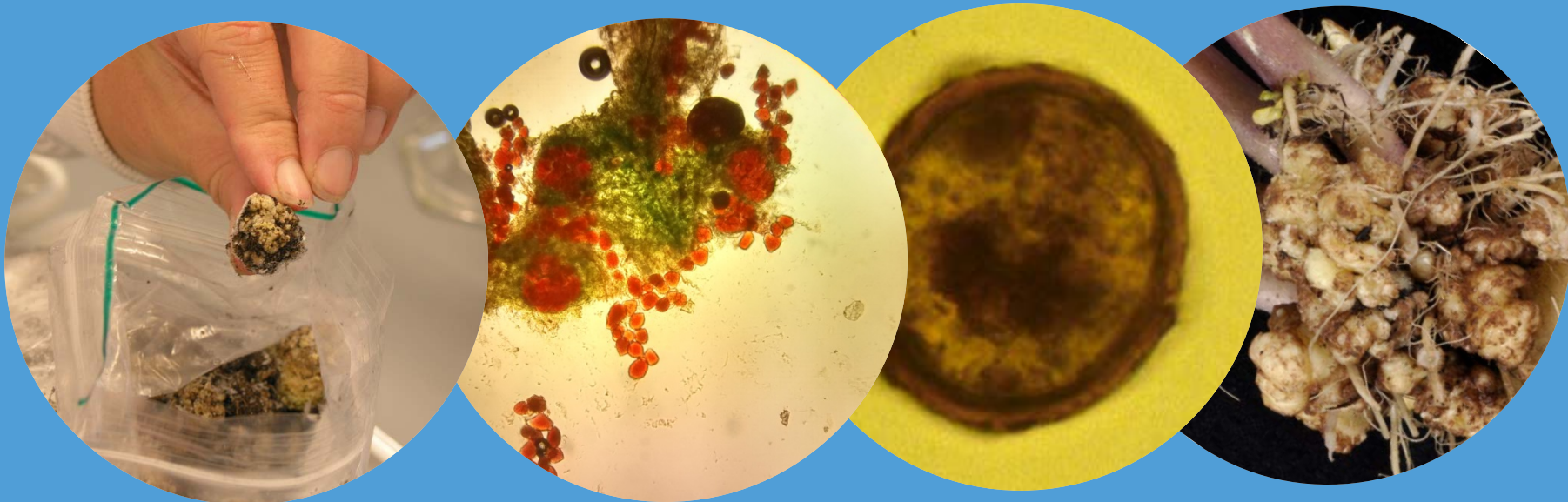# Unravelling the *Synchytrium endobioticum* genome
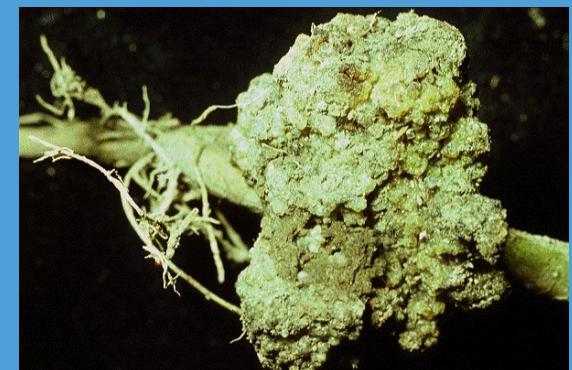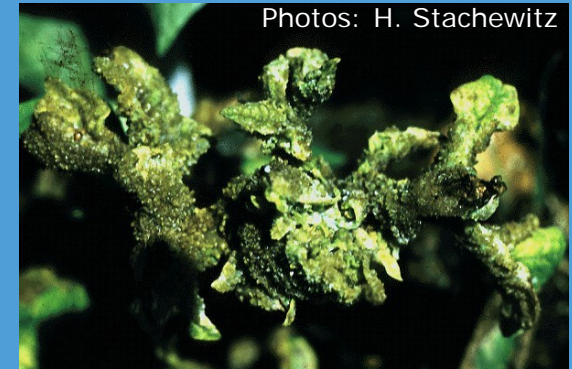
Working towards reliable and rapid molecular pathotype identification

B.T.L.H. van de Vossenberg, M. van Gent-Pelzer, L.V. Bakker, S. Warris, H.C. van de Geest, P. Bonants, C.A. Lévesque, J. Cullis, C.T. Lewis, J.T. Chapados, W. McCormick, K. Dadej, Z. Adam, G.J. Bilodeau, M. Gagnon, D.S. Smith, R.F.G. Visser, J.H. Vossen, T.A.J. van der Lee
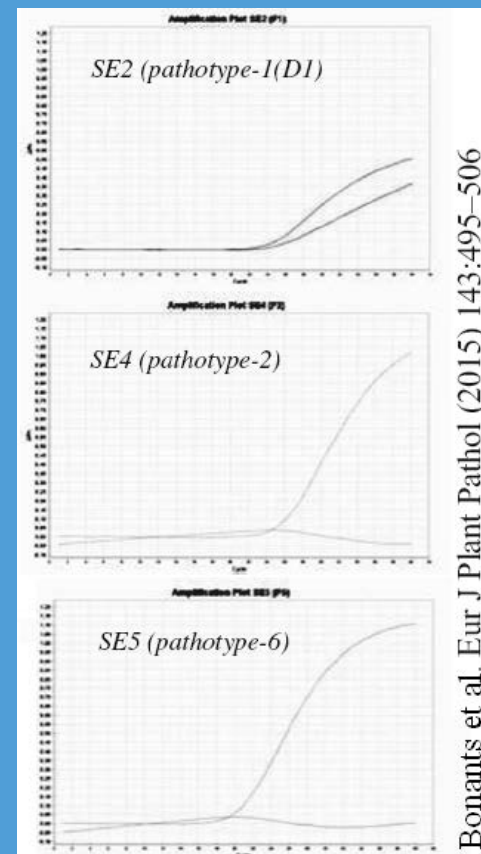
WAGENINGEN UR
For quality of life

Synchytrium endobioticum

Photos: H. Stachewitz

- Soil-borne, obligate parasitic fungus on potato

- Causal agent of potato wart disease

- Induces wart formation upon infection of susceptible potato cultivars

- Production of robust resting spores

- Regarded as one of the most important quarantine disease on cultivated potato

- World-wide quarantine status and included on USA bioterrorism list for plant pathogens

WAGENINGEN **UR**
For quality of life
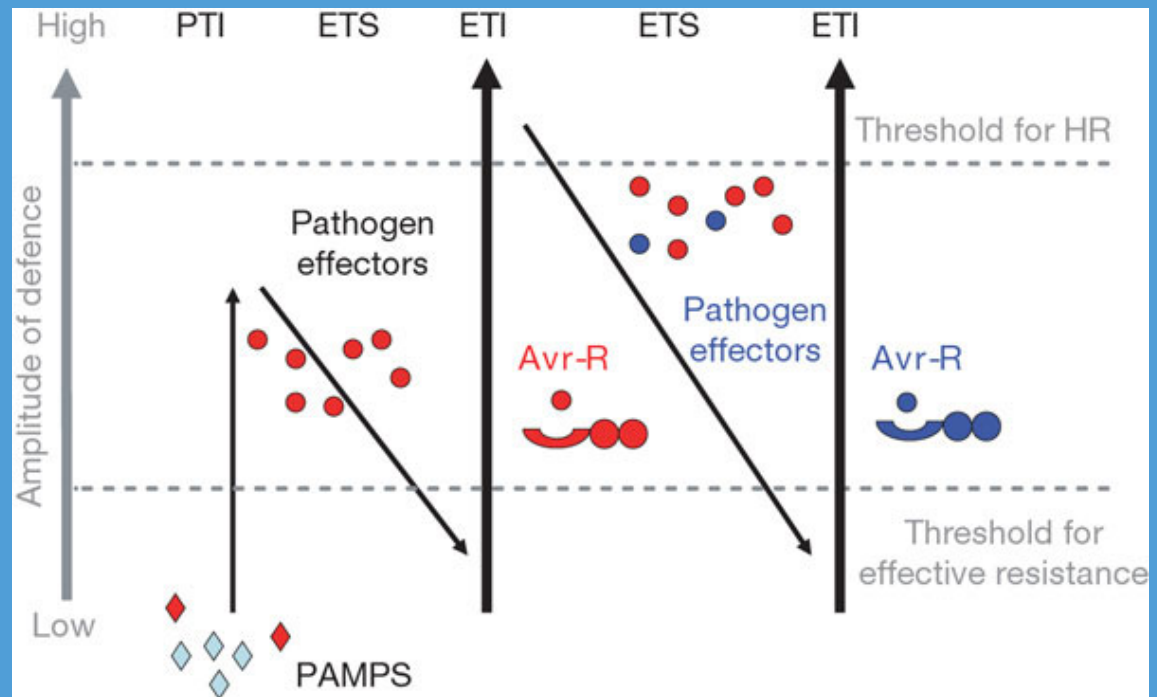
Pathotype identification

- Pathotype identification basis of phytosanitary measures
- More than 30 pathotypes described
- Pathotypes 1, 2, 6 and 18 are most frequently found in Northwest Europe

- Spieckermann and Glynne-Lemmerzahl bioassays
- Time consuming, costly, results not always conclusive no standardised differential set of potato cultivars

- Pathotype identification based on associated SNPs

- Need for trait associated molecular markers

SE2 (pathotype-1(D1))

SE4 (pathotype-2)

SE5 (pathotype-6)

Bonants et al. Eur J Plant Pathol (2015) 143:495–506

Avr and *R* genes: an ongoing arms race

- Plant immune system is triggered by pathogen associated molecular patterns (PAMP) (e.g. chitin for fungi): PTI

- Pathogens produce effectors to suppress PTI (*Avr*): ETS

- Resistant plants respond with ETI (*R*-gene)

Sanchez-Vallet et al. (2014) doi: 10.1093/femsre/fuu003

Jones & Dangl (2006) *Nature* 444, 323-329

Determining the genome of an obligate pathogen

- Non-culturable fungus

- Winter spores (resting spores) are the "purest" form of the fungus we can obtain in sufficient amounts for NGS

S. endobioticum

Potato

Saprobic fungi and bacteria

…

Genome assembly and annotation

```
          ┌─────────────────┐
          │  S. endobioticum │
          │   1(D1) - MB42   │
          │   Winter Spores  │
          └─────────────────┘
```

| RNAseq (Illumina) | 454 (Roche) | PacBio (Pacific Biosciences) | HiSeq (Illumina) |

Filter potato reads
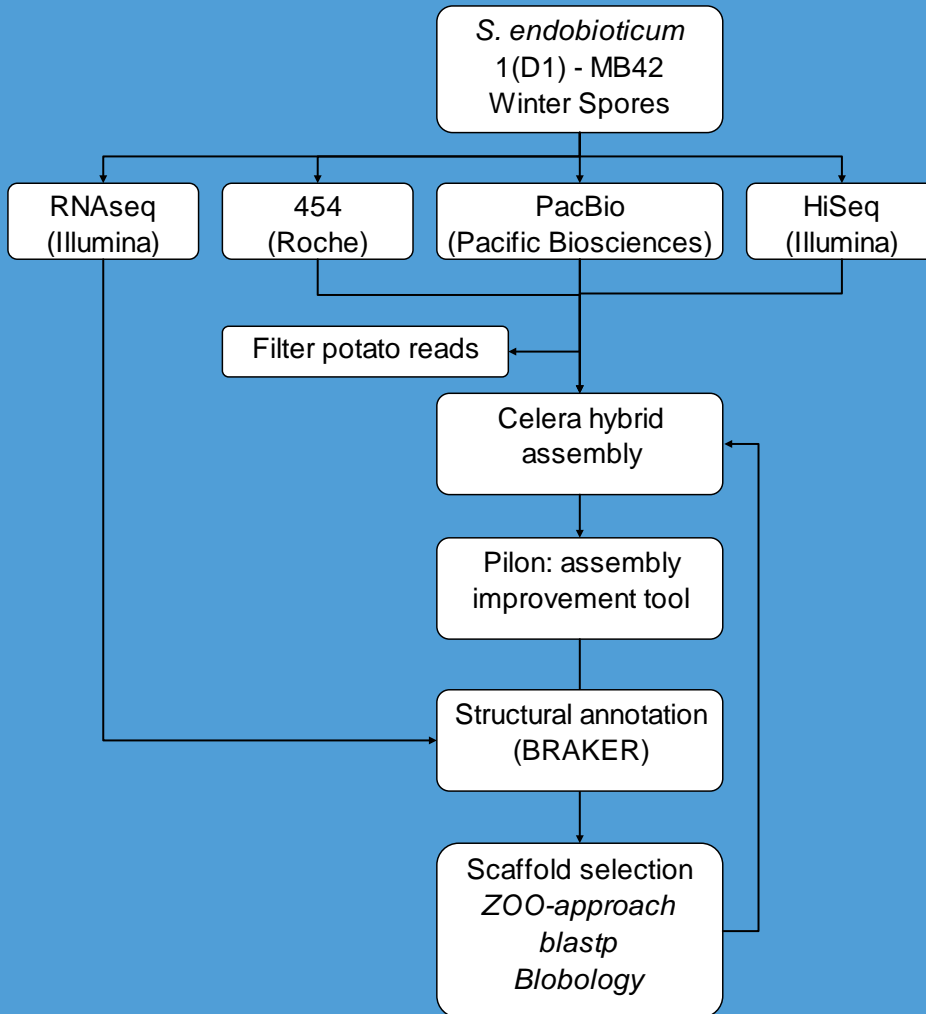
Celera hybrid assembly

Pilon: assembly improvement tool

Structural annotation (BRAKER)

Scaffold selection
ZOO-approach
blastp
Blobology

- Iterative process used to improve:
  - genome assembly
  - structural annotation
  - scaffold selection

- 808 contigs: 21.3 Mb

- N50: 42.7 kb

- 8174 annotated genes

- Current state of the assembly and annotation provides a reliable basis for effector prediction

WAGENINGEN UR
For quality of life

Genome assembly and ZOO approach

- **Several NGS datasets generated with Sendo 1(D1) strain MB42**

  - Illumina HiSeq (DNA and RNAseq), Roche 454, PacBio

- **>2000 scaffolds obtained after subtraction of potato reads**

- **ZOO approach: initial *S. endobioticum* contig selection**

  - Mapping reads of other pathotypes and healthy potato to MB42 scaffolds
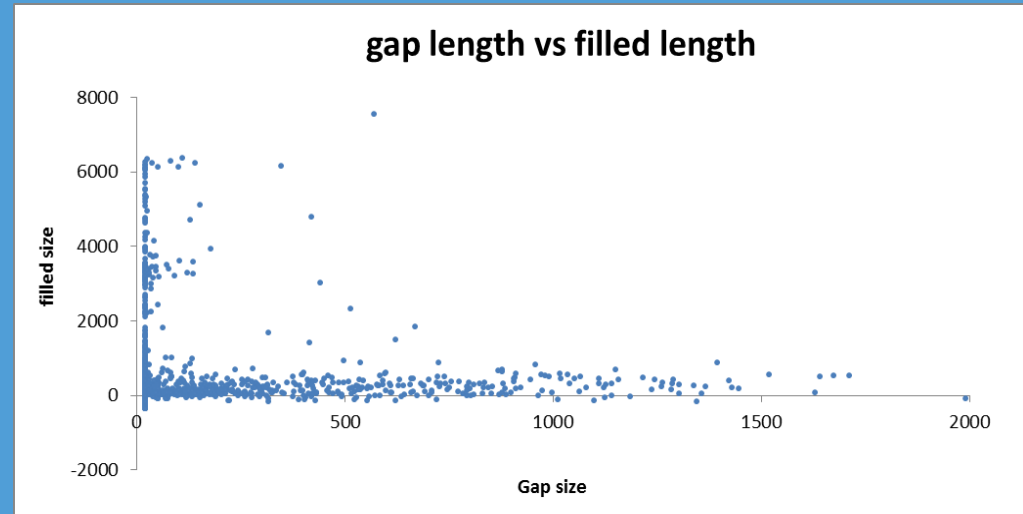
  - Relative read coverage

scf718000	31039	16156	51.6	71.26	69.76	16.48	13.02	14.71
scf718000	37873	22561	57.0	76.5	70.02	18.28	14.56	16.19
scf718000	11069	5682	50.8	67.65	69.05	15.6	12.32	13.82
scf718000	3722	6339	169.7	114.08	112.03	46.07	33.51	51.21
scf718000	15292	8698	56.6	72.98	59.72	18.48	14.18	16.8
scf718000	6947	4461	63.2	82.65	74.81	21.08	17.11	16.72
scf718000	5831	3092	52.7	70.83	52.31	16.15	11.93	13.11
scf718000	21357	10920	50.9	67.63	65.62	15.44	12.93	15.27
scf718000	9935	10177	100.8	66.22	62.83	15.93	18.19	14.12
scf718000	8341	7966	89.3	140.76	150.28	33.15	22.21	31.44
scf718000	10814	5558	51.2	64.9	61.63	15.41	13.13	13.78
scf718000	7653	3701	47.5	58.12	66.14	14.87	10.12	14.19
scf718000	6915	3940	56.3	72.03	71.54	18.08	13.38	16.3
scf718000	27913	16782	59.8	78.02	83.71	19.17	15.48	16.78
scf718000	12332	6504	52.1	68.95	64.09	16.5	12.34	15.17
scf718000	8799	6502	73.0	83.8	82.29	18.88	16.21	17.76
scf718000	15652	9230	58.2	73.47	71.64	18.04	14.52	15.97
scf718000	7581	3054	40.0	65.02	55.07	11.41	11.71	9.27
scf718000	3522	2388	67.3	90.81	91.66	17.6	17.17	20.04
scf718000	20579	12480	59.4	79.13	74.01	18.84	15.44	17.66
scf718000	9620	5392	55.5	72.45	52.66	18.12	14.16	15.78
scf718000	2065	191	4.2	0.14	17	13.19	5.75	29.63
scf718000	1719	84	1.6	0.24	5.83	5.46	2.07	12.71
scf718000	11126	18870	166.5	218.72	120.94	51.26	44.19	47.36
scf718000	9677	6998	69.4	75.11	89.54	19.3	17.19	17.67
scf718000	10313	5546	53.0	71.07	58.94	17.93	13.16	15.15
scf718000	8388	10627	125.5	348.15	268.76	74.82	31.69	71.02
scf718000	26448	16745	62.8	78.64	69.2	18.2	14.56	17.15
scf718000	552	13	0.2	0.26	0.21	0.59	0.24	0.36
scf718000	7828	6167	78.0	92.37	64.89	27.06	16.78	18.74
scf718000	807	22	0.5	0	2.95	4.85	0.43	4.38
scf718000	0	0	-	0	0.13	0.68	0.49	0
scf718000	23101	15127	64.1	109.18	105.05	28.11	17.03	24.45
scf718000	24365	11510	47.0	62.83	62.8	14.35	11.95	12.97
scf718000	39930	35490	87.7	117.12	92.36	26.84	20.93	25.25
scf718000	716	31	0.8	0.41	4.69	3.7	1.4	12.5
scf718000	517	13	0.2	0.11	1.18	2.18	0.92	2.66
scf718000	1773	950	37.0	12.37	183.67	171.98	64.67	266.17
scf718000	12650	9763	76.3	101.24	97.5	22.71	19.32	21.11
scf718000	40778	21715	52.7	67.68	68.61	15.05	12.44	14.57
scf718000	27572	18117	65.3	88.82	85.98	21.07	16.35	19.76
scf718000	59860	33537	55.3	73.63	68	16.93	13.24	16.19

Pilon: assembly improvement tool

- N-stretches (gaps) in assembly complicate structural annotation

- Pilon corrected SNPs, fixed mis-assemblies and filled gaps

- Completely closed gaps (45%), and partially filled (55%) gaps

- Incorrect gap sizes in original assembly

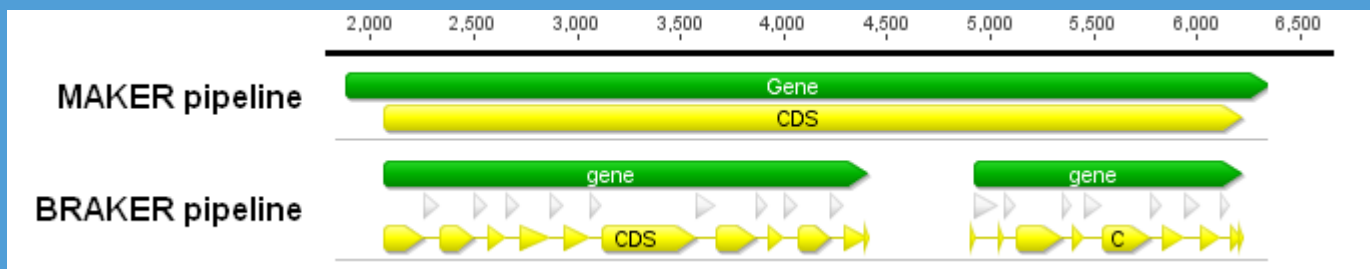|	post-ZOO	Post-ZOO + Pilon	Δ
SNP	11714	10686	-1028
MNP	521	468	-53
Insertion	767	317	-450
Deletion	660	446	-214
Replacement	50	19	-31
Total	13662	11917	-1745
Ns	264022	151866	-112156

gap length vs filled length

Pilon: Walker *et al.* (2014) PLoS One; 9(11)
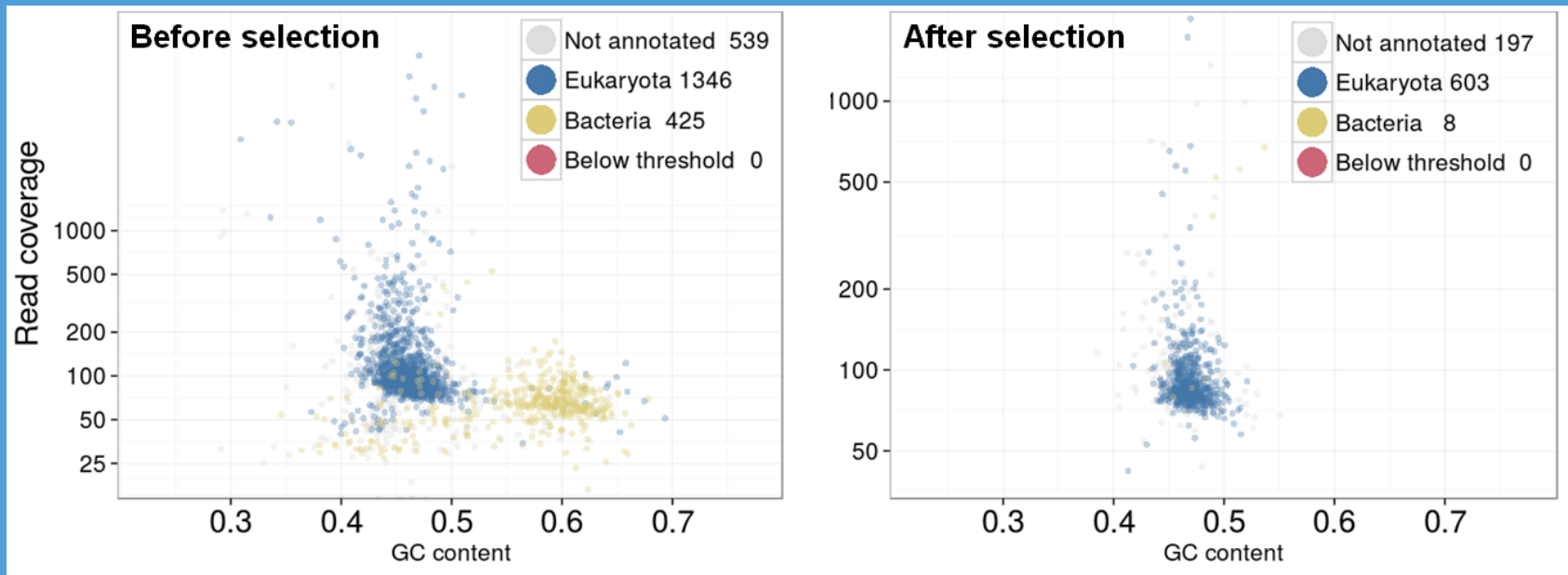
Structural genome annotation

- Fungal gene annotation can be problematic for annotation pipelines

  - Gene-dense: RNAseq data can support CDS spanning more than 2 loci due to overlapping UTRs

  - Short introns: can be missed (annotation pipeline vs. contamination with DNA reads in RNAseq)

- BRAKER successfully used RNAseq data to annotate the gene-dense intron-rich *Synchytrium endobioticum* genome

MAKER: Cantarel et al. (2008) Genome Res. 18(1)
BRAKER (unpublished)
http://exon.gatech.edu/genemark/braker1.html

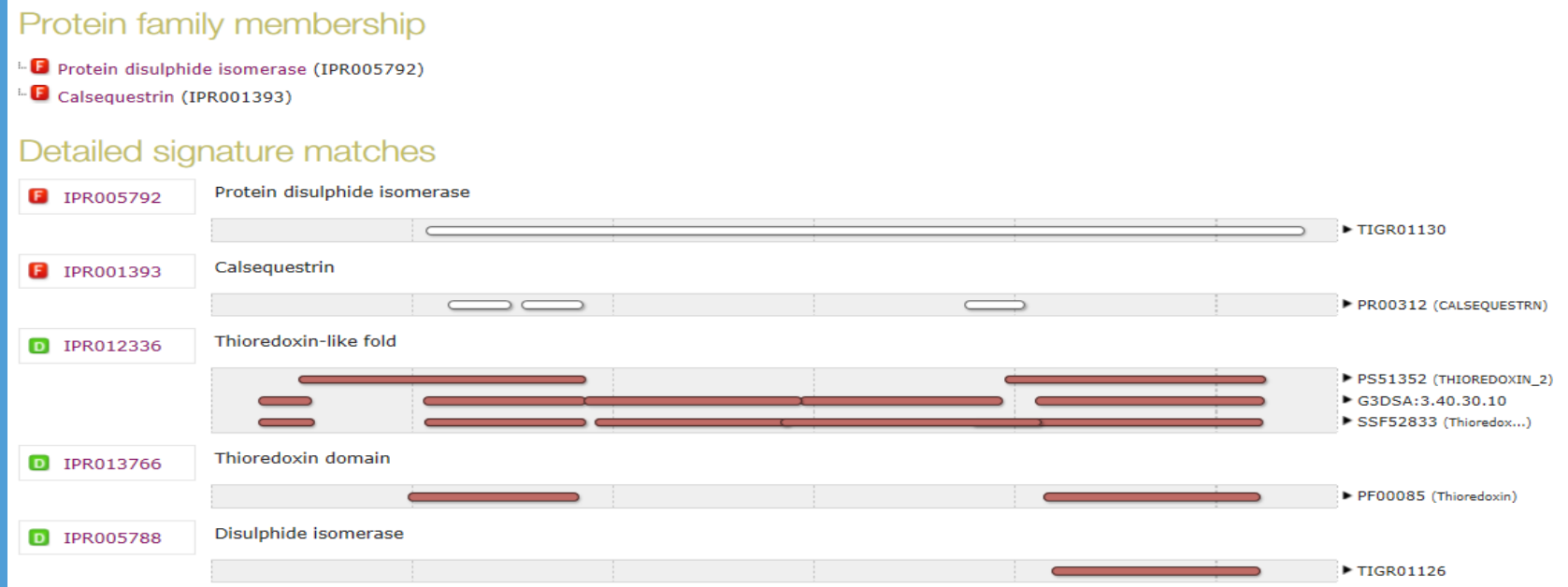# Validation of scaffold selection

- Blobology was used to visualise and validate scaffold selection

- Combines GC content, read coverage and blastn-ID

- Blastn identification on large scaffolds not always informative

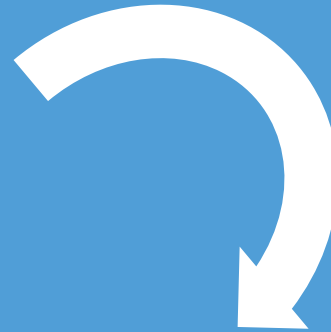Blobology: Kumar *et al.* Front Genet. 2013; 4: 237

Functional annotation - InterProScan

- Classifying proteins into families, and predict domains and important sites

- Combines resources from 11 databases

- Enables effector prediction

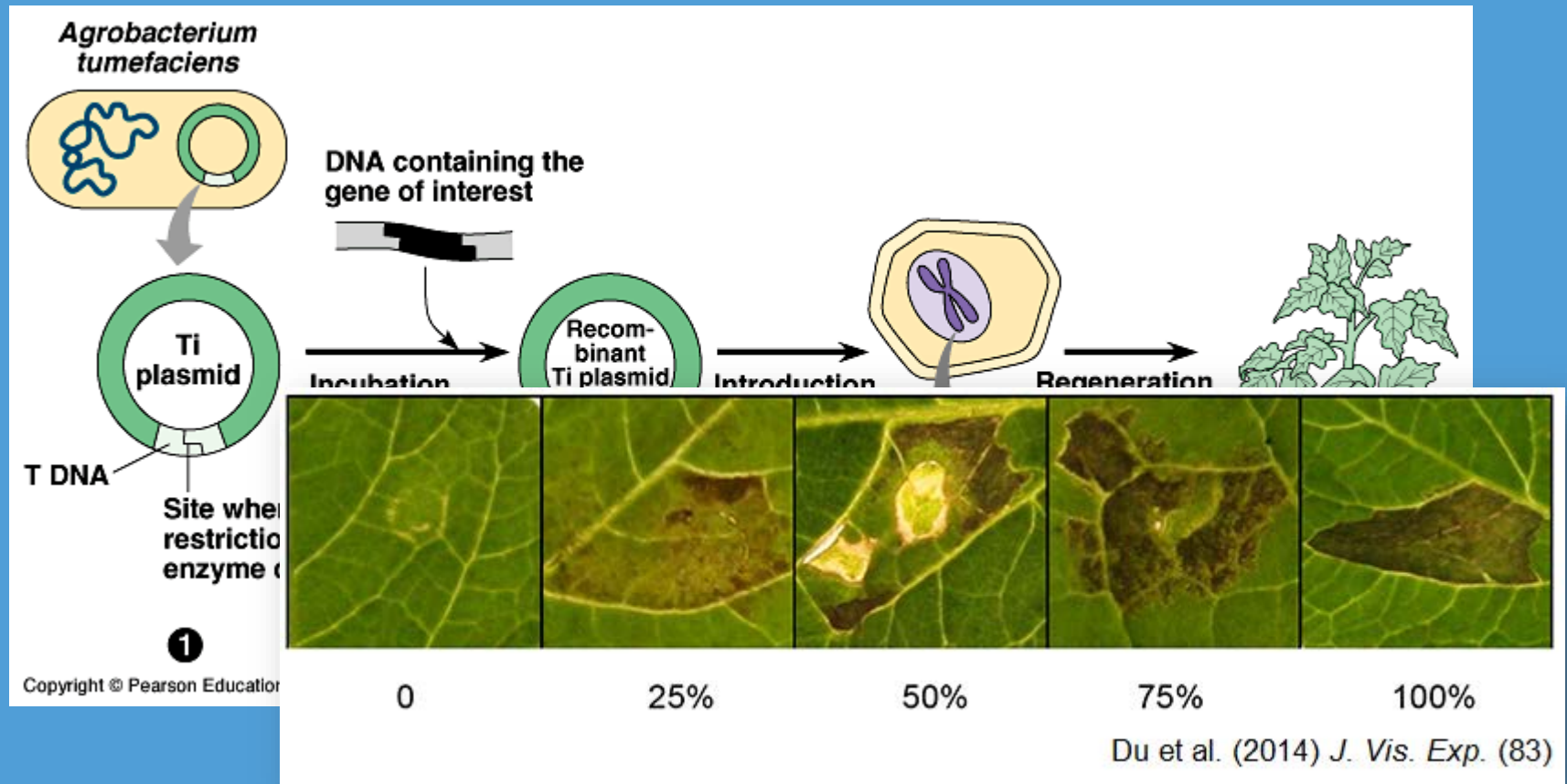InterProScan 5: Jones *et al.* (2014) Bioinformatics; 30 (9)

We have the genome… What's next?

Effector prediction and selection

- There is no single strategy for fungal effector prediction

- However, different traits and criteria can be exploited
  - Secreted, small and cysteine-rich
  - Presence/absence Pfam domains
  - Effector motifs/internal repeats
  - Clustering into (pathogen associated) gene families
  - Species/pathotype/strain specific
  - 3^0 structure and expression

The HR-inducing ability of putative effectors using Agroinfiltration

Du et al. (2014) *J. Vis. Exp.* (83)

Conclusions

- We have sequenced and assembled the genome of *S. endobiotiucum*

- The 21.3 Mb draft genome sequence of comprises 8174 predicted genes

- Pilon fixed mis-assemblies and closed gaps in the genome sequence

- BRAKER was able to predict genes on the gene-dense and intron-rich genome

- The structurally annotated assembly provide a reliable gene/protein set for effector prediction

- There is no straightforward strategy for effector prediction, but combined traits can be used for prediction and selection of promising effector candidates

Further activities

- Identification and selection of putative effectors for *in planta* expression using agroinfiltration

- Agroinfiltration of selected effectors to test their ability to trigger a hypersensitive response in resistant potato cultivars

- Determine trait associated inter-pathotype variation and develop molecular pathotype specific identification tools

WAGENINGEN **UR**
For quality of life

Acknowledgements

- Prof. Dr. Richard Visser, WUR plant Breeding

- Gerard van Leeuwen and Patricia van Rijswick, National Plant Protection Organisation, Wageningen, NL

- Ineke van Holst and Margriet Boerma, HLB, Wijster, NL

- Canadian partners

- Funding

Thank you for your attention!