

NLR STATION

MAR 21, 2021

NLR STATION

MAR 8, 2021

Goals:

What is the structure of a typical NLR gene?

What are some challenges associated with identifying NLR genes?

Why do NLR biologists love pangenomes?

Why is there so much funded research in NLR biology?

Introduction

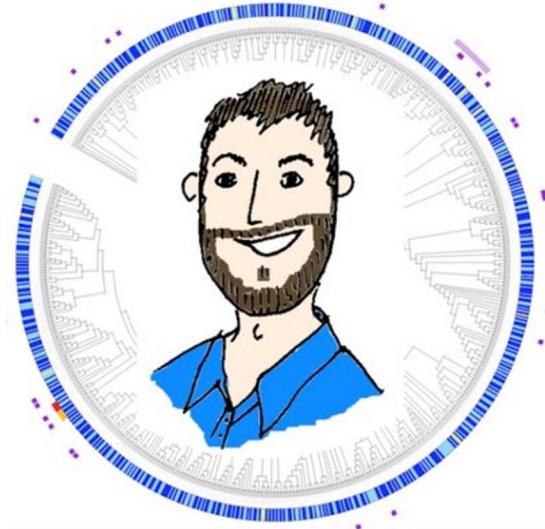
B.A. in Ecology, Evolution, Conservation Bio from
University of Hawai'i at Hilo

- Research in a plant virology lab

Research Associate with Monsanto in Cambridge, MA
-Insect-control proteins from bacteria

PhD in Plant-Microbe Biology from Cornell U.
-Molecular warfare during bacterial infection of rice

NSF PRFB Postdoc Fellow with Nathan Springer
-Epigenetics, what even is it?



@hashtag_read
read0094@umn.edu

Warm-up Question 1:

“NLR” is an acronym for:

- A. Nucleotide-binding and Leucine-rich Repeat
- B. NOD-Like Receptor
- C. Nuclear Localized Resistance gene
- D. I thought these things were called NB-LRRs or R-genes

Warm-up Question 1:

“NLR” is an acronym for:

- A. Nucleotide-binding and Leucine-rich Repeat
- B. NOD-Like Receptor
- C. Nuclear Localized Resistance gene
- D. I thought these things were called NB-LRRs or R-genes

Warm-up Question 2:

According to the reading, which of these plants has the largest NLR repertoire?

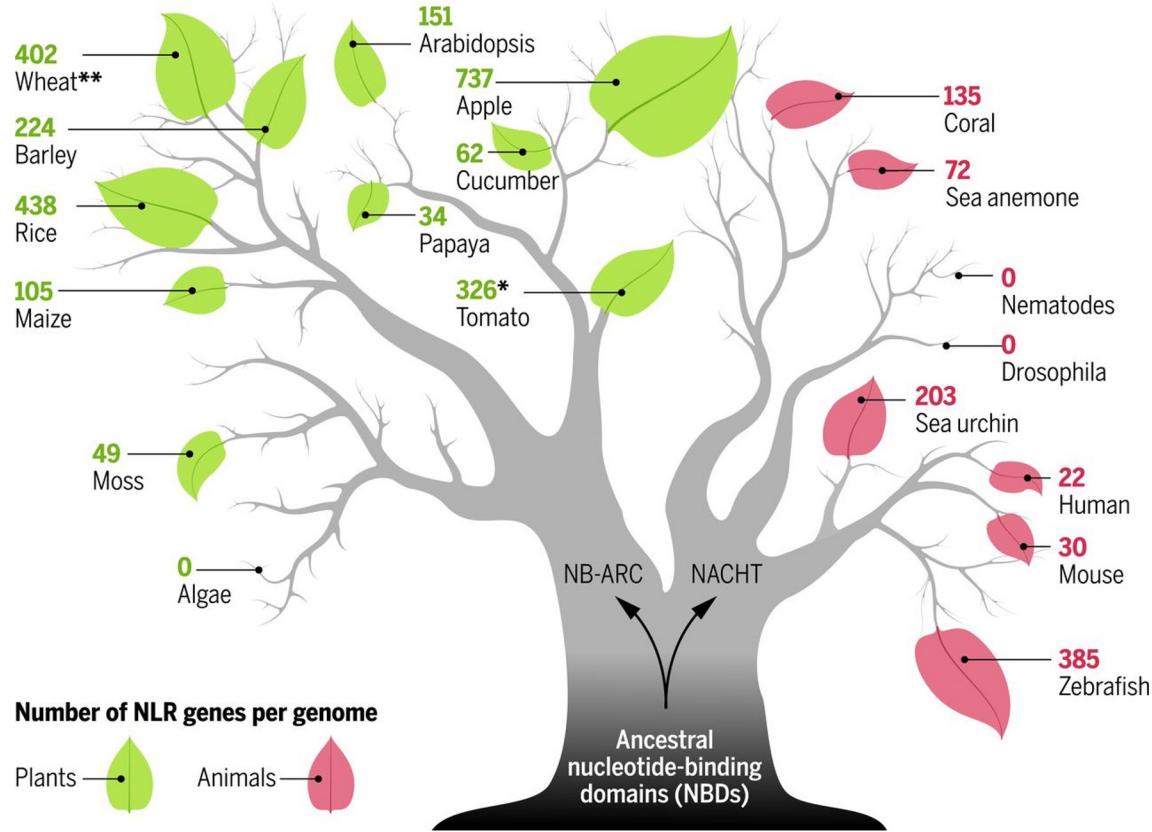
- A. Giant Sequoia
- B. Apple
- C. Wheat
- D. Cucumber

Warm-up Question 2:

According to the reading, which of these plants has the largest NLR repertoire?

- A. Giant Sequoia
- B. Apple*
- C. Wheat***
- D. Cucumber

When corrected for ploidy – Apple wins



Intracellular innate immune surveillance devices in plants and animals

BY JONATHAN D. G. JONES, RUSSELL E. VANCE, JEFFERY L. DANGL
SCIENCE 02 DEC 2016

Warm-up Question 3:

According to the reading, what is the largest NLR cluster?

- A. RGC2 in lettuce
- B. Xo1 in rice
- C. RPPS in *Arabidopsis*
- D. NLR2000 in baobab tree

Warm-up Question 3:

According to the reading, what is the largest NLR cluster?

- A. RGC2 in lettuce **24 cluster members - ~3.5 Megabases!**
- B. Xo1 in rice
- C. RPPS in Arabidopsis
- D. NLR2000 in baobab tree

Life is hard for plants....

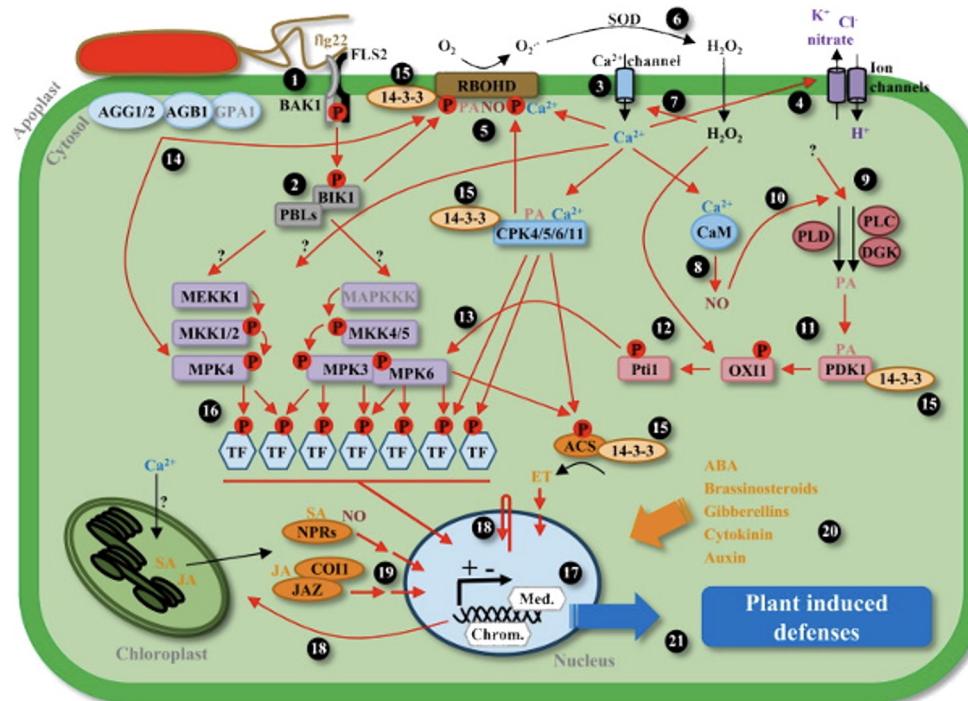
DALE WALTERS

FORTRESS PLANT

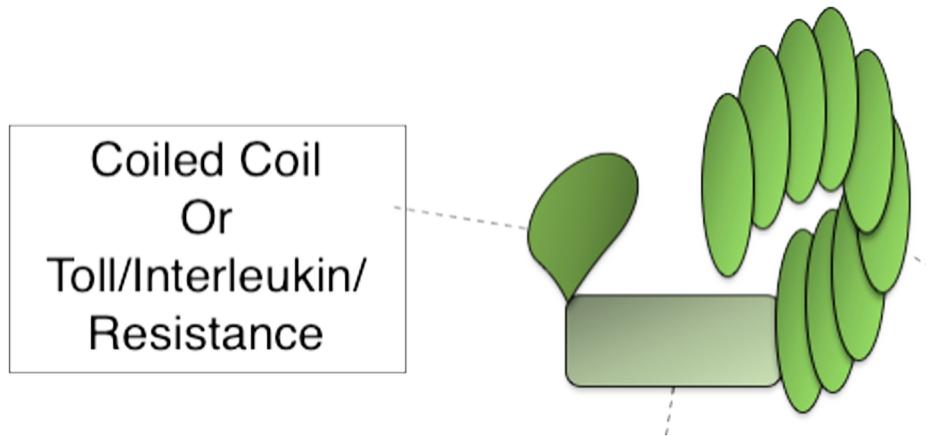
*How to survive
when everything
wants to eat you*



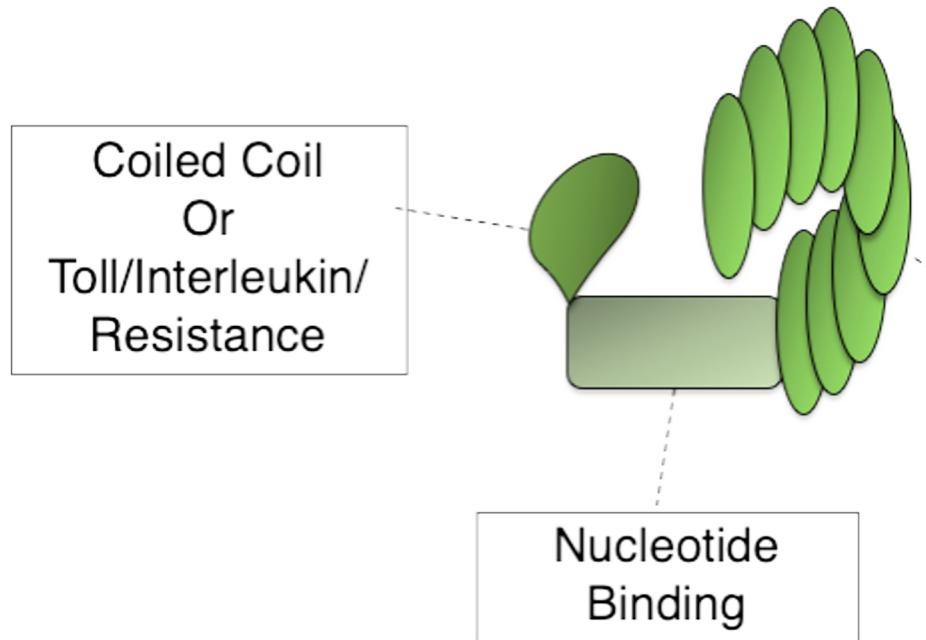
Molecular plant defense is complicated



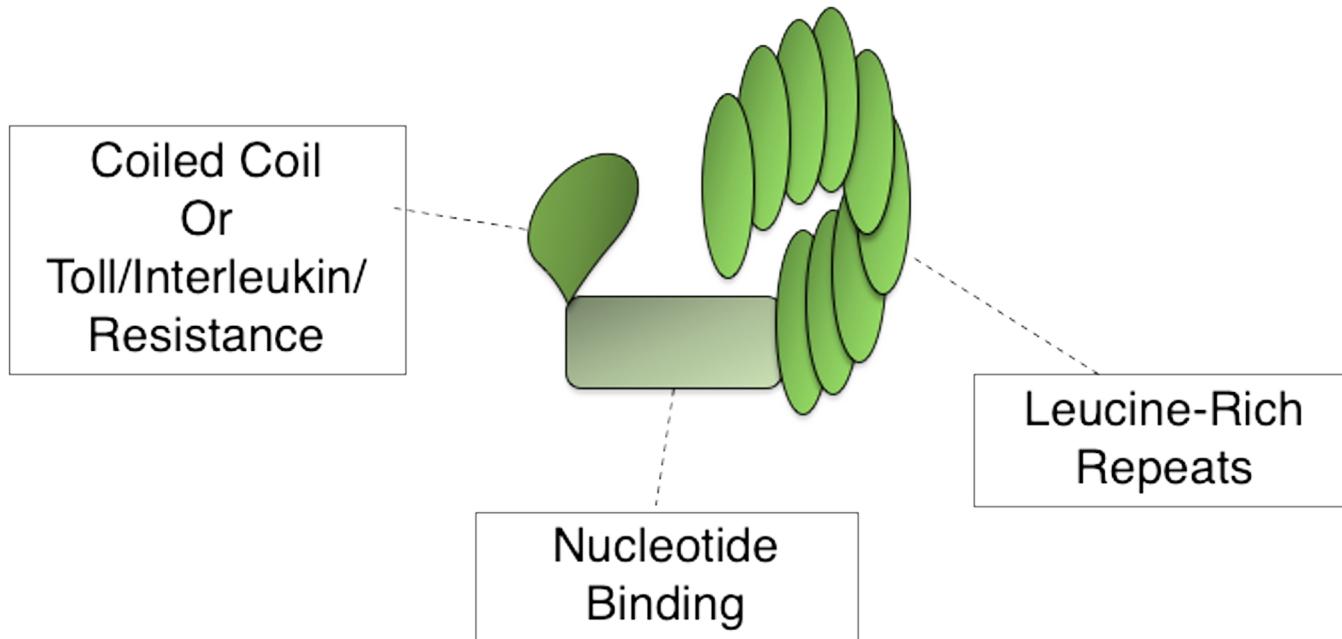
Introducing the NLR gene family



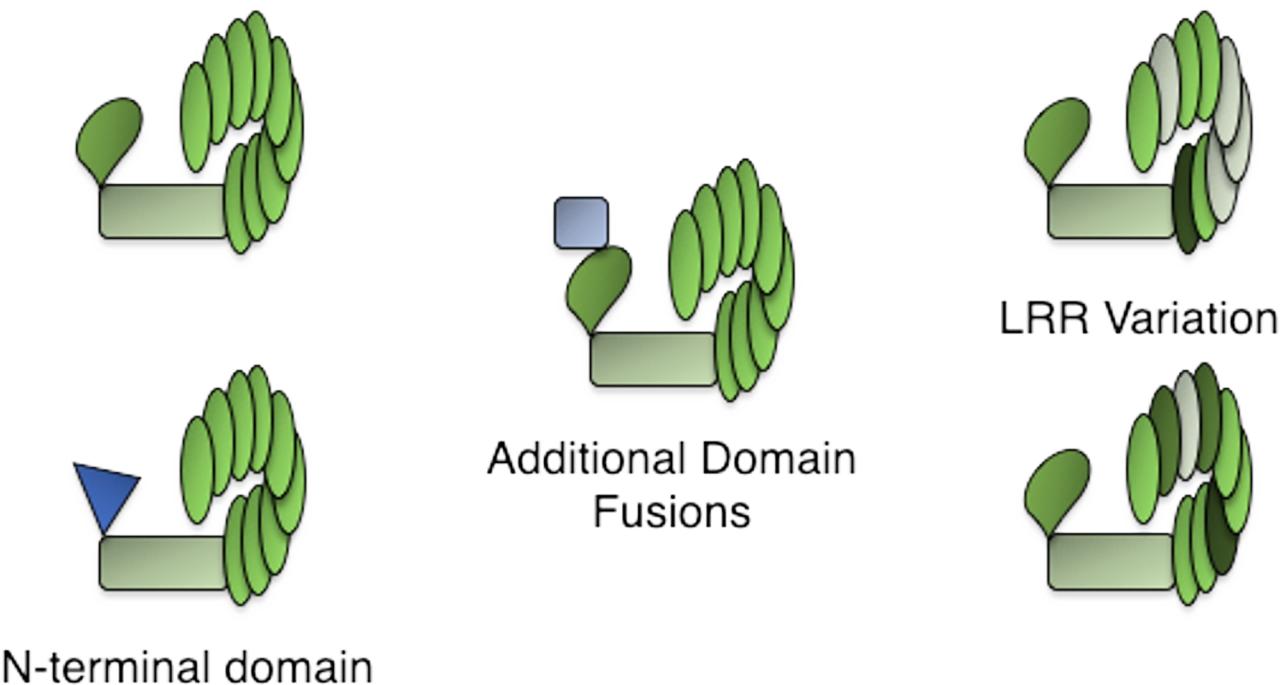
Introducing the NLR gene family



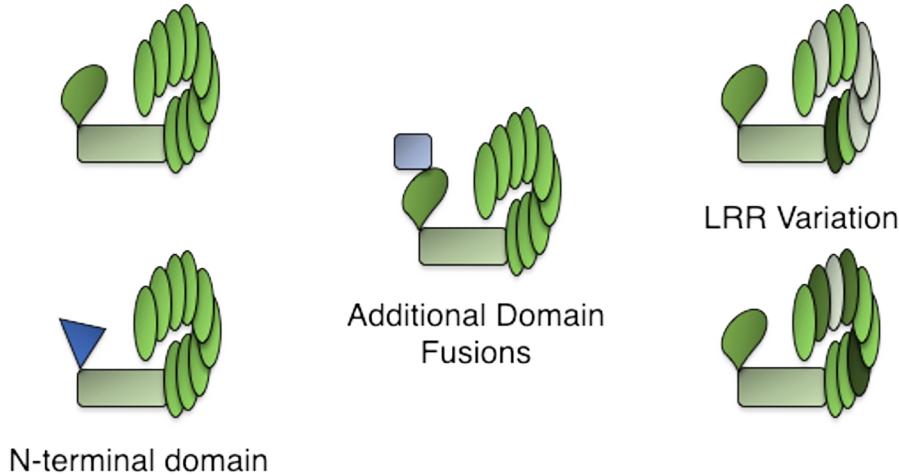
Introducing the NLR gene family



NLRs are among the most diverse plant gene families -
Both within individuals and at the population level

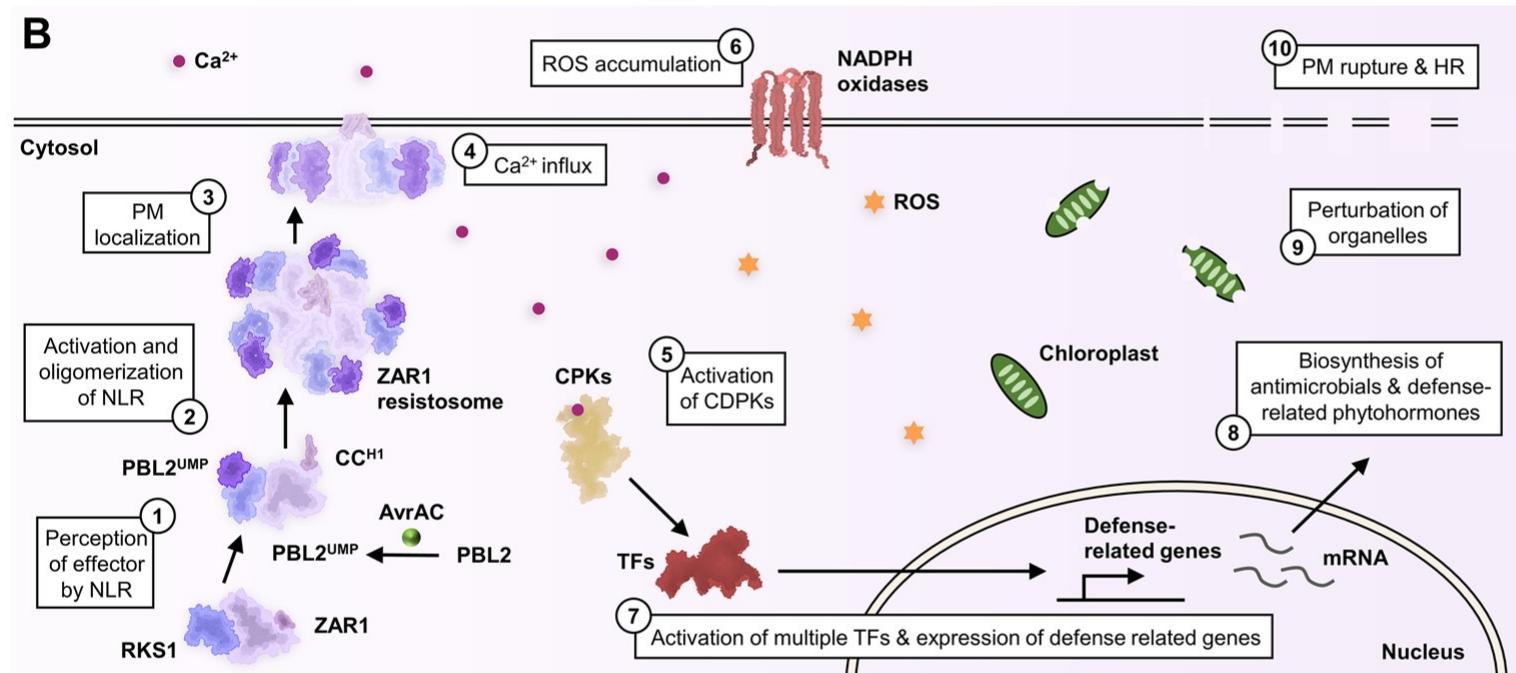


NLRs are among the most diverse plant gene families -
Both within individuals and at the population level



Plants lack an adaptive immune system (antibodies) - the large repertoire of NLRs act as an alternative mechanism to recognize 'non-self'

How do NLRs ‘work’?

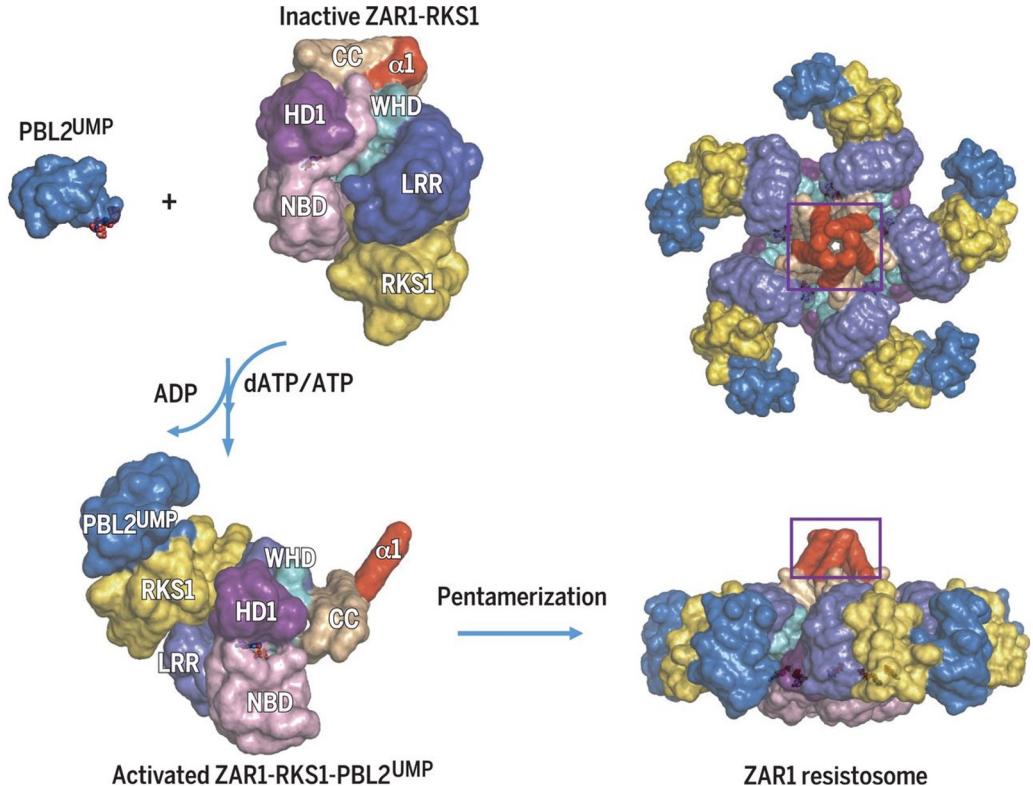
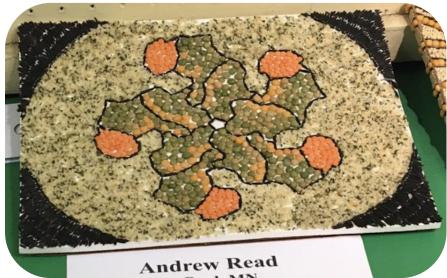


Thirty years of resistance: Zig-zag through the plant immune system

Bruno Pok Man Ngou, Pingtao Ding, Jonathan D G Jones Author Notes *The Plant Cell*, koac041, <https://doi.org/10.1093/plcell/koac041>

Published: 15 February 2022

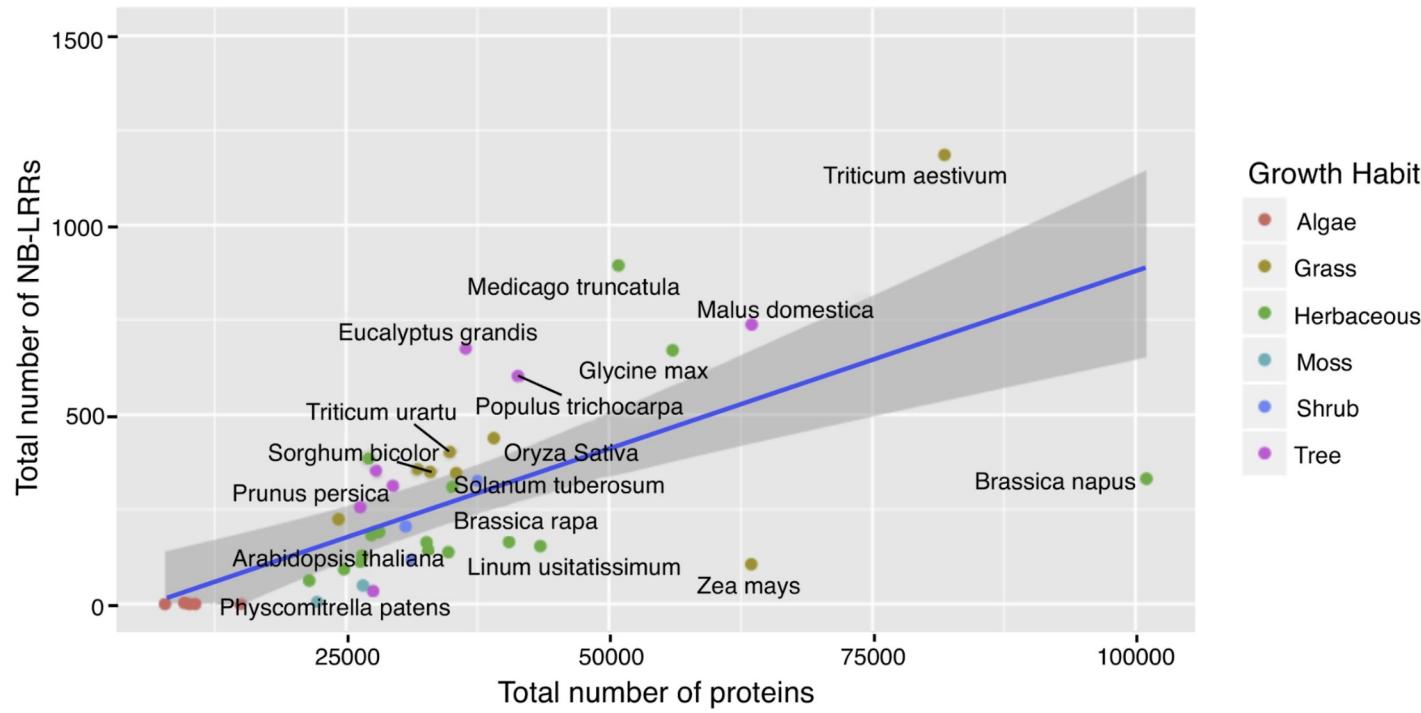
How do NLRs ‘work’?



Reconstitution and structure of a plant NLR resistosome conferring immunity

BY JIZONG WANG, MEIJUAN HU, JIA WANG, JINFENG QI, ZHIFU HAN, GUOXUN WANG, YIJUN QI, HONG-WEI WANG, JIAN-MIN JIJI CHAI
SCIENCE 05 APR 2019

NLRs are diverse within and between individual genomes



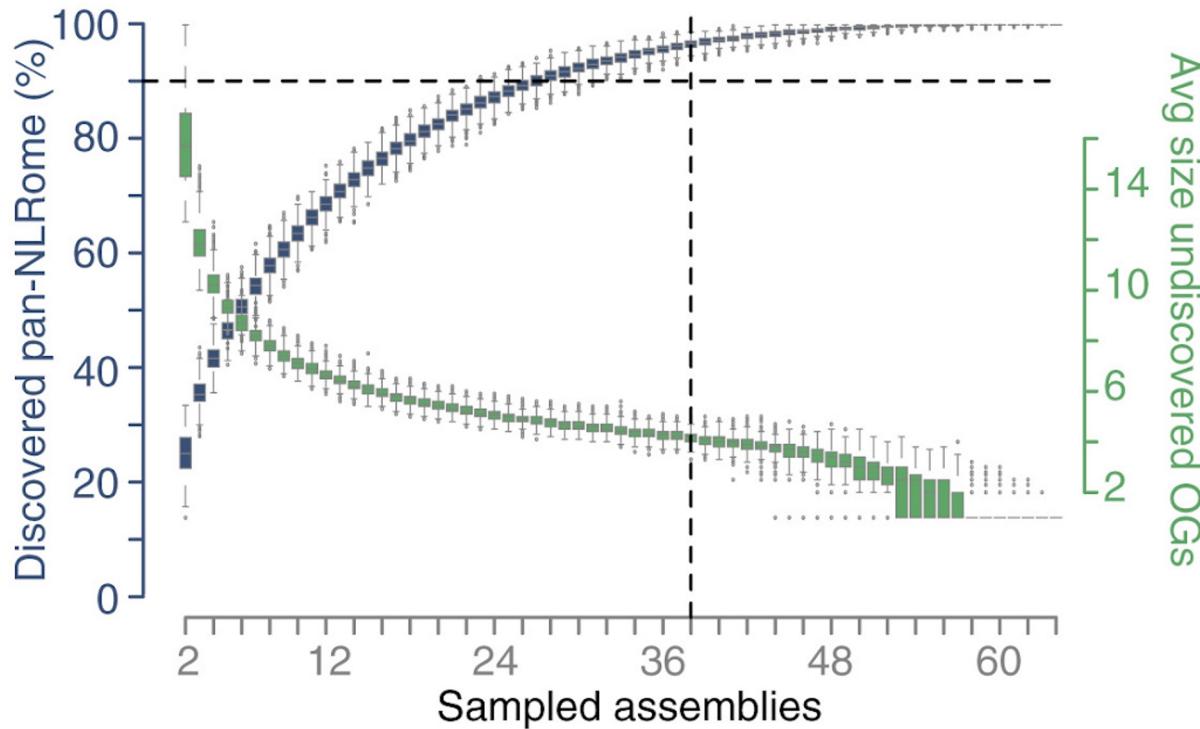
That's a lot of diversity within individuals, What about at the species level?

Table 1

Examples of plant pan-genome or pan-NRLome studies and key observations on NLRs or *R* genes

Plant pan-genome or pan-NRLome studies					
Organism (s)	<i>N</i>	Sequencing technology	Assembly strategy	Reference	Key observations on NLR or <i>R</i> genes
<i>Brassica rapa</i>	3	Short reads	De novo assemblies	Lin et al. (2014)	- NLRs enriched in dispensable genome.
<i>Glycine soja</i>	7	Short reads	De novo assemblies	Li et al. (2014)	- CNVs common in NLRs: candidates of resistance differences between wild and cultivated accessions. - NLRs numbers and domain architectures varying between species. - NLRs enriched in dispensable genome.
<i>Zea mays</i>	503	Short reads	De novo transcriptome	Hirsch et al. (2014)	
<i>Oryza sativa</i>	3	Short reads	De novo assemblies	Schatz et al. (2014)	- NLRs enriched in dispensable genome, e.g. 12% shell versus 0.35% core genes are NLRs.
<i>O. sativa</i>	1483	Short reads	Iterative assembly	Yao et al. (2015)	- NLRs enriched in dispensable genome.
<i>Populus</i> clade	7	Short reads	Map to reference	Pinosio et al. (2016)	- CNVs and SVs enriched for NLRs.
<i>Brassica oleracea</i>	10	Short reads	Iterative assembly	Golicz et al. (2016)	- NLRs enriched in genes showing PAV. - 43% of NLRs dispensable, 45% in clusters, and 60% absent from reference.

How to capture NLR diversity? Enter the panNLR-ome

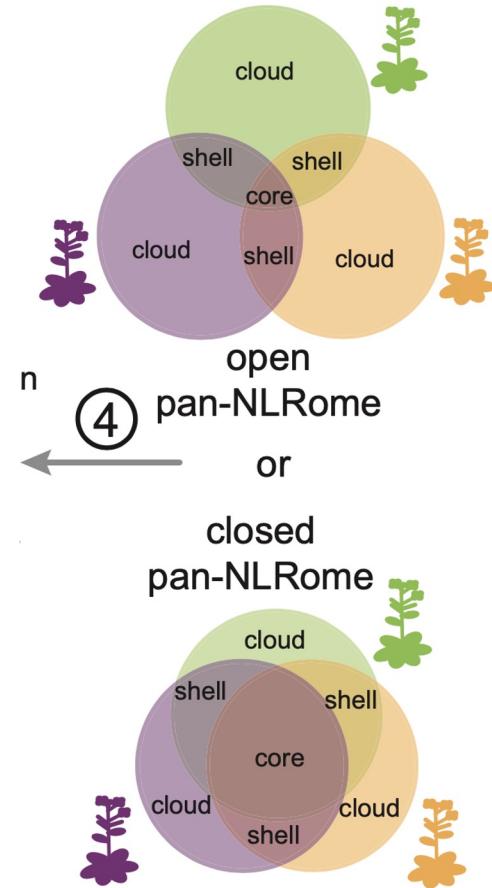
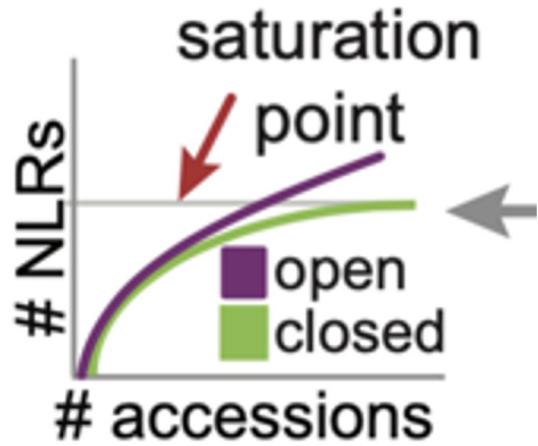


CELL VOLUME 178, ISSUE 5, P1260-1272.E14, AUGUST 22, 2019

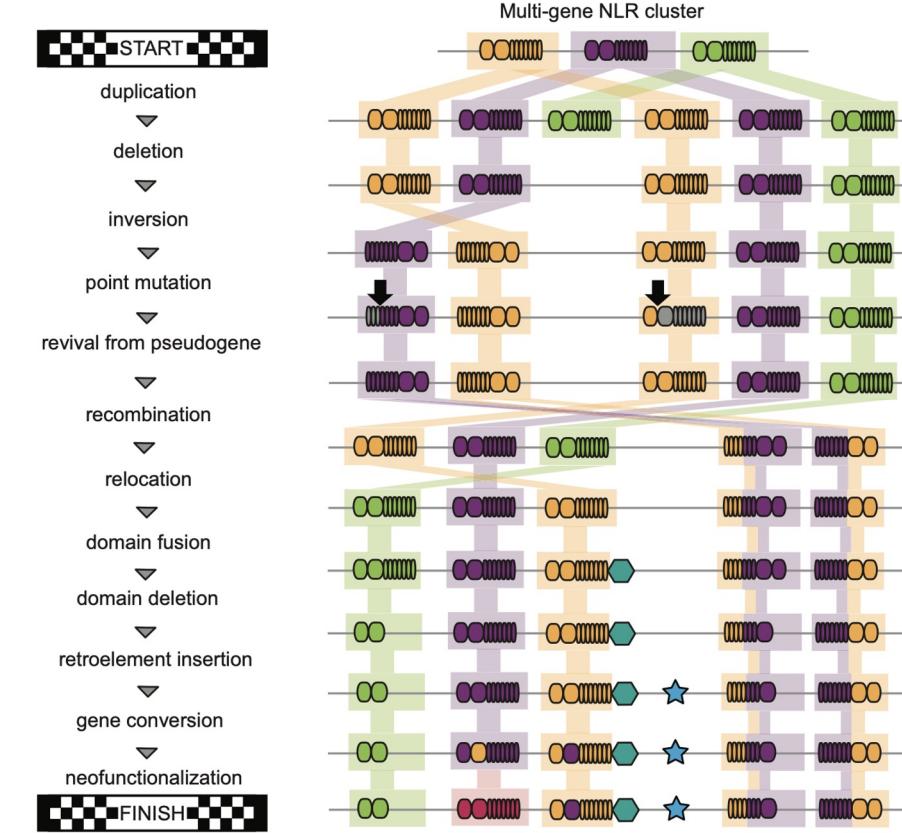
A Species-Wide Inventory of NLR Genes and Alleles in *Arabidopsis thaliana*

Anna-Lena Van de Weyer 8Freddy Monteiro 8Oliver J. Furzer 8Marc T. Nishimura 8Volkan Cevik 8Kamil Witek 8Jonathan D.G. Jones 8Jeffery L. Dangl 8Detlef Weigel 9Felix Bemm 8

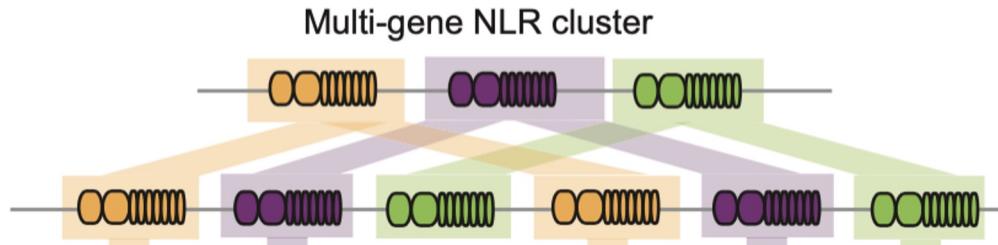
Will all species have similar curves?



But, NLRs are challenging to sequence, assemble, and annotate!



But, NLRs are challenging to sequence, assemble, and annotate!

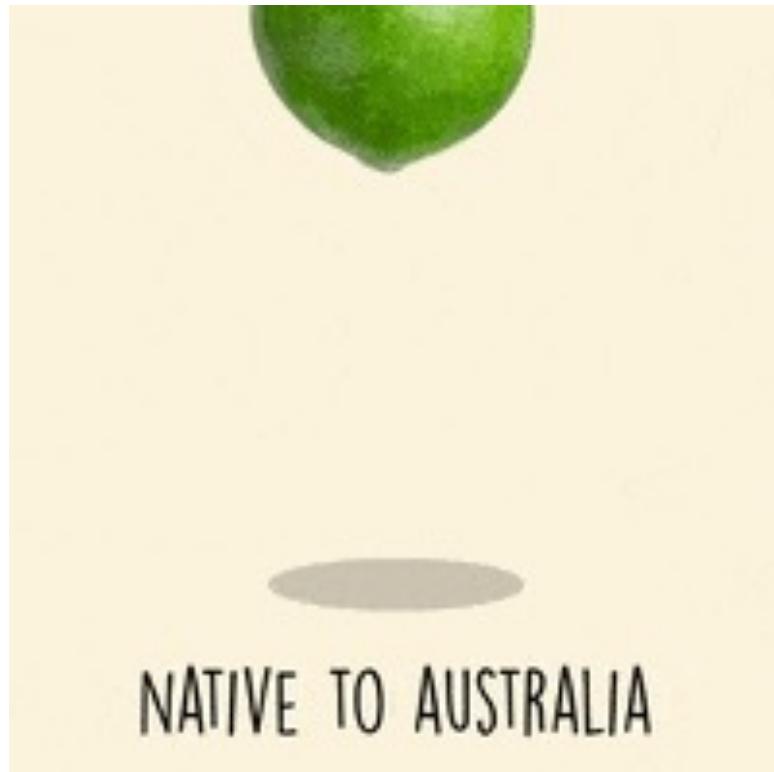


Gedankenexperiment time!

You are given enough money to sequence your favorite plant with one sequencing technology, and your goal is to identify NLR diversity - what technology do you choose? Why? Are there any down-sides to this choice?

- A. Illumina
- B. PacBio
- C. Oxford Nanopore
- D. Other

Let's take a look at NLRs in..... *Macadamia integrifolia*



Macadamia facts:

14 chromosomes

Genome size ~745 megabases

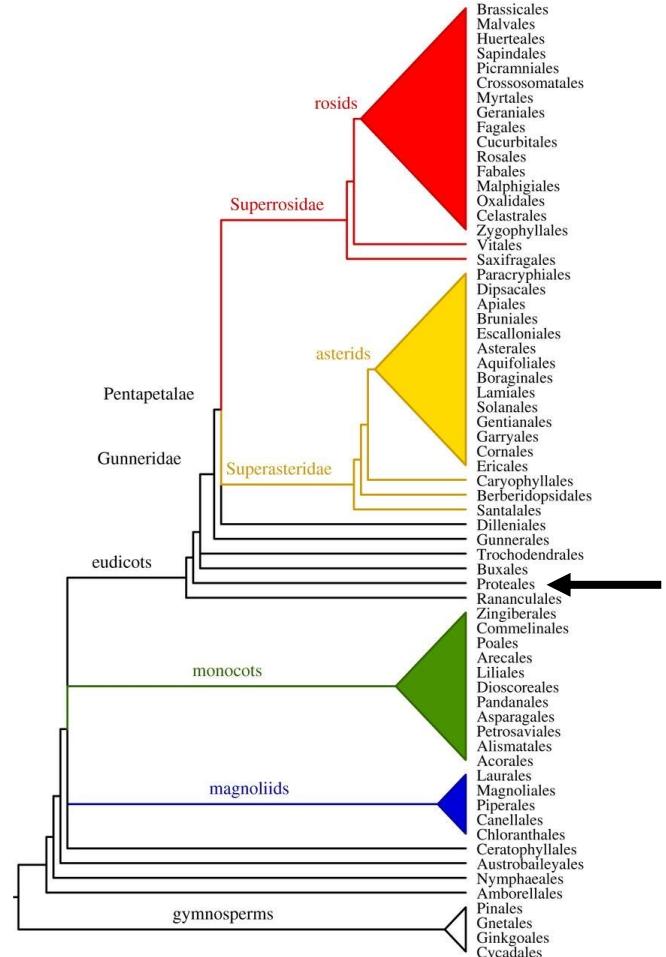
Assembly size 382 megabases

MAKER used for gene annotation –
34,275 predicted genes

TEs and other repeats make up 55%
of the assembly

Hardest nut in the world to crack!!!

Any NLR number guesses?



We have a genome - time to find some NLRs

Evidence-based

Our basic pipeline

0. Obtain protein sequences of species of interest and organise them into a directory.

We follow the Phytozome organisation of `master_dir/species/annotation/species_version_proteins.fa` where each species is denoted by the first letter of the genus name and all letters in the species names, for example `Athaliana`

1. Pfam-based annotation of domains

usage: `bash run_pfam_scan.sh dir`

Dependencies:

- HMMER software (<http://hmmer.janelia.org/>) including pfam_scan.pl (part of HMMER) Move in same directory as this script or set path at command string.
- Pfam database (<http://pfam.xfam.org/>)
- File names should be consistent with Phytozome and include Species_*_protein.fa
- perl modules specified in the scripts (best to install with cpan: <http://www.cpan.org/modules/>)

2. Parsing the pfamscan output with K-parse_Pfam_domains_v3.1.pl

- The script parses the output of pfam_scan.pl
- The script extracts all domains for each proteins and removes redundant nested hits with larger e-values.
- Domains are printed out in the order of appearance in the query.
- By default, Pfam_B domains are skipped.

Ab initio NLR-Annotator

github.com/steuernb/NLR-Annotator

1. ChopSequence.jar -- chop the genome into 20kb fragments with 5kb overlaps
2. NLR-Parser.jar -- 6 frame translation of each genomic fragment
 - Translations scanned for NLR domains
3. NLR-Annotator.jar -- 20kb fragments are merged
 - Potential NLR-containing regions are identified
 - NLR domains are listed and NLRs are classified
(Complete/Partial, Pseudo)
 - Optional NB domain alignment for building trees

NLR-Annotator – Scripts and Outputs

```
(base) read0094@ln0004 [~/Scriptz/NLR-Annotator] % ls -lh
total 3.6M
-rw-----. 1 read0094 springer 140K Mar  6 16:54 ChopSequence.jar
-rwx-----. 1 read0094 springer 455K Mar  6 16:54 mast
-rw-----. 1 read0094 springer 1.7M Mar  6 16:54 meme.xml
-rw-----. 1 read0094 springer 140K Mar  6 16:54 NLR-Annotator.jar
-rw-----. 1 read0094 springer 50K Mar  6 16:54 NLR-Parser3.jar
-rw-----. 1 read0094 springer 140K Mar  6 16:54 NLR-Parser.jar
-rw-----. 1 read0094 springer 4.7K Sep 28 17:00 readme.md
```

```
read0094@C02D82DPML7L Macadamia % ls -lh
total 3653608
-rw-r--r--@ 1 read0094  staff  720M Mar  2 13:42 GCF_013358625.1_SCU_Mint_v3_genomic.fna
-rw-r--r--  1 read0094  staff  93M Mar  2 15:08 Macadamia.NLR.xml
-rw-r--r--  1 read0094  staff  950M Mar  2 13:48 Macadamia_CHOPPED.fna
-rw-r--r--  1 read0094  staff   93K Mar  2 16:41 NBARC_Macadamia.fasta
-rw-r--r--  1 read0094  staff   67K Mar  2 15:27 output.txt
```

NLR-Annotator output

How many NLRs? - NLR-Annotator-based

From our *ab initio* analysis - NLR-Annotator

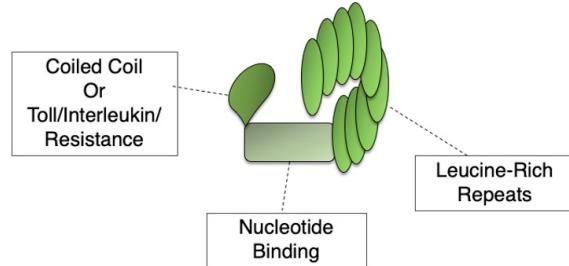
640 total NLRs

450 on chromosomes

190 on unplaced contigs...

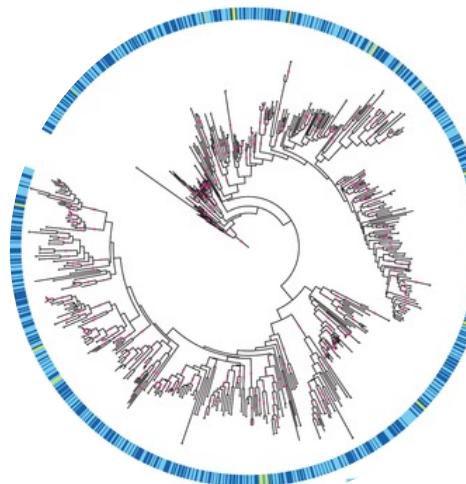
Row Labels	Count (on chr)
complete	341 (248)
complete (pseudogene)	166 (118)
partial	63 (36)
partial (pseudogene)	70 (48)

NLR-Annotator output - sequence alignment of NB-domain



NLR phylogenies usually use the Nucleotide Binding domain because it evolves in a ‘normal’ way

CLUSTAL O(1.2.4) multiple sequence alignment



Advantages of the different NLR annotation methods

Evidence-based

More likely to get the ‘true’ or ‘active’ NLRs - this represents the plant’s potential defense gene repertoire (important for breeders/pathologists)

Sequence-based

No genome masking required

Can identify pseudogenes - not important for current resistance but may be building blocks for future genes or teach us about evolutionary history

But - identifies regions, not genes - is not splice-aware

Misses certain classes of NLRs

For pan-genome comparisons - probably fine as long as you are consistent in your method across samples

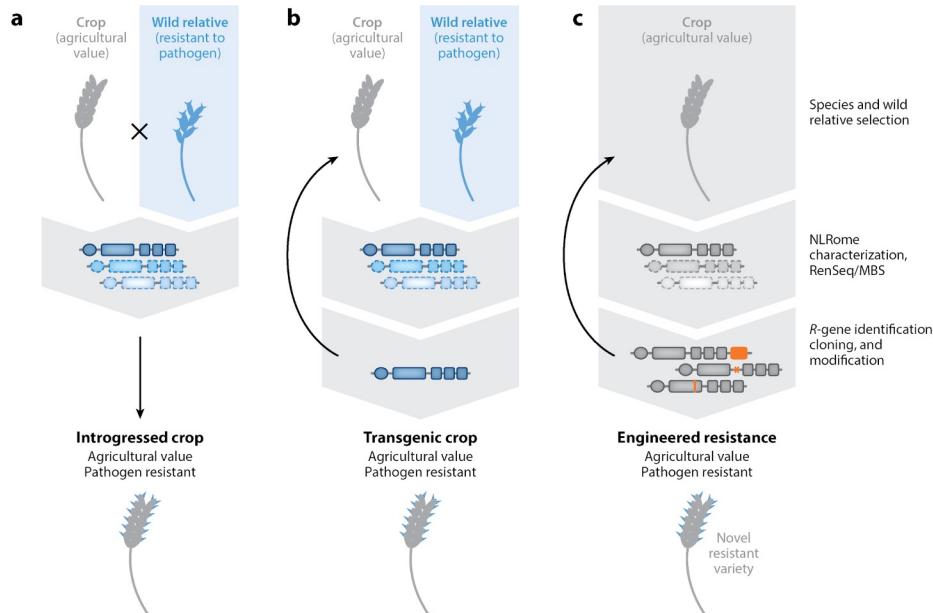
Additional Challenges:

How to resolve evidence-based and sequence-based predictions?

How many genomes do you need for a pangenome?

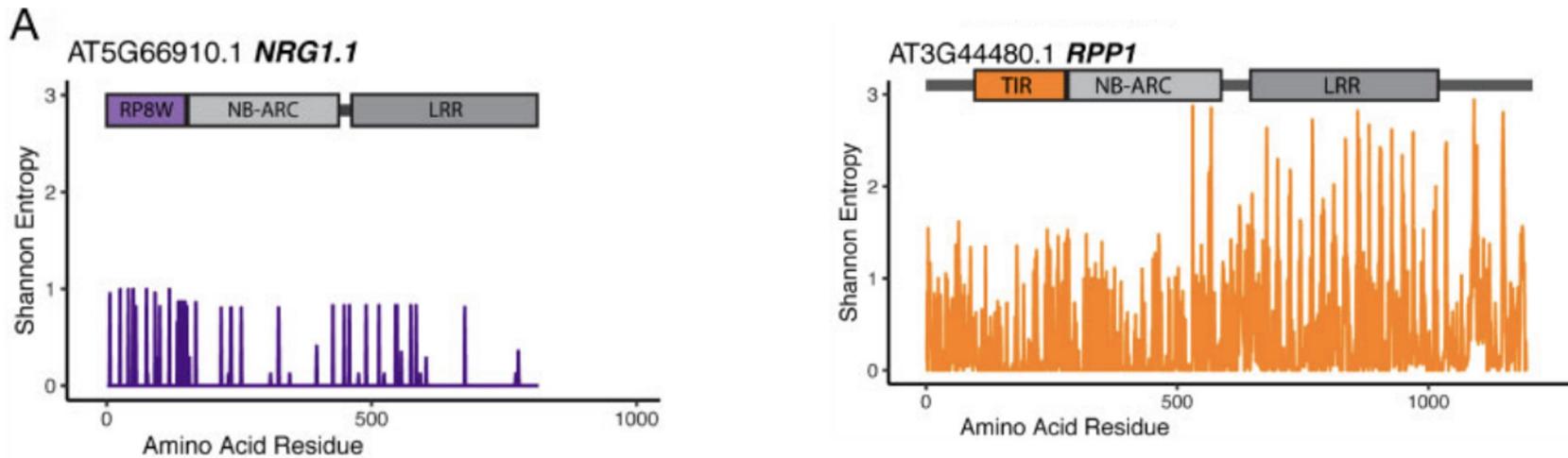
How will we leverage the data?

How will we leverage the data: Pan-genomes can inform disease resistance breeding



Monteiro F, Nishimura MT. 2018.
Annu. Rev. Phytopathol. 56:243–67

How will we leverage the data: Engineered recognition - a disease resistance ‘grail’



High entropy values correlate strongly with surface exposure and hydrophilic character (Liao et al., 2005) and can be used to predict rapidly evolving ligand-binding sites

Thanks for riding!

What is the structure of a typical NLR gene?

What are some challenges associated with identifying NLR genes?

Why do NLR biologists love pangenomes?

Why is there so much funded research in NLR biology?

