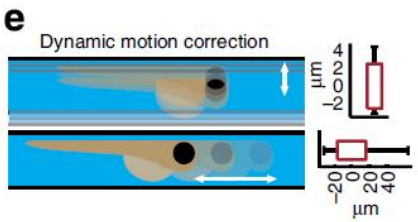
Organ-scale phenotyping

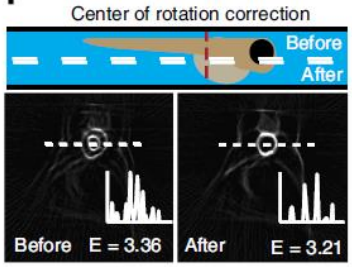


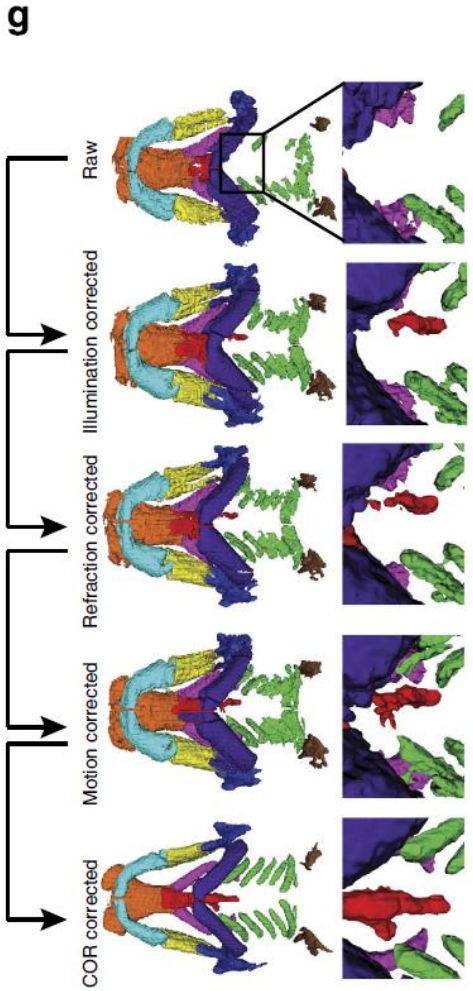






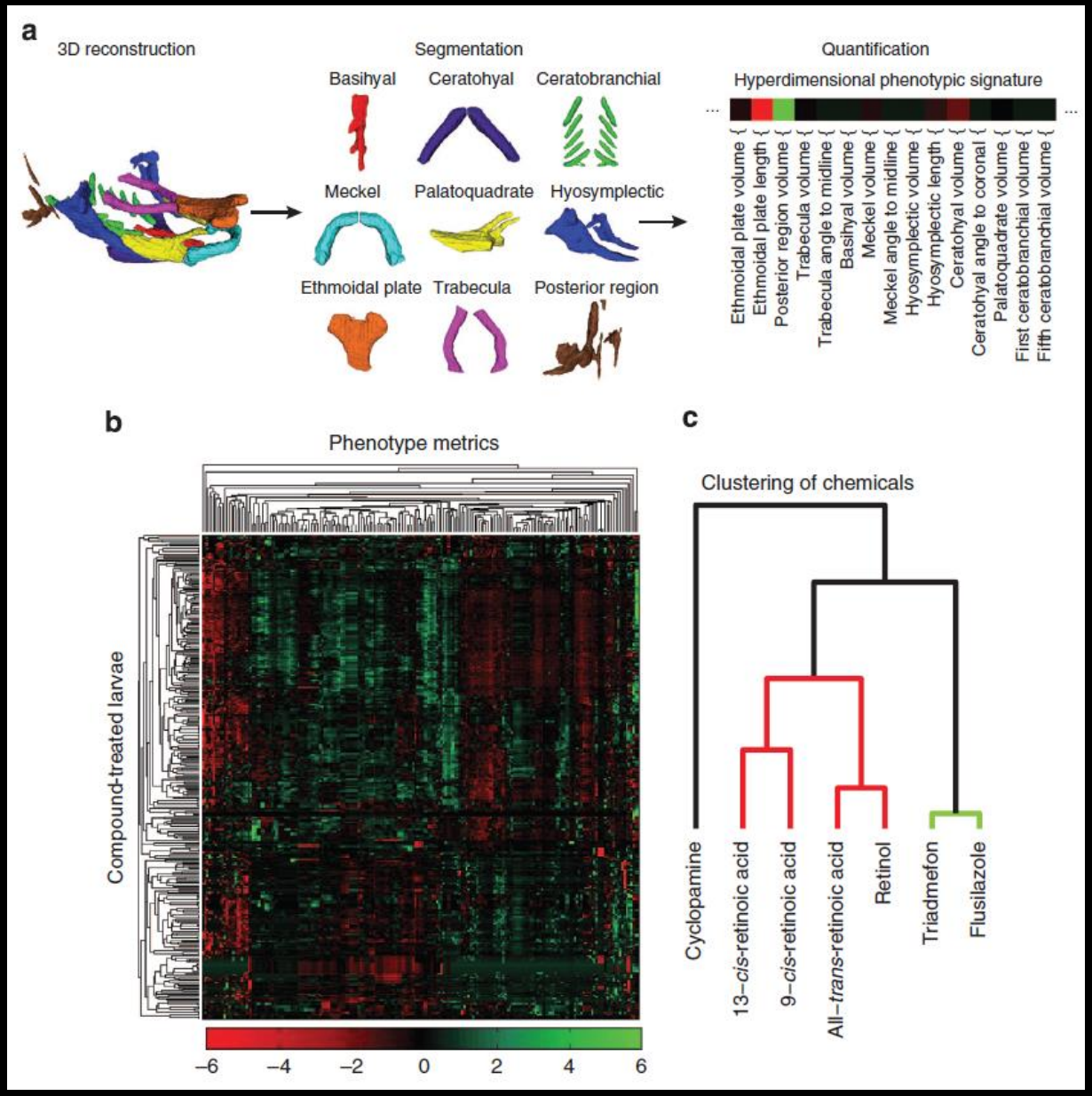






Pardo-Martin et al. *(2013)*

Results – teratogen studies



Organism-level phenotyping

- Root systems architecture may represent untapped genetic resource to improve modern crops
- Screen mapping populations for desirable root traits (e.g.: deep for water and nitrate, shallow for phosphorus)

Wheat root phenotyping – mapping populations

Savannah

- Group 4
- Feed wheat
- Very High Yield

X

Rialto

- Group 1
- Bread wheat
- High quality

- 132 doubled haploid lines

- All genotyped using iSelect 80k SNP array – publically available maps for 44k of those SNPs

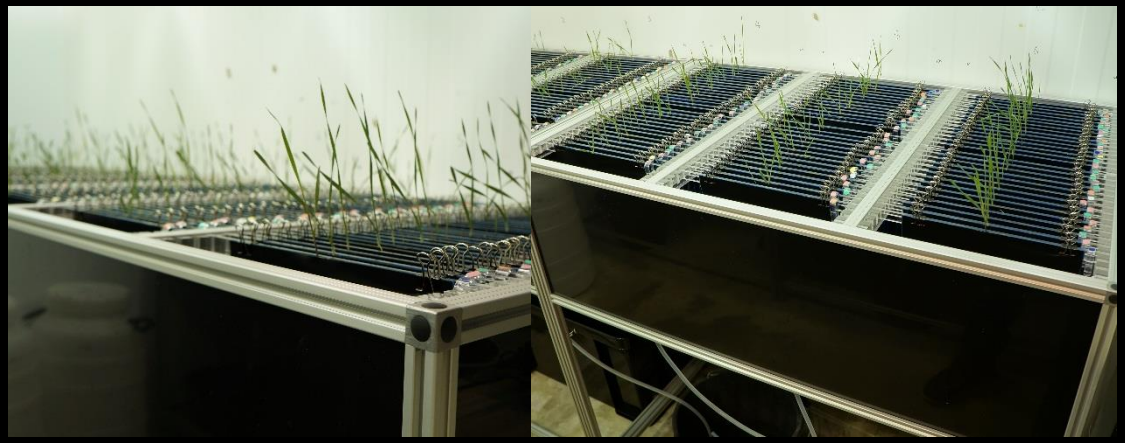
- **96 lines (20 replicates) phenotyped using a 2D-imaging pipeline**

Limagrain
from earth to life

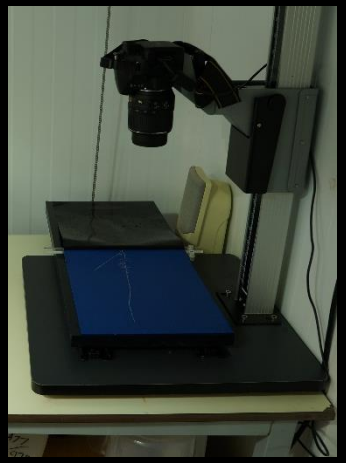
2D Phenotyping pipeline

4 components of the 2D root phenotyping pipeline…

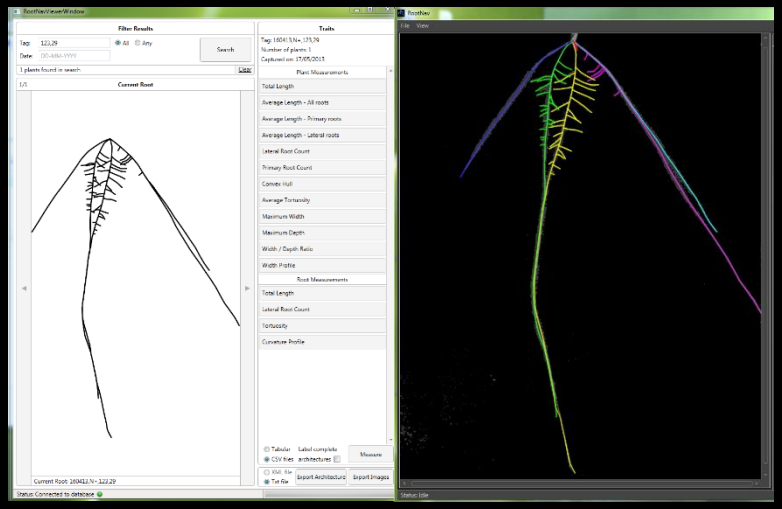
1. Plant growth system



2. Image capture



3. Image analysis & trait quantification



4. QTL analysis

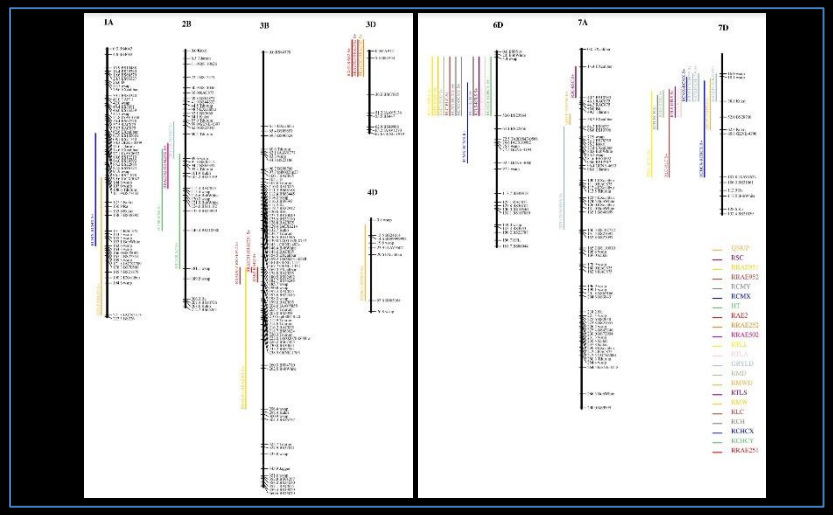
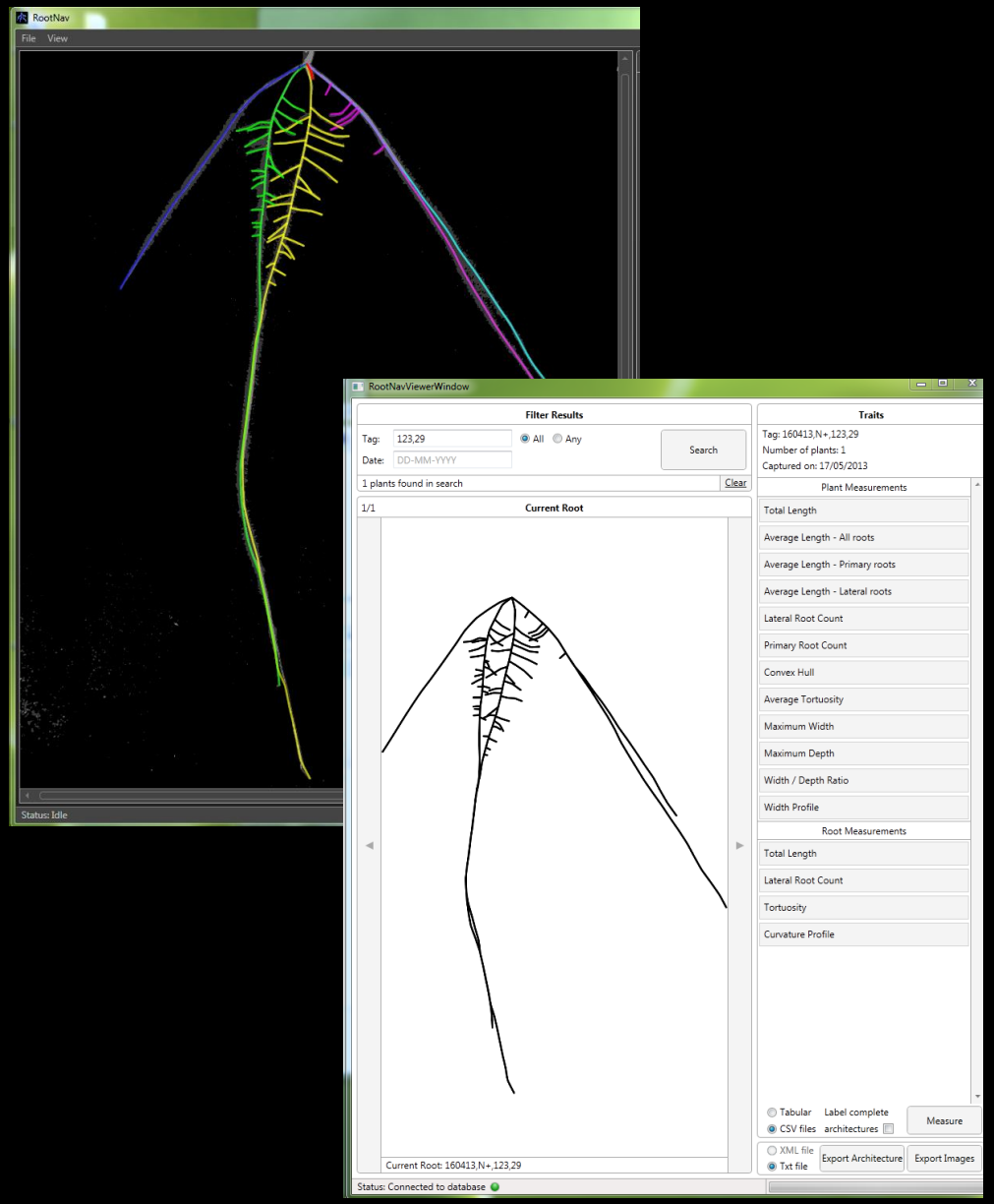
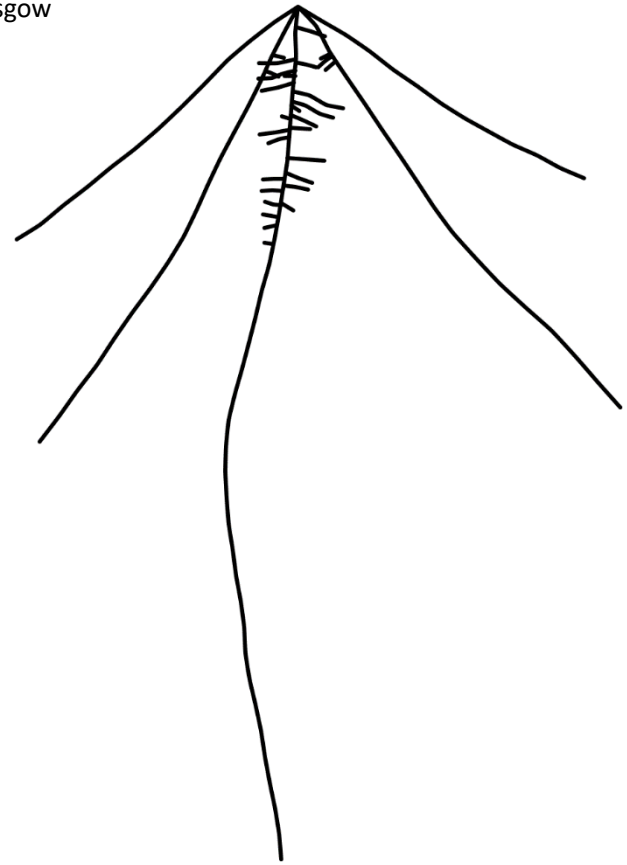


Image analysis - RootNav

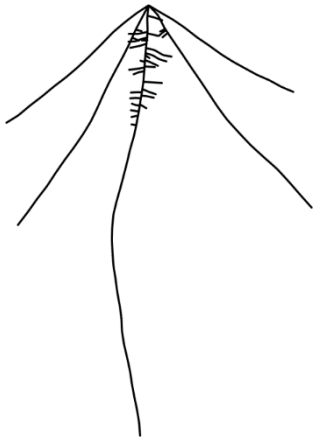
- Semi-automatic analysis and quantification of RSA
- 30s – 2 minutes per image
- Spline data stored on RootNav server
- Root data can be queried using the viewer tool.
- Traits quantified and exported via RSML



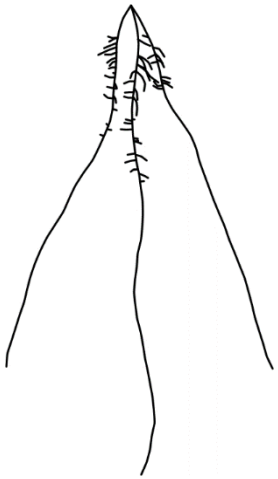
Winter Wheat var:
Glasgow



Glasgow

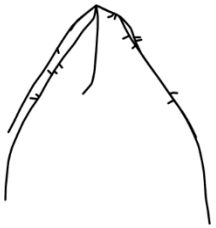


Ae. buncialis



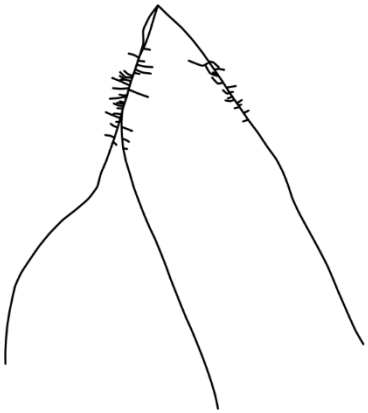
Ae. uniaristata

T. dicocchodies



Ae. variablis

Ae. genticulata



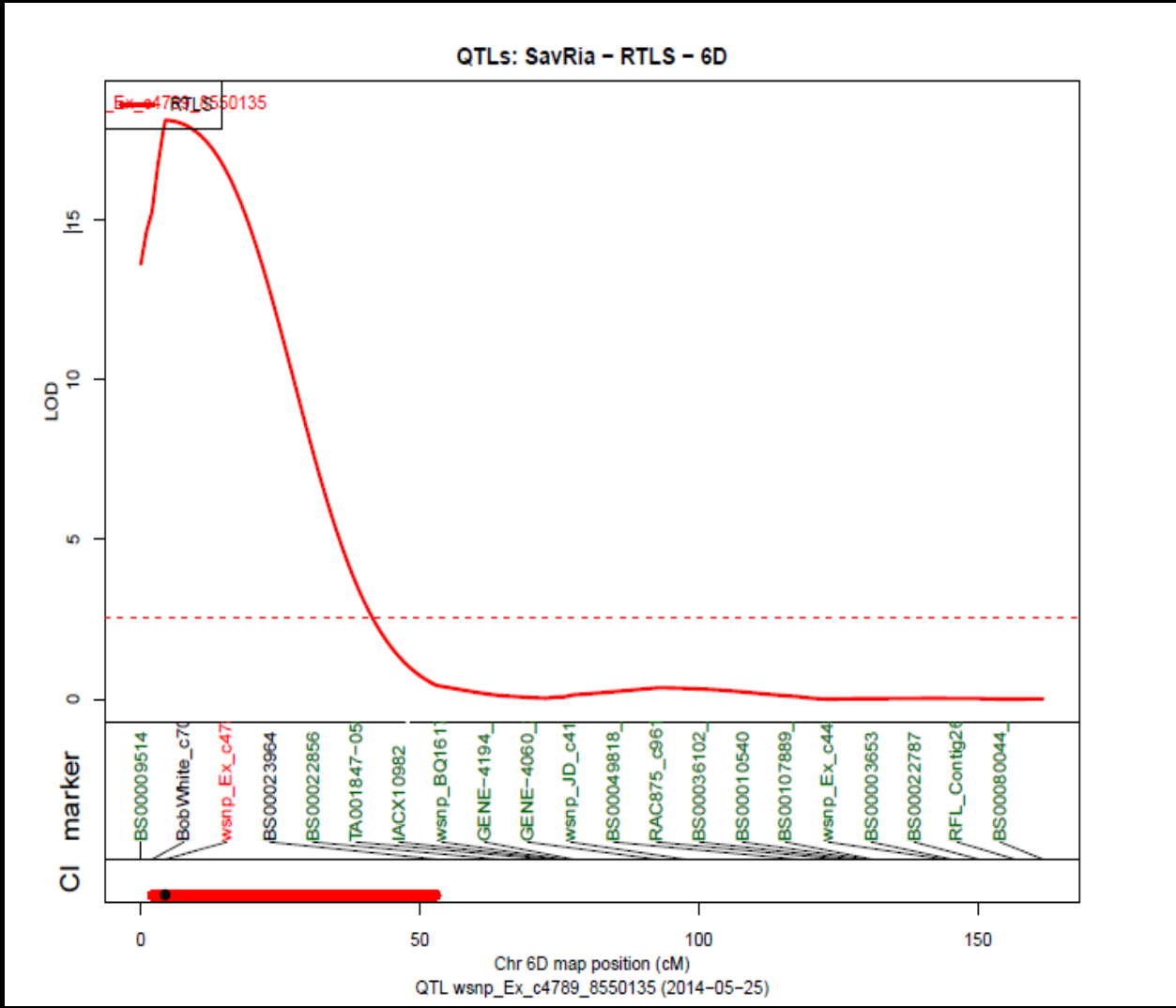
Ae. markgafii

Ae. columnaris

T. urartu

Ae. peregrina

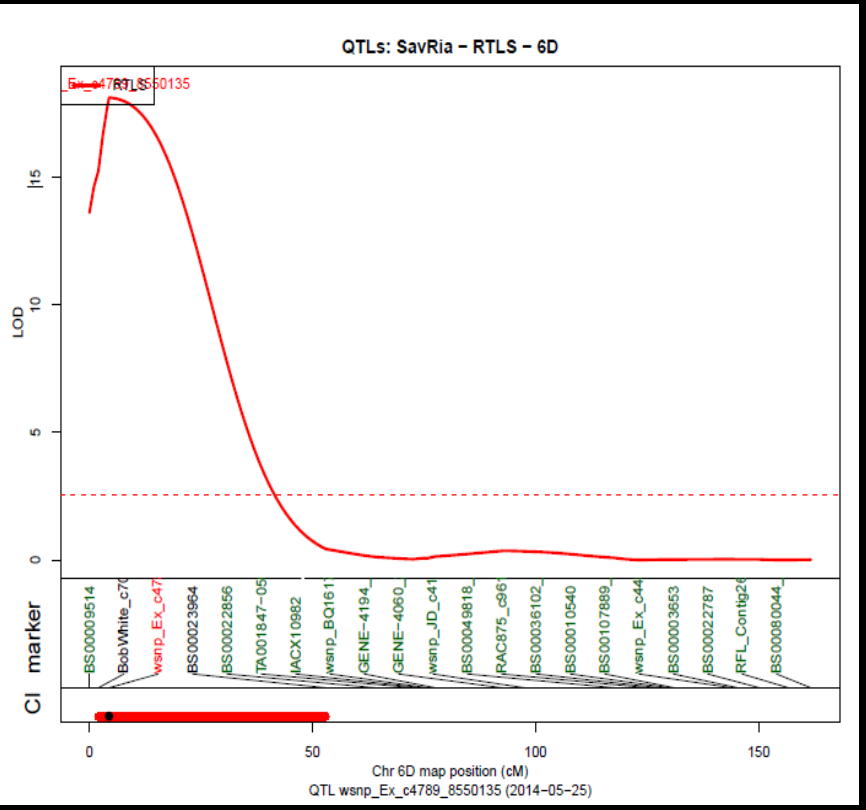
QTL Analysis Results



Indicates the presence of a major effect gene regulating seedling root architecture/vigour

Atkinson et al. *(2015)*

2D Seedling Root Phenotyping



RNAseq

Collaboration with Laura Gardiner (Earlham)

18 candidate genes

NILs

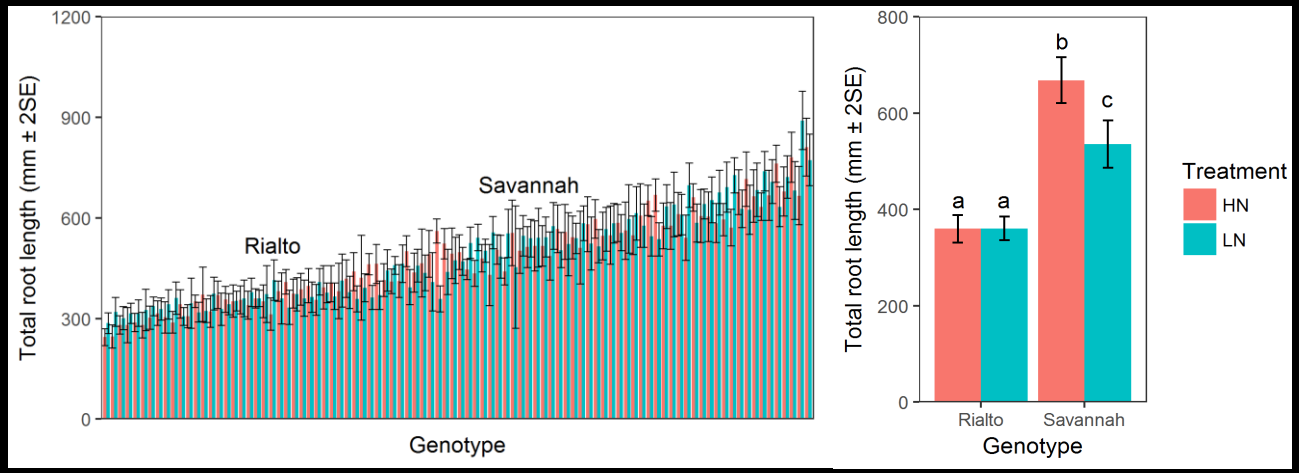
Collaboration with Limagrain

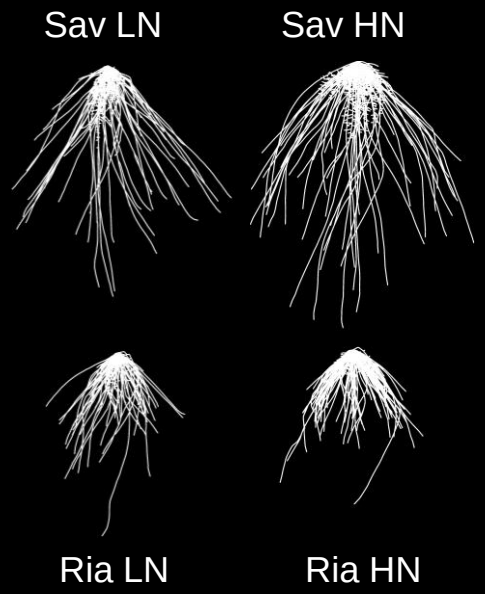
- Currently phenotyping
- Being evaluated in the field
- 18 cM introgression which confers the phenotype

Jonathan Atkinson *(unpublished)*

2D Seedling Root Phenotyping

Variable N (3.8 mM vs 1 mM)





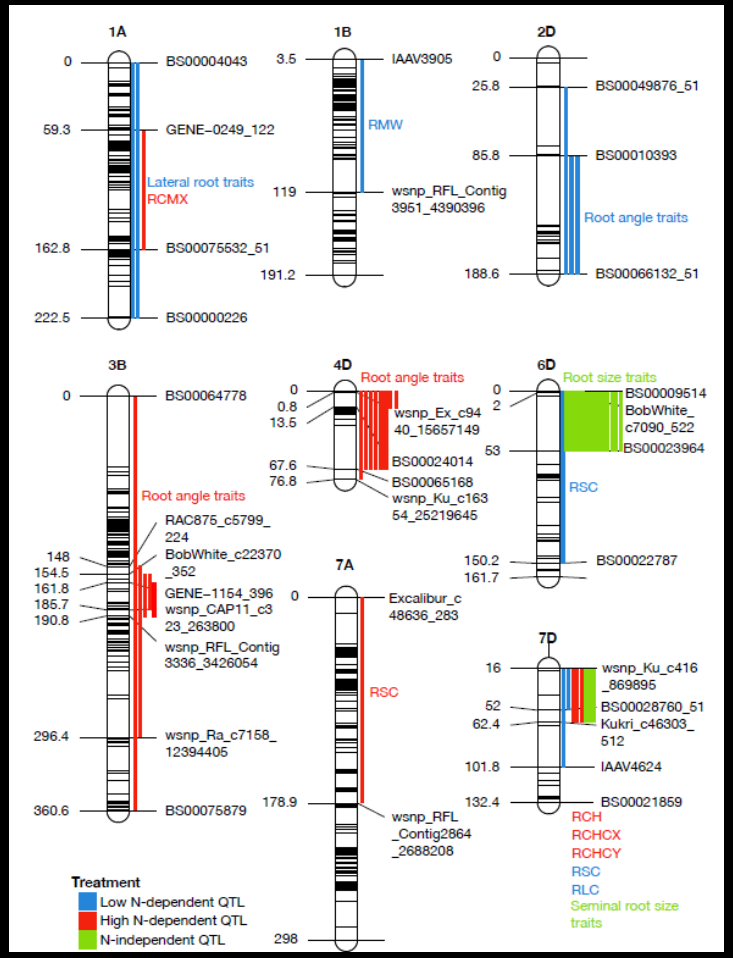
Marcus Griffiths *(unpublished)*

2D Seedling Root Phenotyping

QTL for response to variable N (3.8 mM vs 1 mM)

Low-, high- and N-independent QTL discovered

RNA-Seq for QTL on Chr 2D



Marcus Griffiths *(unpublished)*

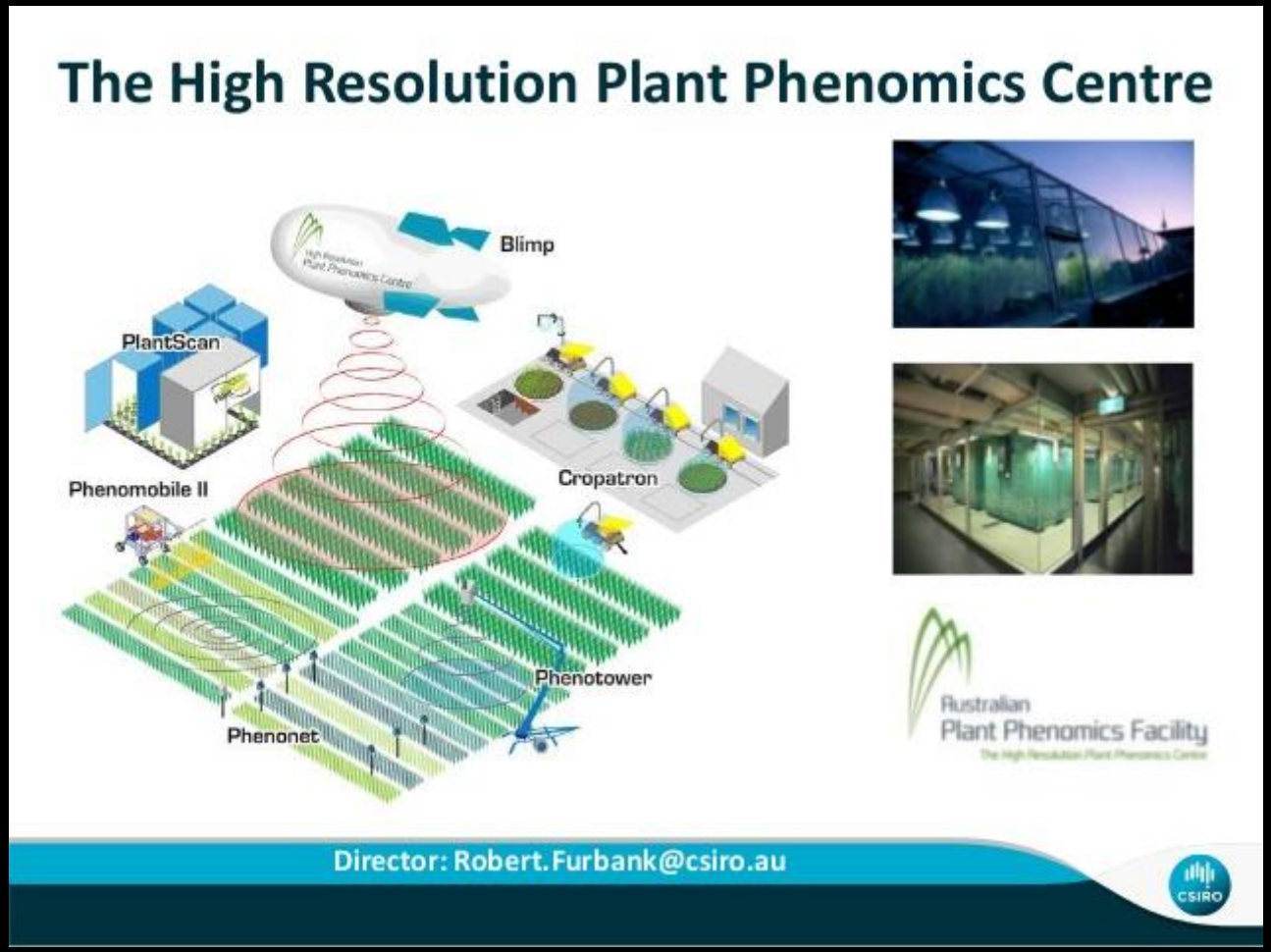
2D Seedling Root Phenotyping

RNA-seq for RRAE251/1001 QTL on chromosome 2D, 3299 differentially expressed genes were identified between wheat lines with QTL (Group A) and without QTL (Group B)

Gene	Log_2 fold change	Adjusted p value	Annotation	Functional annotation
MSTRG.39093	1.73	0.001733593	ref_gene_id_"TRIAE_CS42_2DL_TGACv1_158120_AA0509790"	Peroxidase
TRIAE_CS42_2DL_TGACv1_158042_AA0507110	1.45	0.013427885	ref_gene_id TRIAE_CS42_2DL_TGACv1_158042_AA0507110	Cysteine/Histidine-rich C1 domain family protein
MSTRG.42598	1.31	0.040518291	TGACv1_scaffold_178694_2DS	Unknown
TRIAE_CS42_2DL_TGACv1_158637_AA0523690	1.29	0.036921878	ref_gene_id_"TRIAE_CS42_2DL_TGACv1_158637_AA0523690";	P-loop containing nucleoside triphosphate hydrolases family protein
MSTRG.40726	2.12	2.52E-05	TGACv1_scaffold_160426_2DL	Unknown
TRIAE_CS42_2DS_TGACv1_178283_AA0593780	2.21	9.50E-06	ref_gene_id_"TRIAE_CS42_2DS_TGACv1_178283_AA0593780";	Unknown
MSTRG.40576	1.53	0.007831497	ref_gene_id_"TRIAE_CS42_2DL_TGACv1_160112_AA0546950"	Wound-responsive family protein
MSTRG.41621	1.29	0.036320236	ref_gene_id_"TRIAE_CS42_2DS_TGACv1_177373_AA0574940"	Zinc-finger protein
MSTRG.41900	1.36	0.035609902	ref_gene_id_"TRIAE_CS42_2DS_TGACv1_177679_AA0582290";	Lectin-domain containing receptor kinase A4.3
MSTRG.40281	1.44	0.012862779	ref_gene_id_"TRIAE_CS42_2DL_TGACv1_159581_AA0540110";	Heavy metal transport/detoxification superfamily protein
MSTRG.40366	2.02	8.89E-05	TGACv1_scaffold_159729_2DL	Unknown
MSTRG.41870	1.38	0.025822143	ref_gene_id_"TRIAE_CS42_2DS_TGACv1_177631_AA0581600"	Peroxidase
MSTRG.41588	1.66	0.002127833	TGACv1_scaffold_177335_2DS	Zinc finger BED domain-containing protein RICESLEEPER 2-like
MSTRG.39748	1.66	0.001812471	ref_gene_id_"TRIAE_CS42_2DL_TGACv1_158826_AA0526780";	Basic helix-loop-helix (BHLH) Transcription Factor
MSTRG.39185	1.41	0.025442551	ref_gene_id_"TRIAE_CS42_2DL_TGACv1_158211_AA0512570"	Peroxidase
MSTRG.41090	1.48	0.013212786	TGACv1_scaffold_161672_2DL	Unknown
MSTRG.40833	1.88	0.000358374	ref_gene_id_"TRIAE_CS42_2DL_TGACv1_160694_AA0553200"	Nitrate transporter 1.2

Marcus Griffiths *(unpublished)*

Population scale: field phenotyping



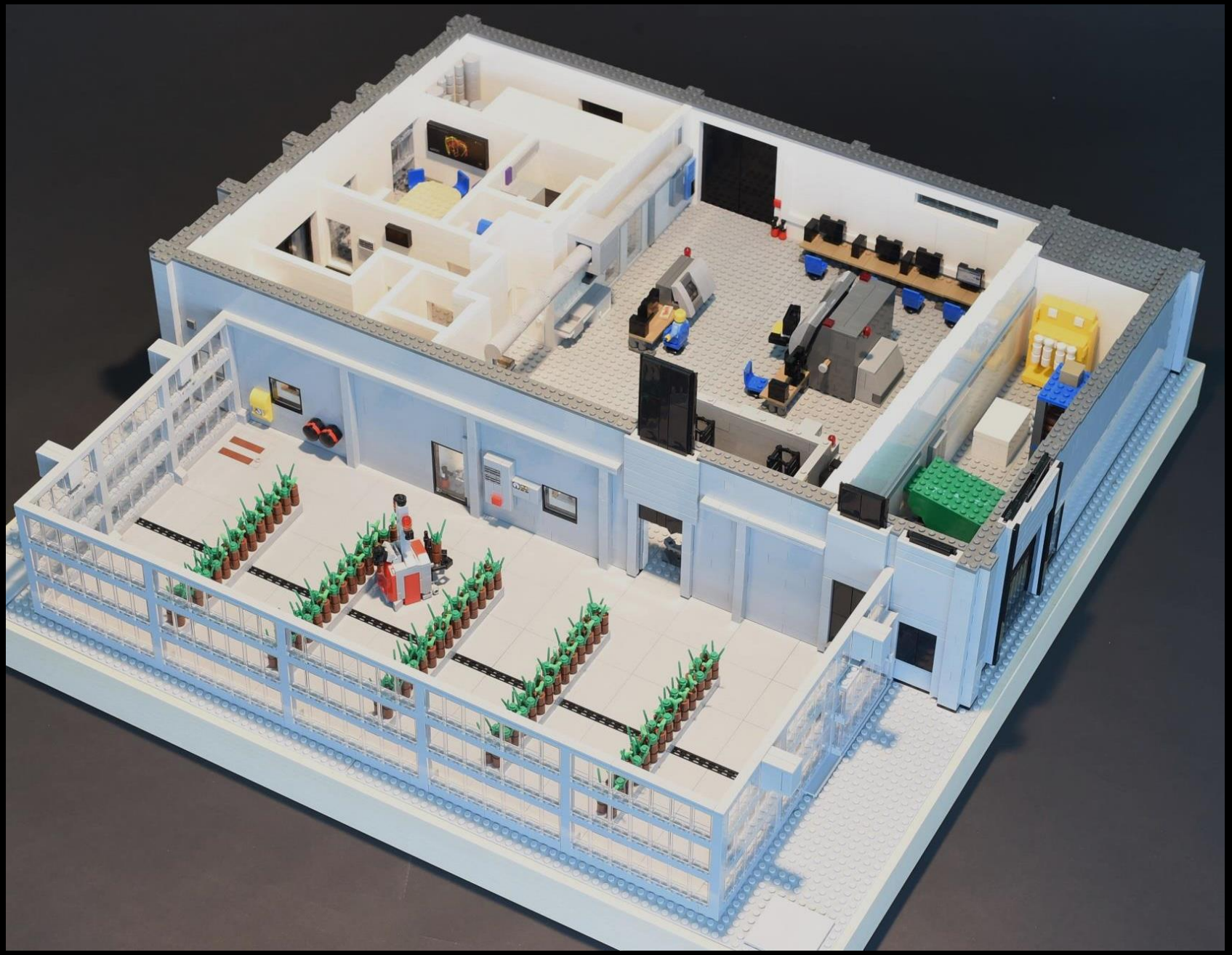
Population scale: field phenotyping

https://www.youtube.com/watch?v=Wj-U0QH5J_M

Case study: developing a phenotyping facility at UoN

- Rationale – micro-computed X-ray tomographic (µCT) scanning allows imaging of root systems in soil.

- Use of mesocosms of realistic dimensions at reasonable throughput presents many technical challenges



Hounsfield Facility

Mesocosm preparation: standardising soils

Sample handling

- Glasshouse capacity 140 columns
- Column dimensions 1 m x 25-30 cm OD
- Weight ~60-90 kg
- Minimal disturbance (reduce growth effects, settling of soil etc.)
- 24/7 automated operation

- Laser guided vehicle (LGV)
  - autonomous robot


ocme

Scanning

Modified large scanner

high power 320kV minifocus X-ray tube

heavy duty manipulator stage

high contrast digital detector

Volumes automatically stitched and reconstructed on acquisition

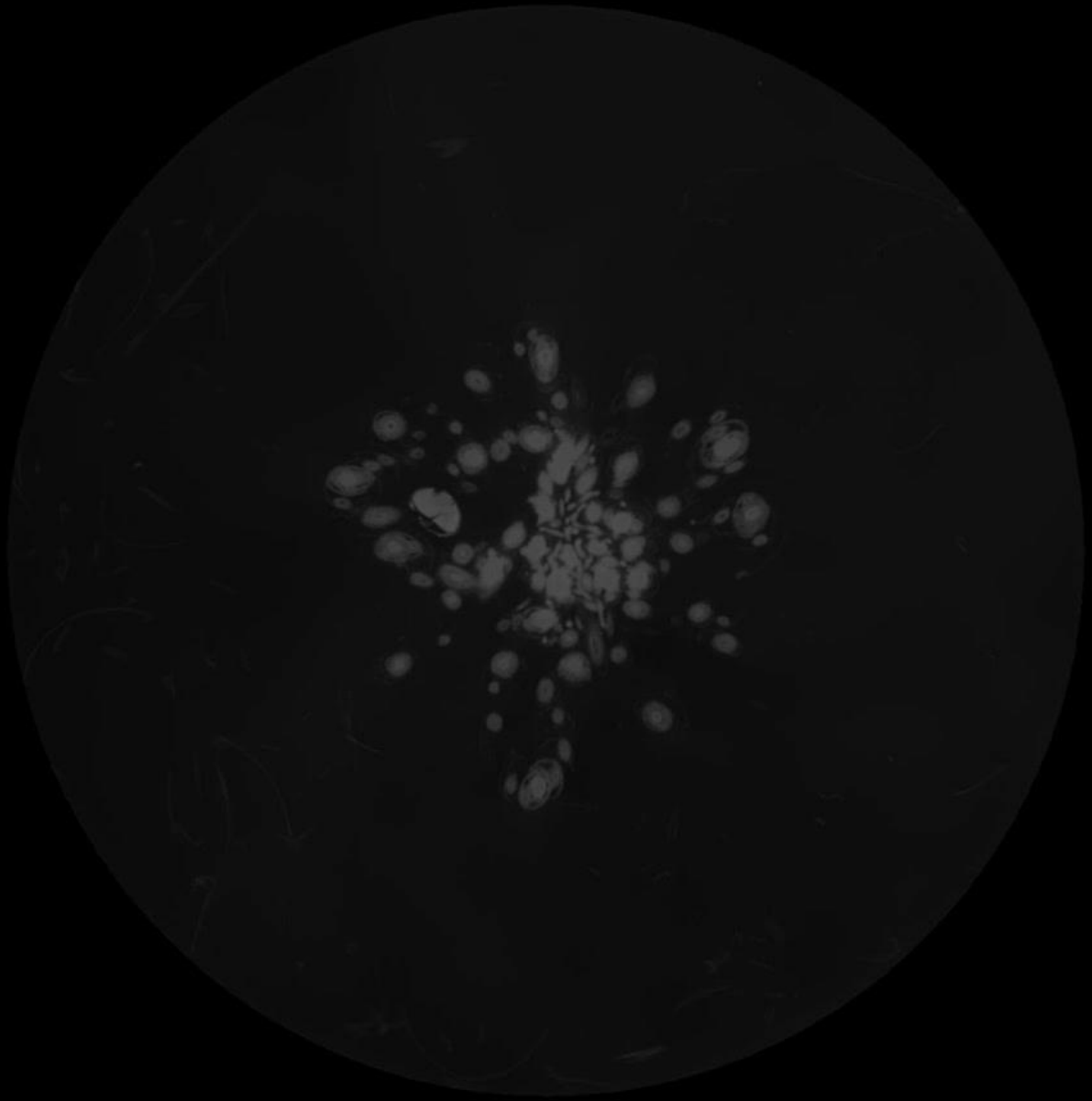
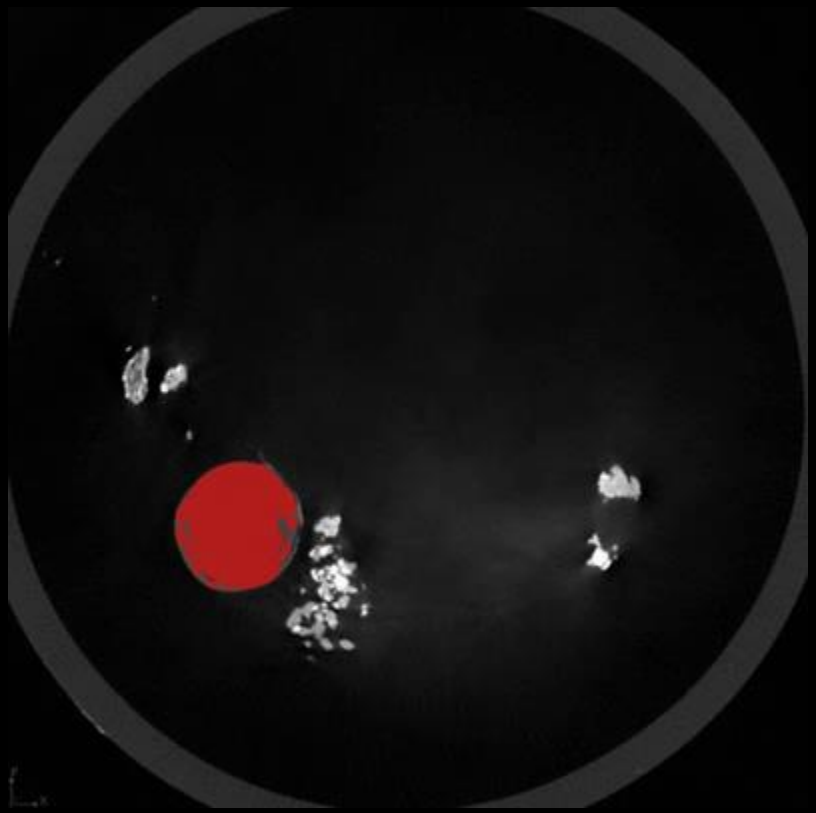
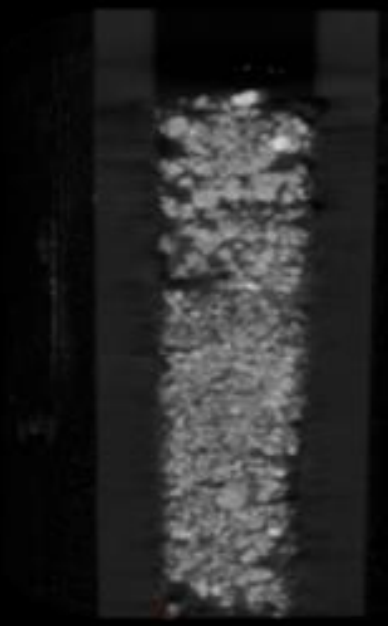


Image analysis- RooTrak





Maize in sandy loam, resolution 30µm

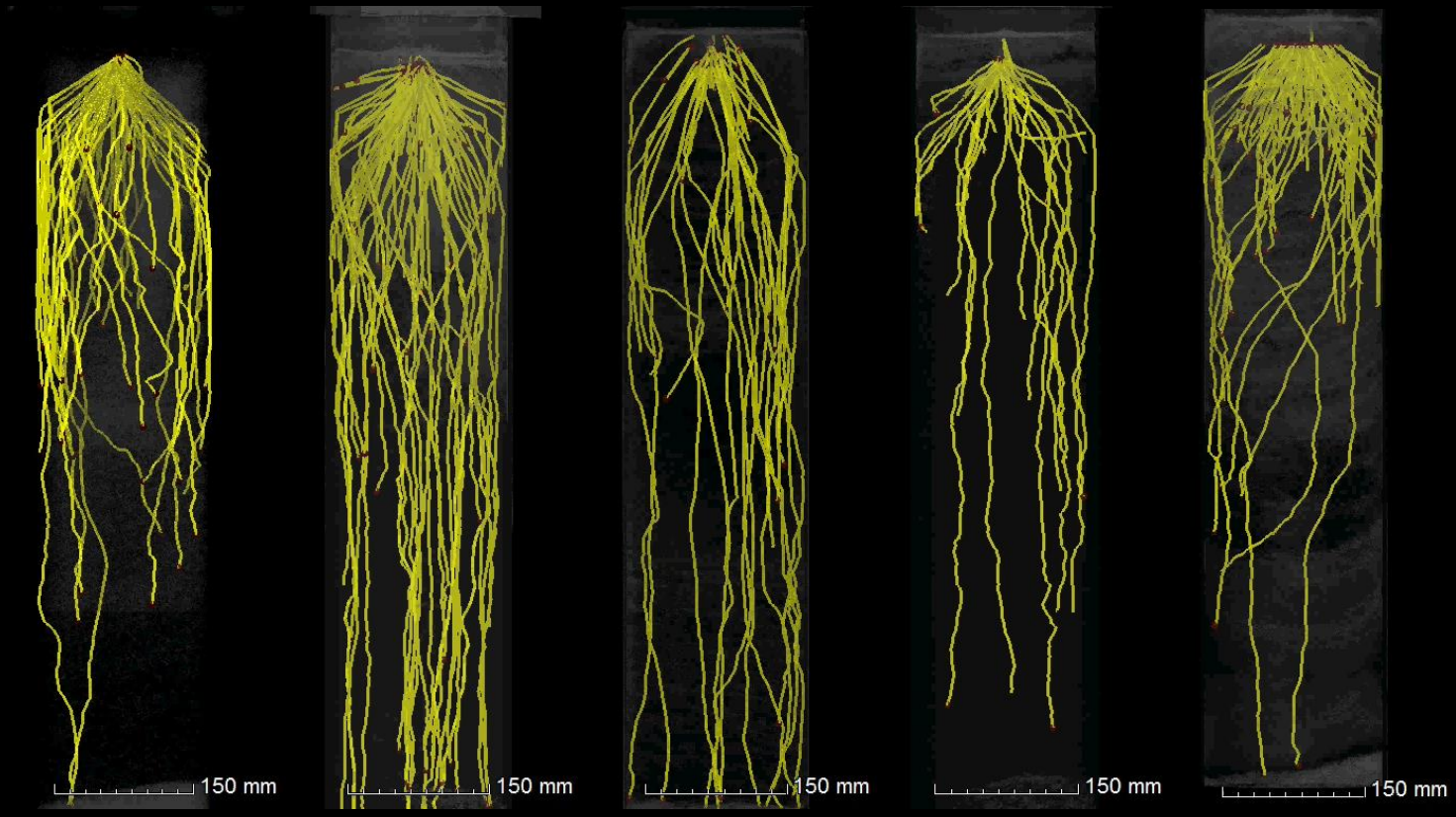
Mairhofer et al. *(2012); Mairhofer* et al. *(2013)*

X-ray Computed Tomography

Ancient relatives of wheat

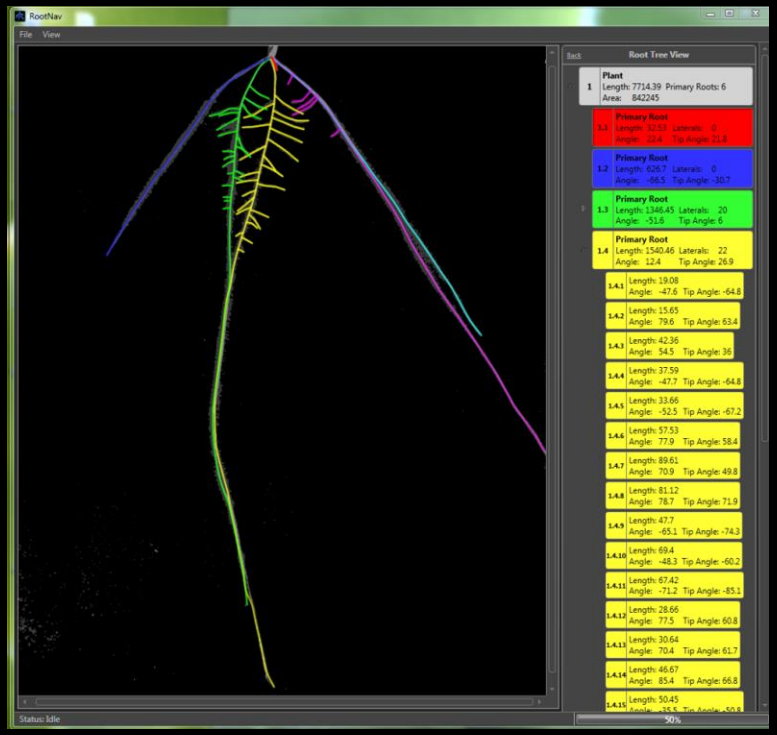
- Variation for useful agricultural traits such as more roots at depth
- Select WISP/DFW introgression panels from the Ian and Julie King which may have beneficial root architecture traits

Atkinson & Atkinson, unpublished



T. urartu *Ae. speltodies* *T. dicoccoides* *Ae. tauschii* *T. aestivum*

Unblocking the image analysis bottleneck



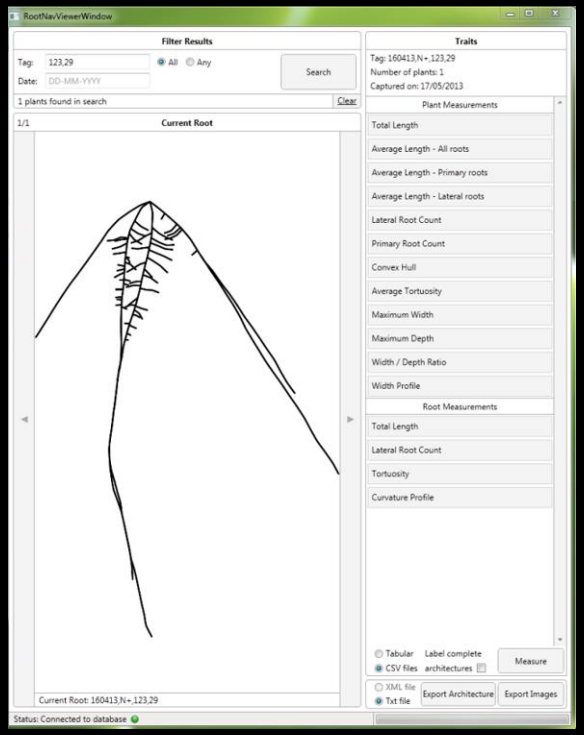
In *Arabidopis*, image analysis unblocked the phenotyping bottleneck.

In larger, more complex plants – image analysis is again a limiting step.

Can new techniques help?

Machine learning

Unblocking the image analysis bottleneck



Machine learning for image analysis

1. 'Traditional' machine learning
   - PRIMAL - Random Forest

2. Deep learning
   - Convolutional neural networks (CNNs)

Machine learning

Chr	Trait	Manual (RootNav)	Automatic (RiaJ)	Primal (600 images)
4D	W/D ratio	2.7	2.71	2.5
6D	Seminal count			3.3
	Total root length	24	17	16.0
	Mean seminal length	22.2		14.0
	Lateral count	9.1		17.0
	Total lateral length	6.4		12.6
	Total seminal length	25.6		15.2
	Width	6.4	13.5	13.1
	Depth	22.7	13.6	15.0
	W/D ratio			2.2
7A	Seminal number	2.1		
7D	Lateral number	2.4		5.0
	Seminal number			3.4
	Total lateral length	2		4.2
	Total root length	9	4.1	3.1
	Total seminal length	9.7		2.8

- Requires around 600 training images to be analyzed to achieve an R^2 of ~0.9
- 12/13 QTL discovered using PRIMAL vs RootNav
- Does sometimes create false positives with low LOD scores, but these often co-localise with other 'real' QTL

Atkinson *et al.*, 2017. Gigascience

Lobet *et al.*, 2017. Front Plant Sci.

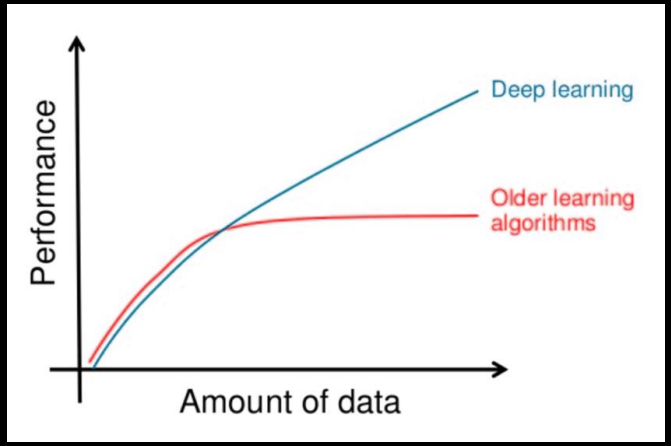
Deep machine learning

Relies on training a network using a large number of annotated images

- The more training data you use, the better it becomes

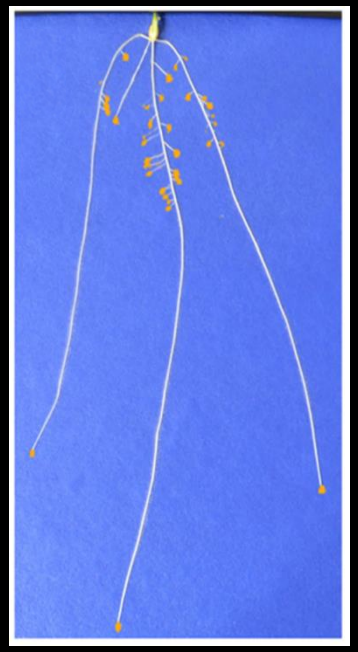
Does not use pre-computed features

Once trained, the network can annotate new images



Deep machine learning

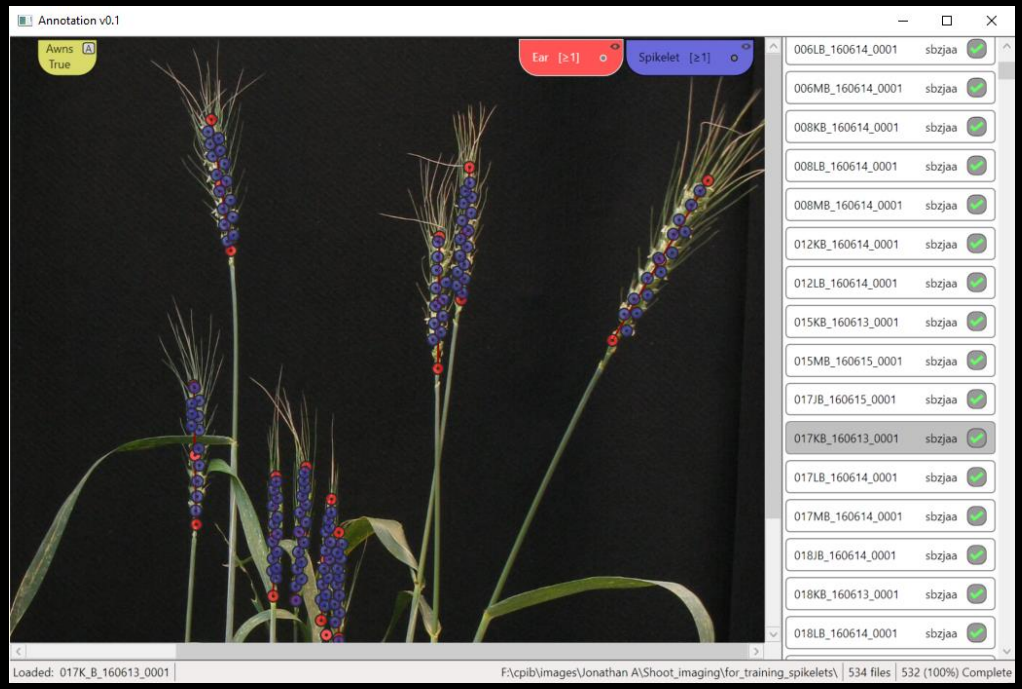
- >97% accuracy in most of the example uses tested to date
- LeMuR: Plant Root Phenotyping via Learned Multi-resolution Image Segmentation (AutoRootNav)



Pound *et al.* (2017)
Pound *et al.* (2017, ICCV)

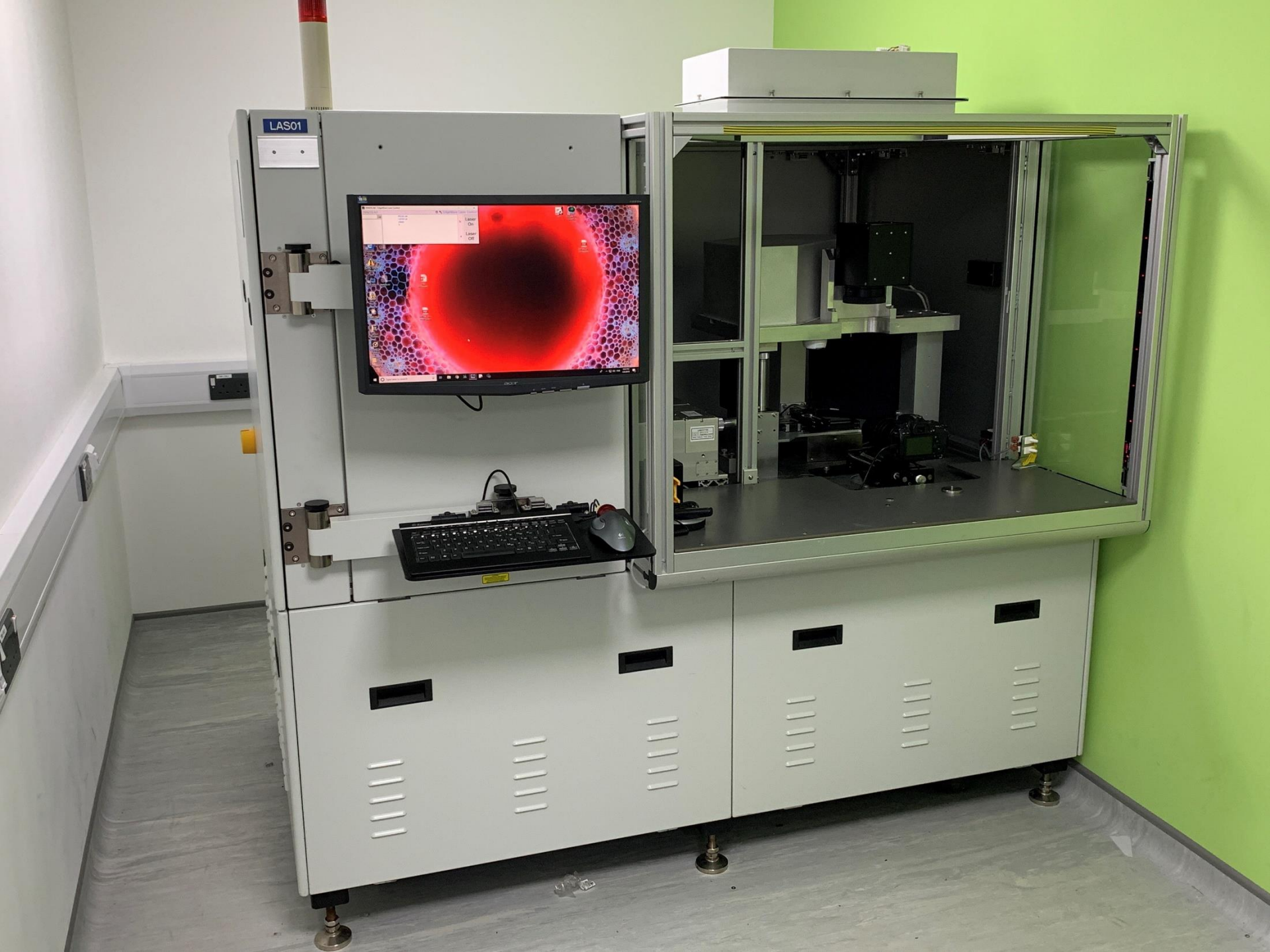
Deep machine learning

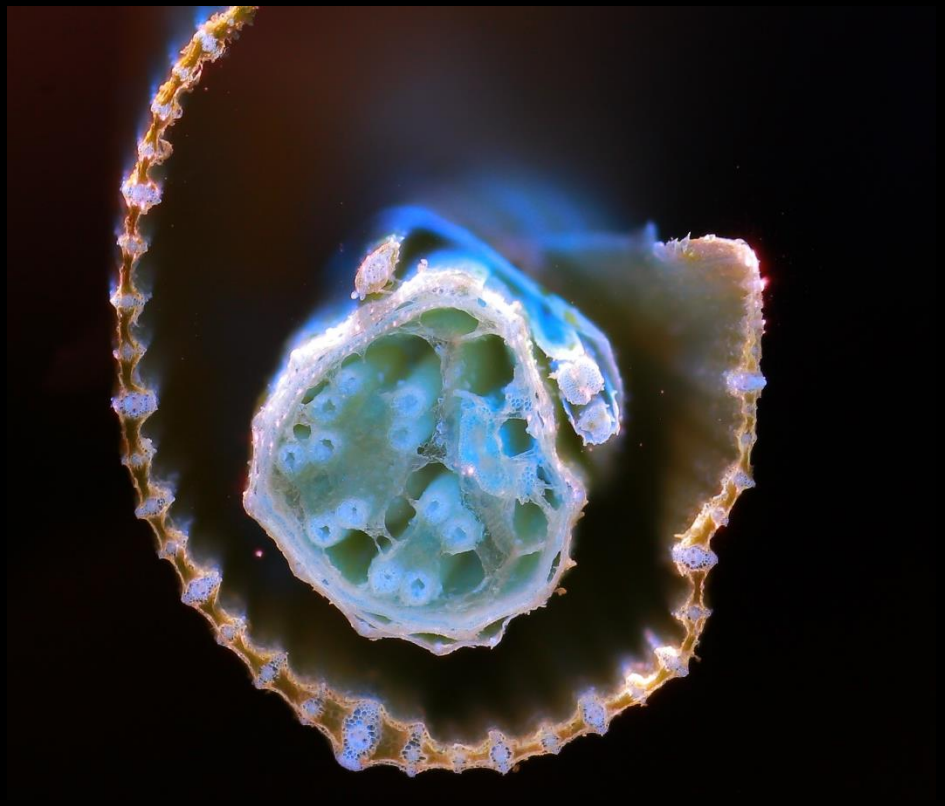
Annotation tool

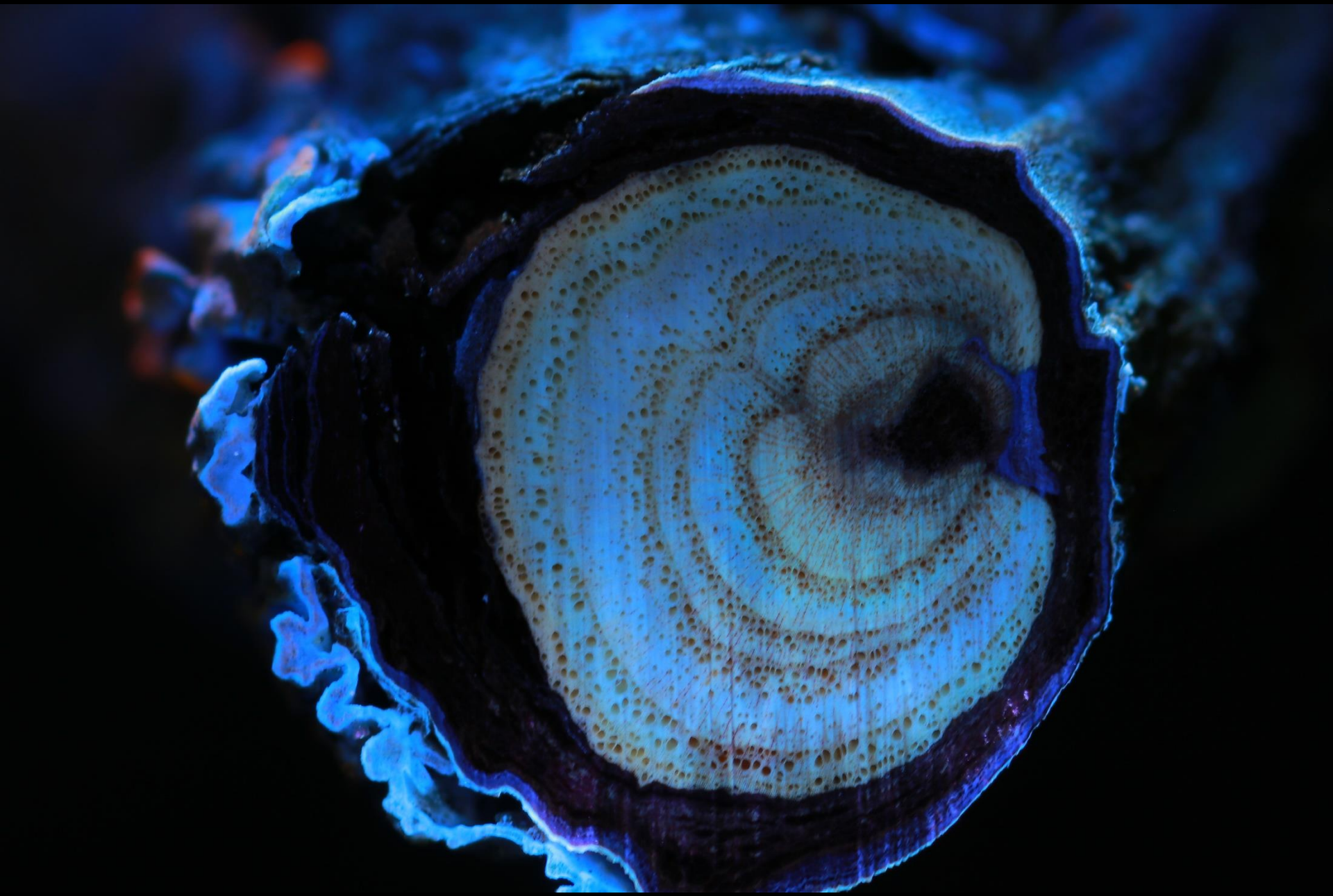


Coming soon:

Laser Ablation Tomography (LAT)


LAS01
Laser On
Laser Off





Summary

- Phenomics aims to bridge the “genotype-phenotype gap”
- Phenomics involves high throughput acquisition and analysis of multi-dimensional data
- Phenomic pipelines utilise multiple disciplines and technologies

References/further reading

Houle D., Govindaraju D.R., Omholt S. (2010) Phenomics: the next challenge. Nature Reviews Genetics 11 (12): 855–66

Furbank RT, Tester M. (2011) Phenomics-technologies to relieve the phenotyping bottleneck. Trends Plant Sci. 16(12):635-44

Tardieu F, Cabrera-Bosquet L, Pridmore T, Bennett M. (2017) Plant Phenomics, From Sensors to Knowledge. Curr Biol. 27:R770–R783

Atkinson JA, Pound MP, Bennett MJ, Wells DM. (2019) Uncovering the hidden half of plants using new advances in root phenotyping
Curr. Opinion Biotech. 55(8):1-8

Databases: http://www.phenomicdb.de/
Resources: http://www.plant-phenomics.ac.uk/en/resources/