Special Topics in Genomics - The Hazelnut Blight Genome Project

Class	Date	Topic(s)
1	Oct. 14th	-Why is this course possible? (introduction to next-gen sequencing - David Guttman) -What is Hazelnut Blight? (introduction to bacterial plant pathogens - David Guttman) -Where can we see the data? (introduction to servers, directories and files - Alan Moses) -Organization of the course. -Assignment of short read assembly papers
2	Oct. 21st	-Assembly of short reads (Mike Brudno) -Discussion of short read assembly papers -Formation of assembly teams
3	Nov. 4th	-Student teams present assembly results
4	Nov. 11th	-Gene finding and annotation (Alan Moses) -Assignment of annotation (gene finding) papers
5	Nov. 18th	-Discussion of gene-finding papers -Formation of annotation teams
6	Dec. 2nd	-Student teams presentation of annotation results -Assignment of final projects

David Guttman david.guttman@utoronto.ca
Alan Moses alan.moses@utoronto.ca
http://www.moseslab.csb.utoronto.ca/alan/STIG-HBGP.html

Special Topics in Genomics - The Hazelnut Blight Genome Project

Grading:

- Class participation (50%)
- Class presentations (25%)
- Comparative genome analysis (25%)

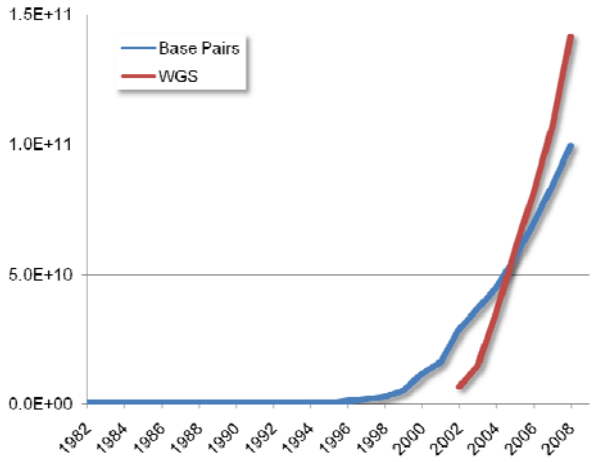
David Guttman david.guttman@utoronto.ca
Alan Moses alan.moses@utoronto.ca
http://www.moseslab.csb.utoronto.ca/alan/STIG-HBGP.html

Special Topics in Genomics - The Hazelnut Blight Genome Project

Outline

- Next-gen genomics background
- Applications of next-gen genomics
  - *Pseudomonas syringae* genomics
  - Massively parallel analysis of type III effector interactions
  - Next-generation mapping of *Arabidopsis thaliana* cell wall accessibility mutants
- *P. syringae* pv. avellanae and Hazelnut Blight

Growth of Genomics

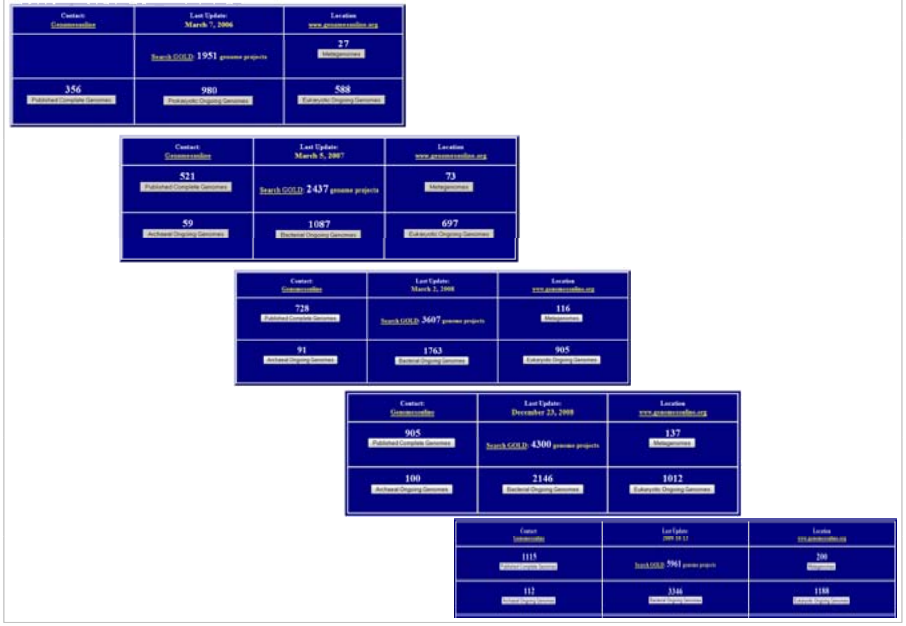


December 2008

GenBank
99,116,431,942 bp
98,868,465 entries

WGS
141,374,971,004 bp
48,394,838 entries

Growth of Genomics



Growth of Genomics

illumina

Solexa

Advancing genetic analysis one billion bases at a time.

Solexa Genome Analysis System

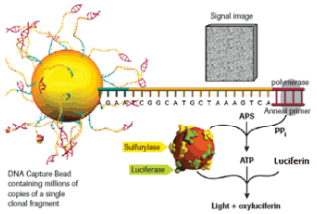
Applied Biosystems SOLiD™ System
— *THE* next-generation in genomic analysis

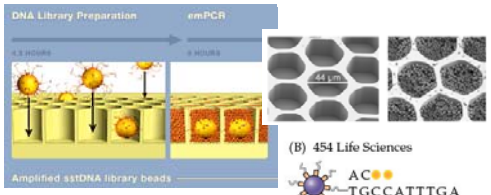
Platform	Read Lgth (bp)	Run Time (days/Gb)	DNA Frag / Run	Data / Run	Cost / Gp
Sanger	1000	500	96-384	75-300 Kb	$1M
454	450	1	500K	450 Mb	$20,000
Solexa	100	0.5	120M	25 Gb	$1000
SOLiD	50	0.5	500M	50 Gb	$1000

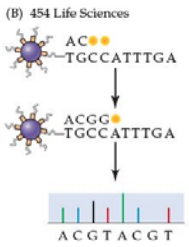
approximate values and costs

Roche 454 Pyrosequencing

- Fragment DNA
- Immobilize to bead
- Deposit in well of picoliter plate
- Amplify
- Add one nucleotide (A,T,C,or G).
- If binds, releases pyrophosphate which drives the production of light.
- The amount of light is proportional to the number of nucleotides bound (repeat stretch).
- Degrade unincorporated nucleotides, and starts again with a different nucleotide.



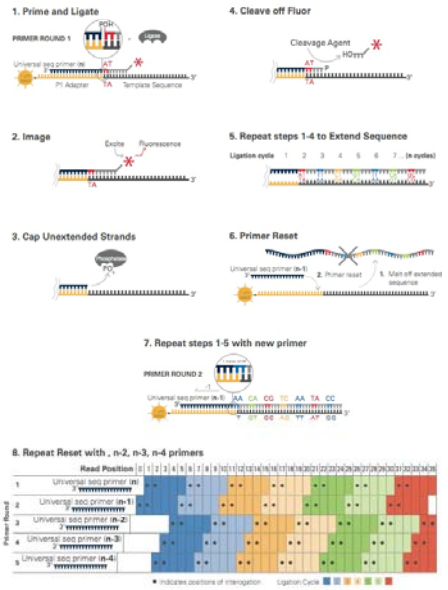




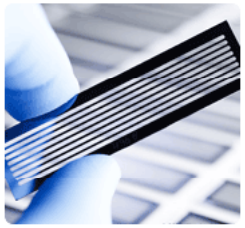
Applied Biosystems SOLiD (Supported Oligonucleotide + Ligation Detection)

Sequencing by Ligation

- Prepare amplified clonal bead populations
- Add sequencing primer.
- Add four fluorescently labeled di-base probes.
- Ligation to the primer will occur if first two bases match.
- Detect color tag.
- Cleave off tag and repeat cycles.

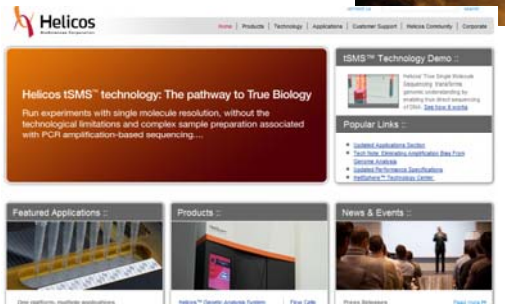


Illumina (Solexa) Genome Analyzer

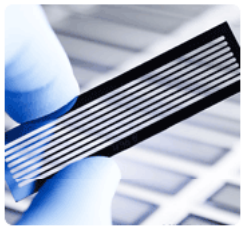


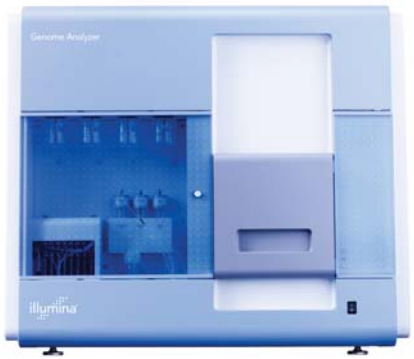
Future of Genomics

True direct sequencing from a single molecule of DNA or RNA



Illumina (Solexa) Genome Analyzer





Sequencing Process

1 *Library prep (~ 6 hrs)*

- Fragment DNA
- Repair ends / Add A overhang
- Ligate adapters
- Select ligated DNA

2 *Automated Cluster Generation (~ 5 hrs)*

1-8 samples

- Hybridize to flow cell
- Extend hybridized oligos
- Perform bridge amplification

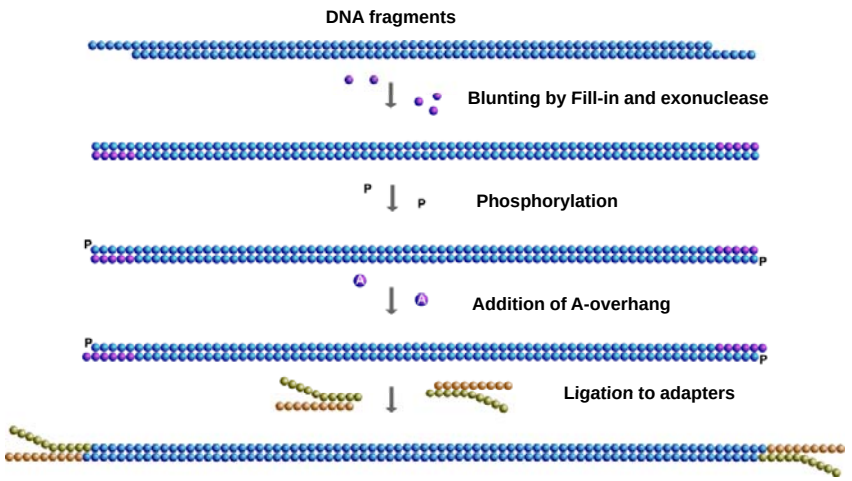
3 *Sequencing (~ 4 days*)*

1-8 samples

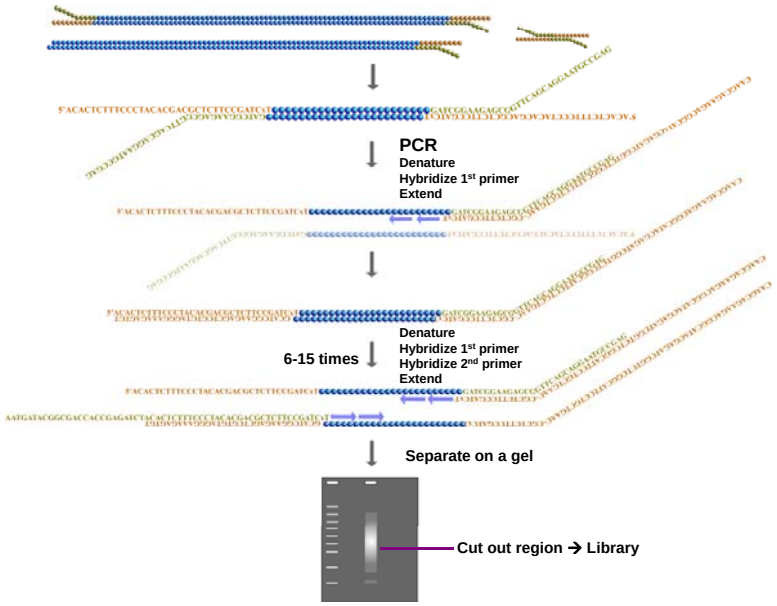
- Perform sequencing on forward strand
- Re-generate reverse strand
- Perform sequencing on reverse strand

* < 2 days for single read sequencing

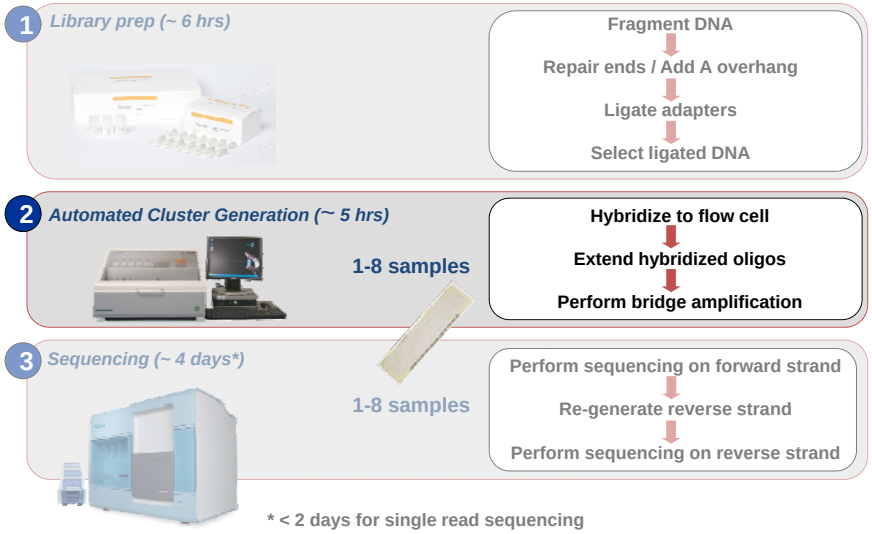
Genomic DNA Library Prep



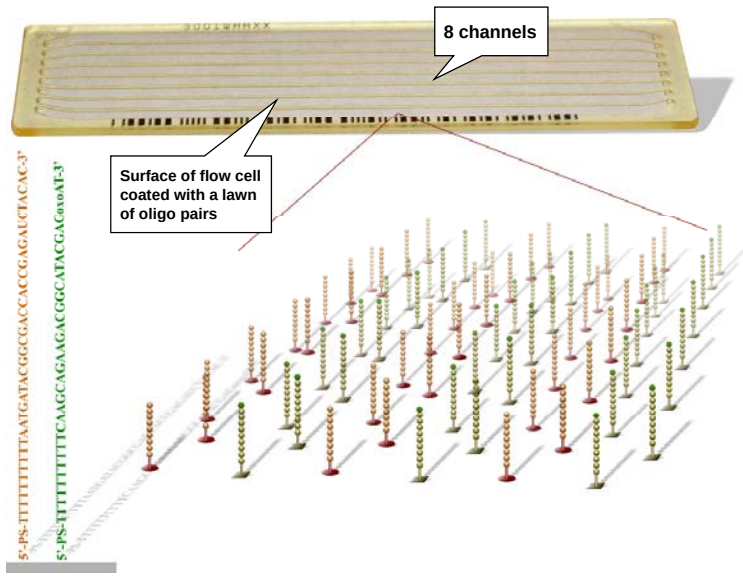
Selecting fragments attached to adapters



Sequencing Process

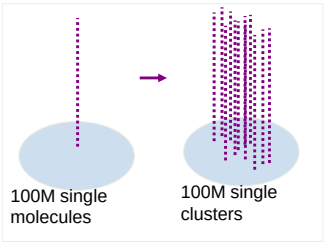


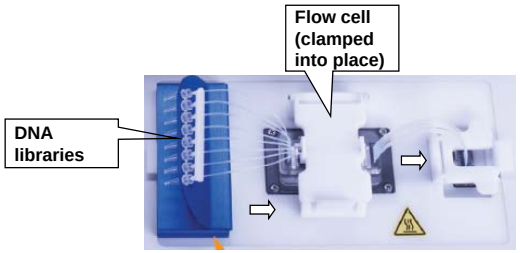
Flow cell



Cluster station

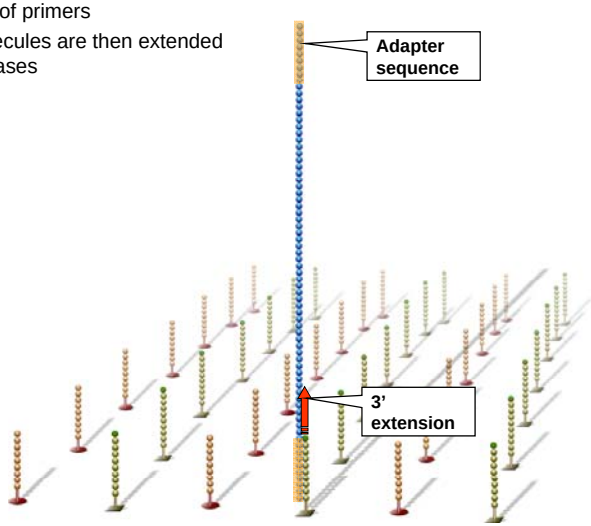
- Aspirates DNA samples into flow cell
- Automates the formation of amplified clonal clusters from the DNA single molecules





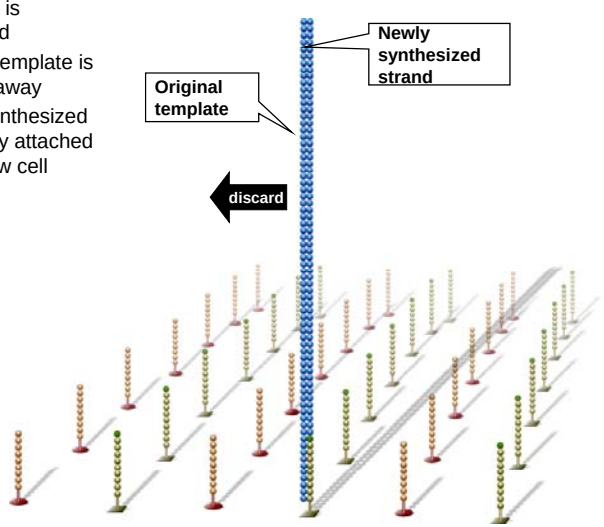
Cluster Generation: *Hybridize Fragment & Extend*

- > 100 M single molecules hybridize to the lawn of primers
- Bound molecules are then extended by polymerases

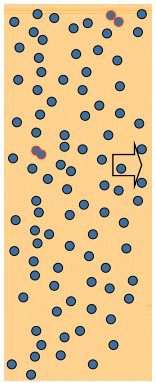


Cluster Generation: *Denature Double-stranded DNA*

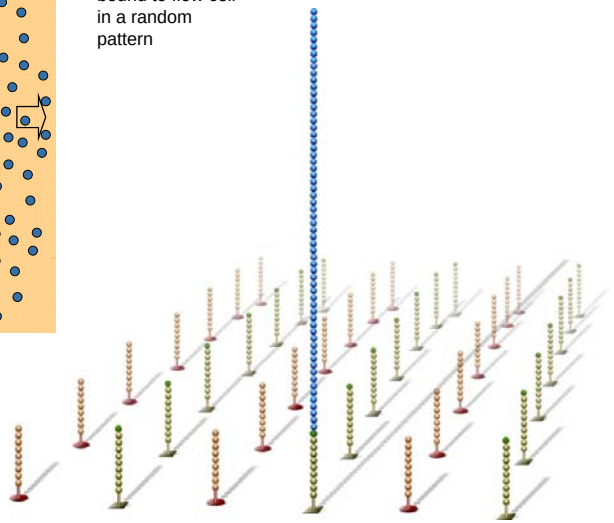
- Double-stranded molecule is denatured
- Original template is washed away
- Newly synthesized covalently attached to the flow cell surface



Cluster Generation: *Covalently-Bound Spatially Separated Single Molecules*

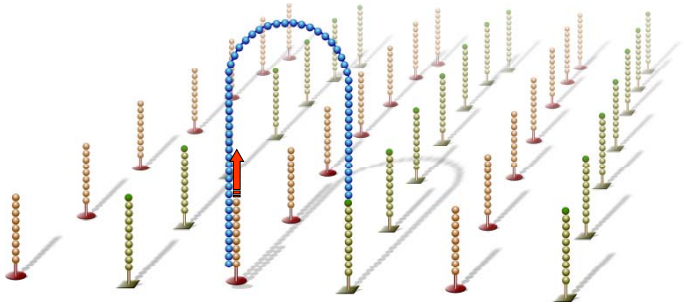


Single molecules bound to flow cell in a random pattern



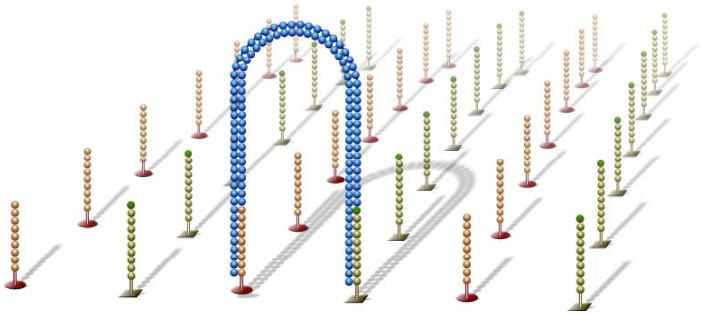
Cluster Generation: *Bridge Amplification*

- Single-strand flips over to hybridize to adjacent primers to form a bridge
- Hybridized primer is extended by polymerases



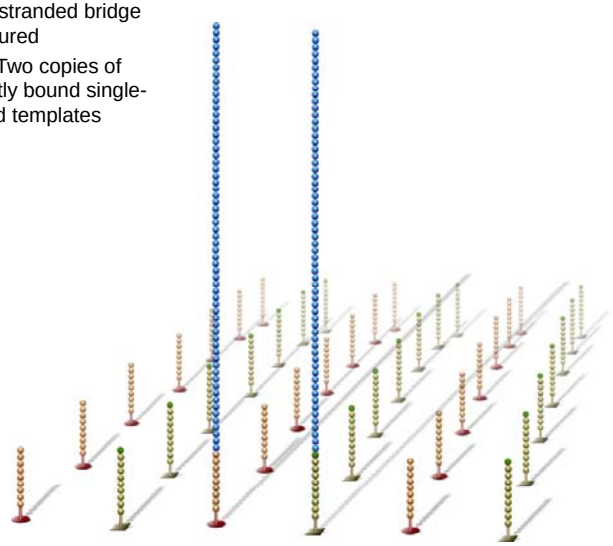
Cluster Generation: *Bridge Amplification*

- Double-stranded bridge is formed



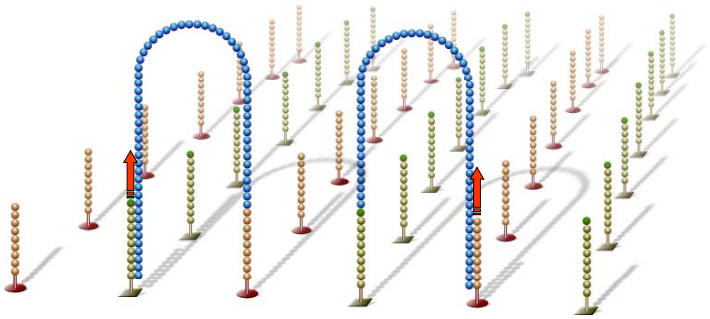
Cluster Generation: *Bridge Amplification*

- Double-stranded bridge is denatured
- Result: Two copies of covalently bound single-stranded templates



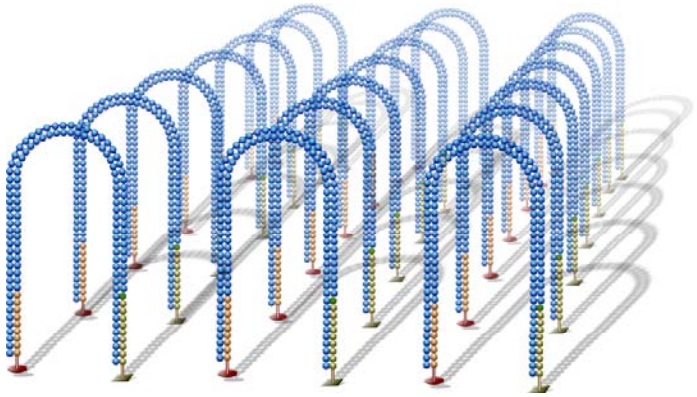
Cluster Generation: *Bridge Amplification*

- Single-strands flip over to hybridize to adjacent primers to form bridges
- Hybridized primer is extended by polymerase



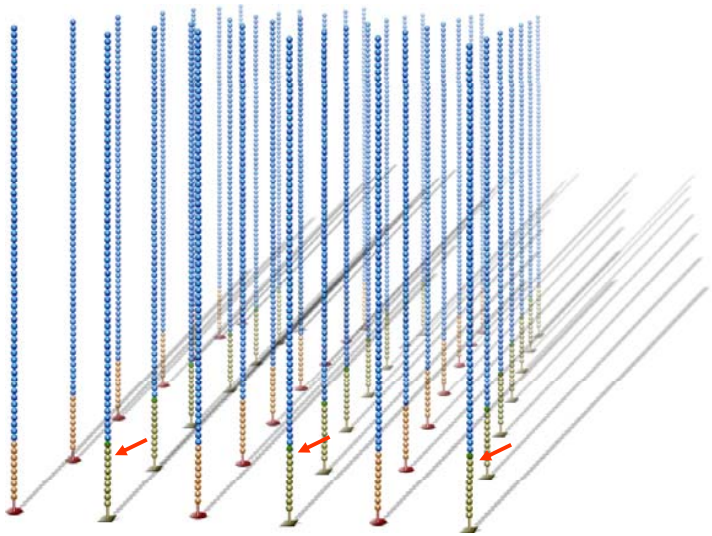
Cluster Generation: *Bridge Amplification*

- Bridge amplification cycle repeated until multiple bridges are formed



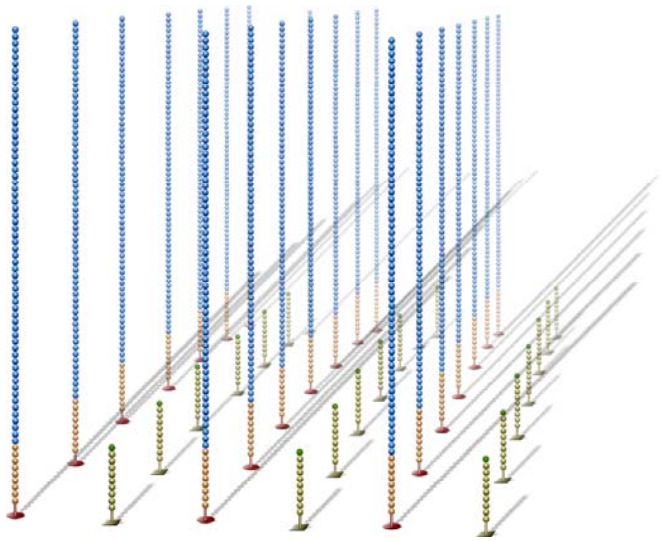
Cluster Generation

- dsDNA bridges denatured
- Reverse strands cleaved and washed away



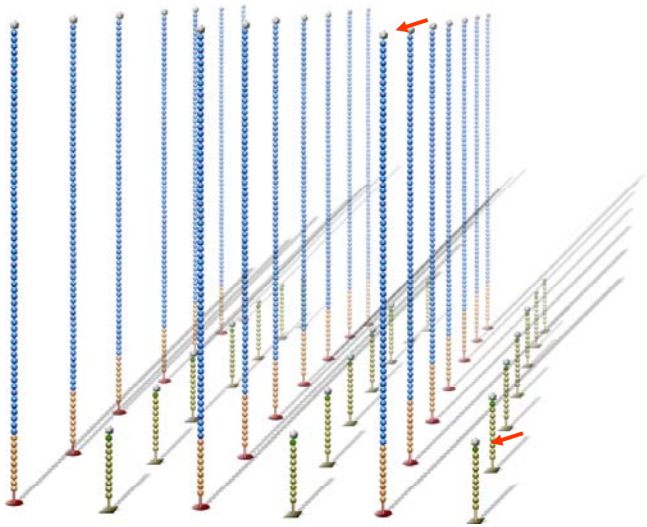
Cluster Generation

- …leaving a cluster with forward strands only



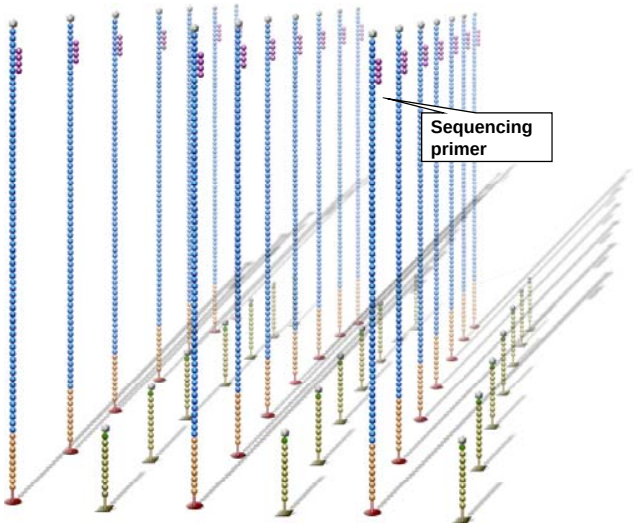
Cluster Generation

- Free 3' ends are blocked to prevent unwanted DNA priming



Sequencing

- Sequencing primer is hybridized to adapter sequence



Sequencing Process

1 *Library prep (~ 6 hrs)*

Fragment DNA
↓
Repair ends / Add A overhang
↓
Ligate adapters
↓
Select ligated DNA

2 *Automated Cluster Generation (~ 5 hrs)*

1-8 samples

Hybridize to flow cell
↓
Extend hybridized oligos
↓
Perform bridge amplification

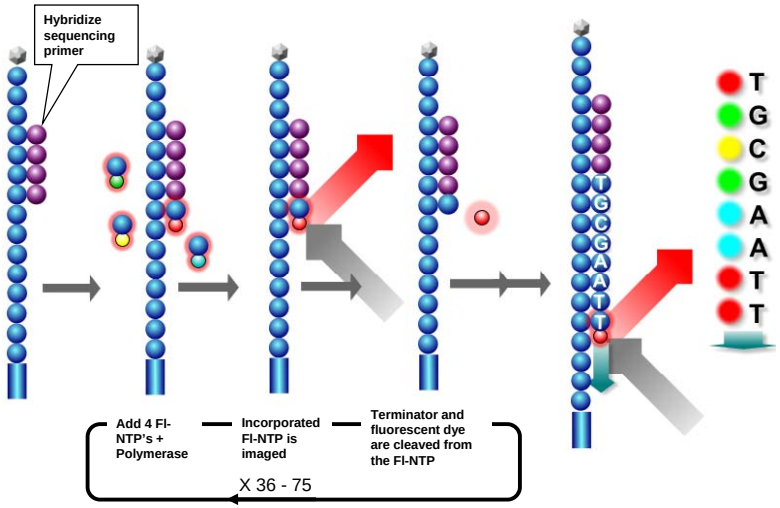
3 *Sequencing (~ 4 days*)*

1-8 samples

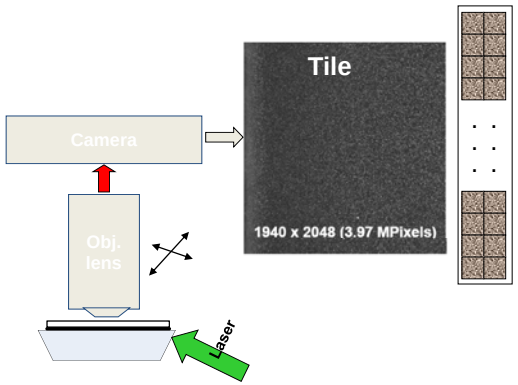
Perform sequencing on forward strand
↓
Re-generate reverse strand
↓
Perform sequencing on reverse strand

*** < 2 days for single read sequencing**

Sequencing



Genome Analyzer Imaging



3-4.5 TB/run

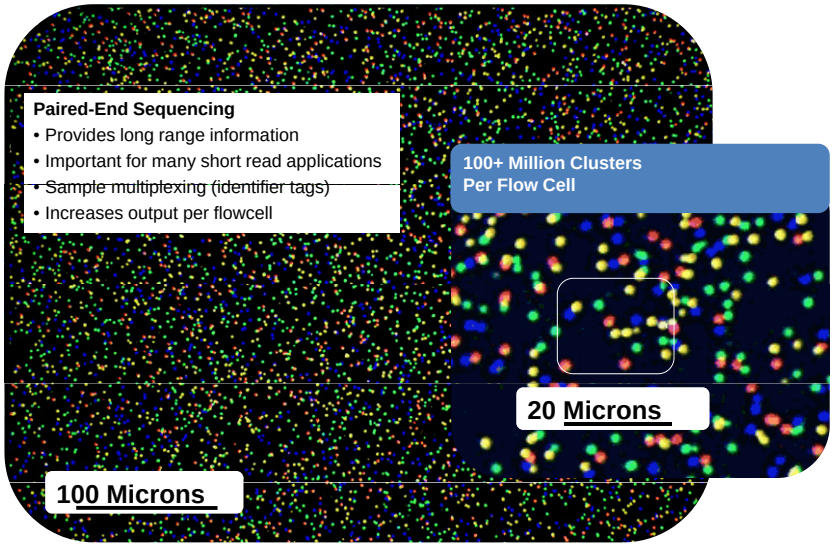
640,000 images x 7 MB/image

75-100 x 2 bases

4 images/base

8 channels/flow cell
2 columns/channel
50 tiles/column

Sequencing

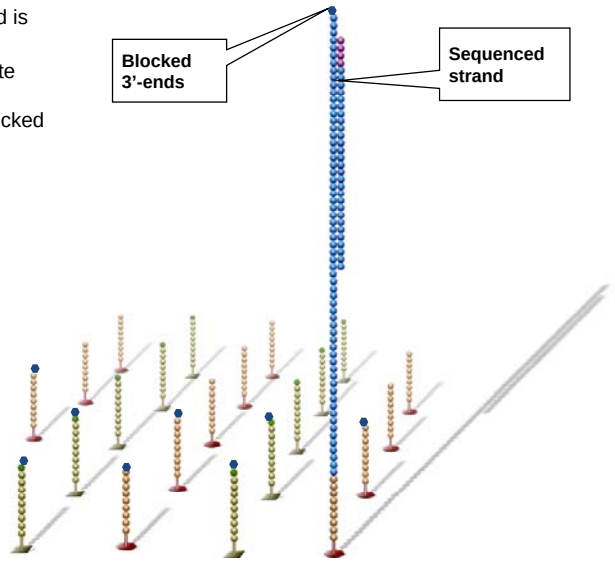


Sequencing with Paired Ends

Reference	This is really the best way to do sequencing
Single-reads	This is
…	is really
…	really the
…	the best
…	sequencing
Paired-reads	This is(----------26 characters----------) sequencing

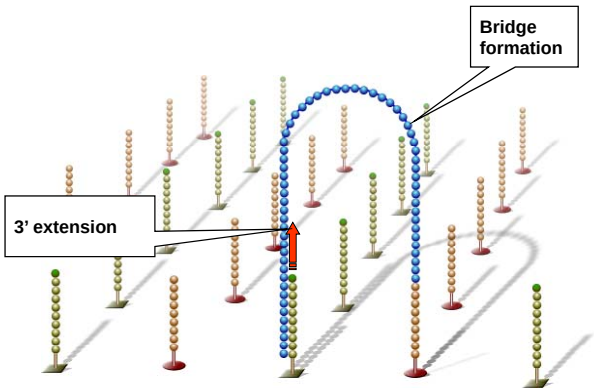
Paired End Sequencing

- Sequenced strand is stripped off
- 3'-ends of template strands and lawn primers are unblocked

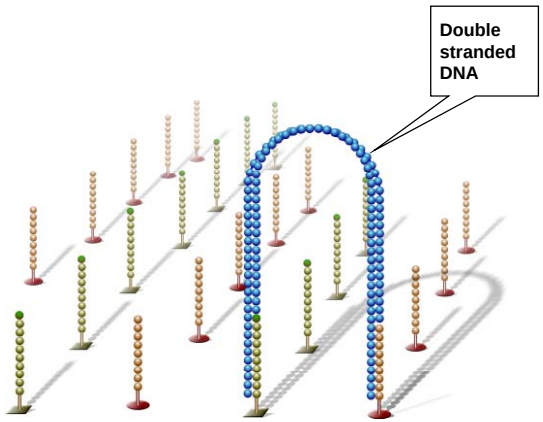


Paired End Sequencing

- Single-stranded template loops over to form a bridge by hybridizing with a lawn primer
- 3'-ends of lawn primer is extended

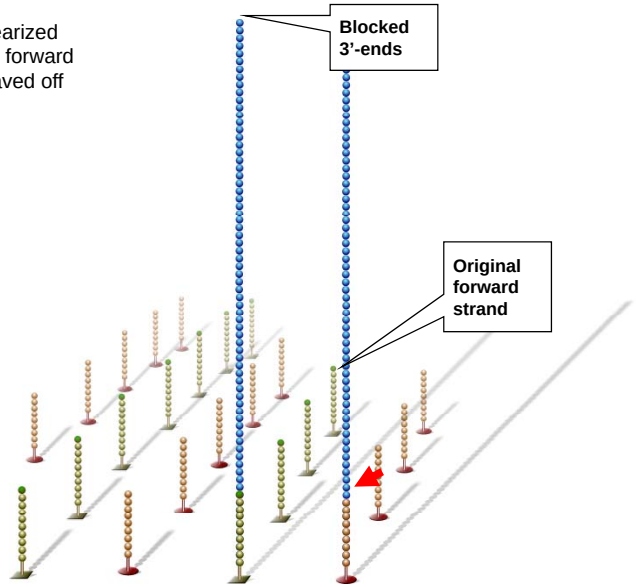


Paired End Sequencing



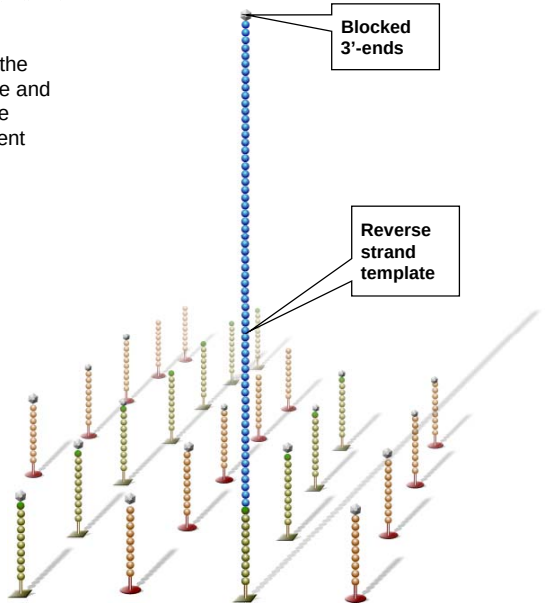
Paired End Sequencing

- Bridges are linearized and the original forward template is cleaved off



Paired end sequencing

- Free 3' ends of the reverse template and lawn primers are blocked to prevent unwanted DNA priming

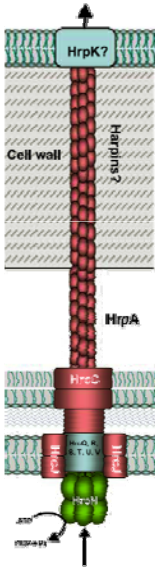


Special Topics in Genomics - The Hazelnut Blight Genome Project

Outline

- Next-gen genomics background
- Applications of next-gen genomics
  - *Pseudomonas syringae* genomics
  - Massively parallel analysis of type III effector interactions
  - Next-generation mapping of *Arabidopsis thaliana* cell wall accessibility mutants
- *P. syringae* pv. avellanae and Hazelnut Blight

Type III Secreted Effector Proteins



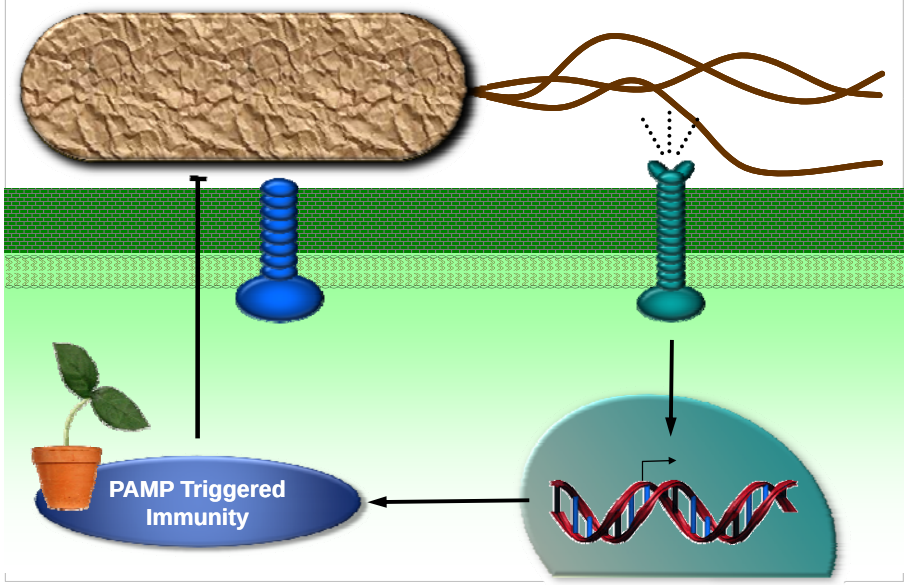
Type III Secretion System:

- animal pathogens:
  - *Escherichia coli*
  - *Salmonella enterica*
  - *Shigella spp.*
  - *Yersinia spp.*
  - etc.
- plant pathogens:
  - *Pseudomonas syringae*
  - *Xanthomonas spp.*
  - *Ralstonia spp.*
  - *Erwinia spp.*
  - etc.

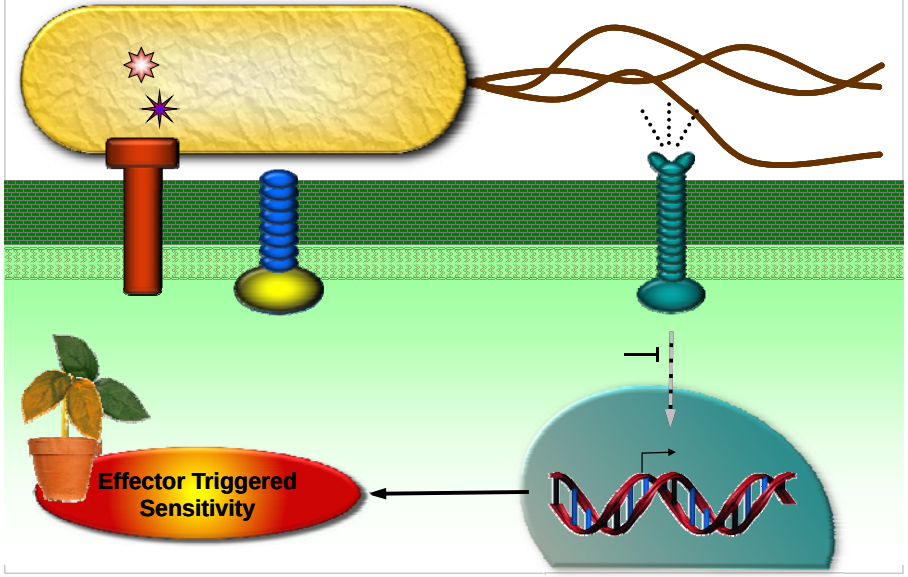
Type III Secreted Effectors:

- inhibition of innate immunity
- disruption of signal transduction
- cytoskeleton rearrangements
- cytotoxicity
- inducers of host immunity

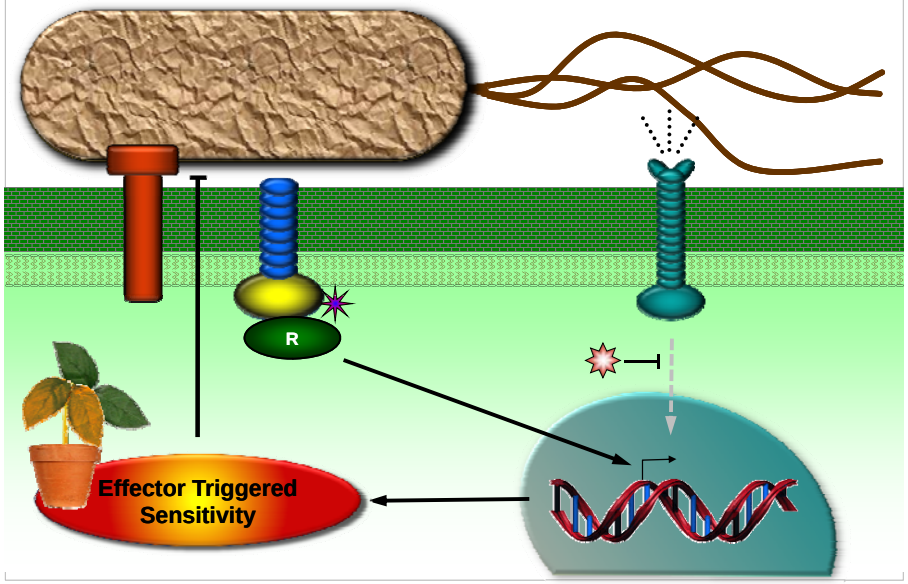
Host-Pathogen Interactions



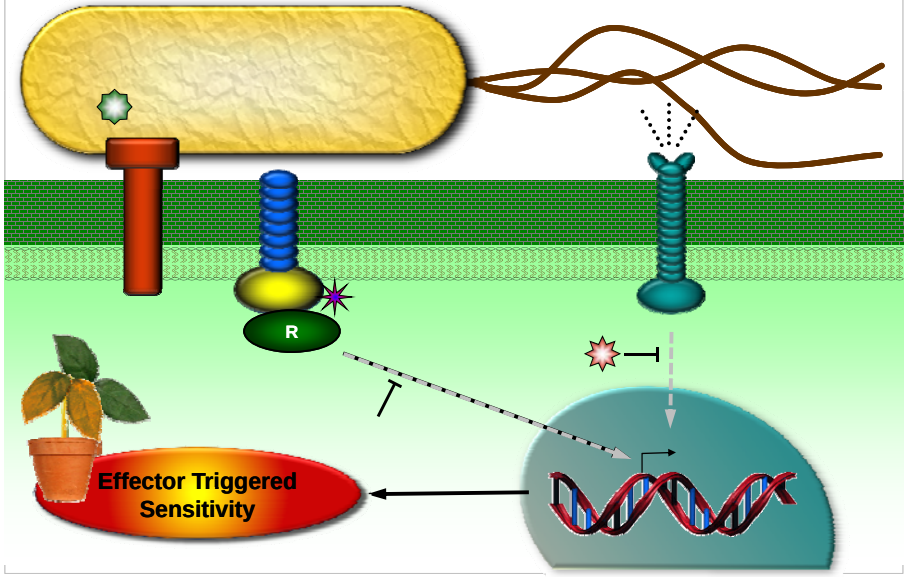
Host-Pathogen Interactions



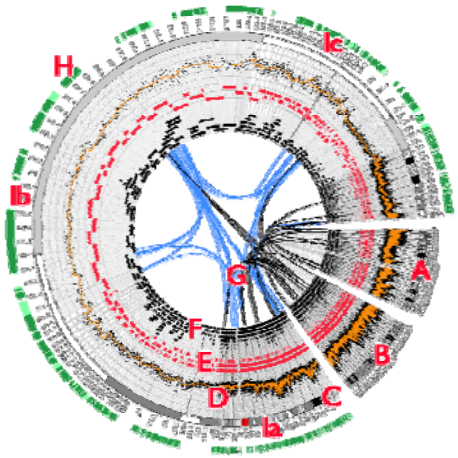
Host-Pathogen Interactions



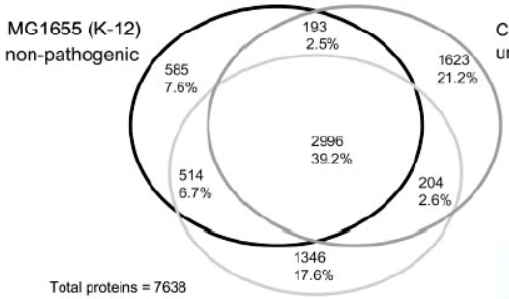
Host-Pathogen Interactions



Pseudomonas syringae Genomics

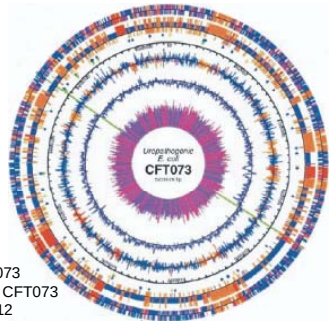


Dynamic Bacterial Genomes



Total proteins = 7638
2996 (39.2%) in all 3
911 (11.9%) in 2 out of 3
3554 (46.5%) in 1 out of 3

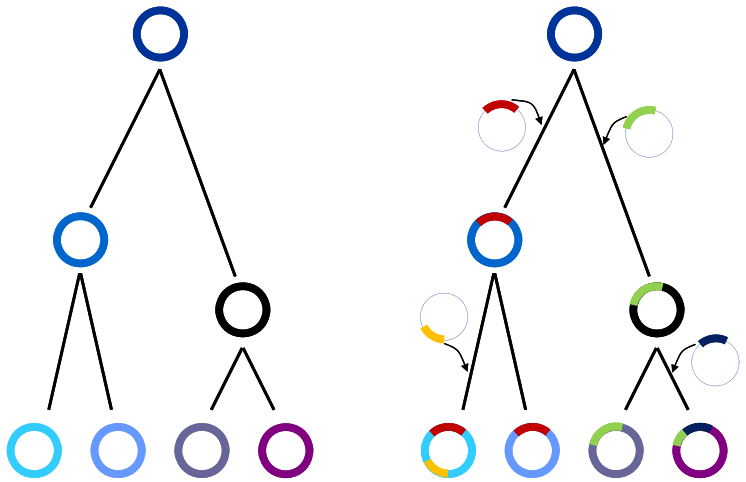
EDL933 (O157:H7)
enterohaemorrhagic



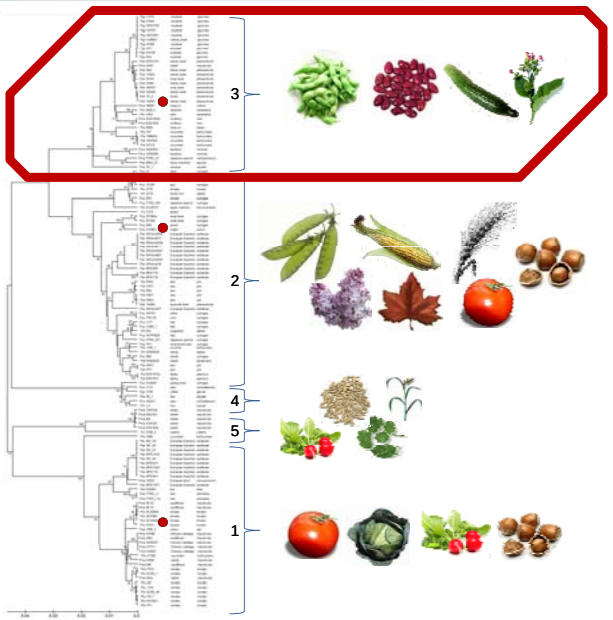
2nd circle
blue – backbone
red – insertions in CFT073
orange – substitution in CFT073
purple – insertions in K12

Welch, *et al.* (2002) PNAS 99:17020

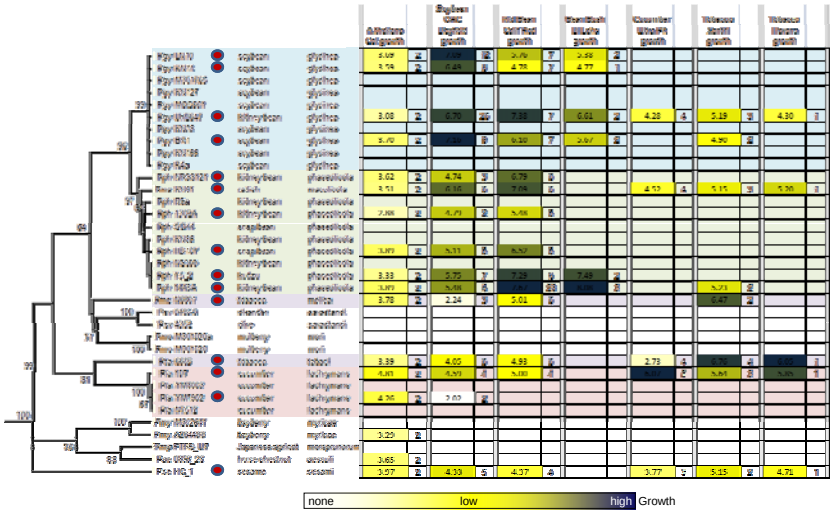
Vertical vs. Horizontal Evolution



The *P. syringae* Species Complex



Pseudomonas syringae phylogroup 3

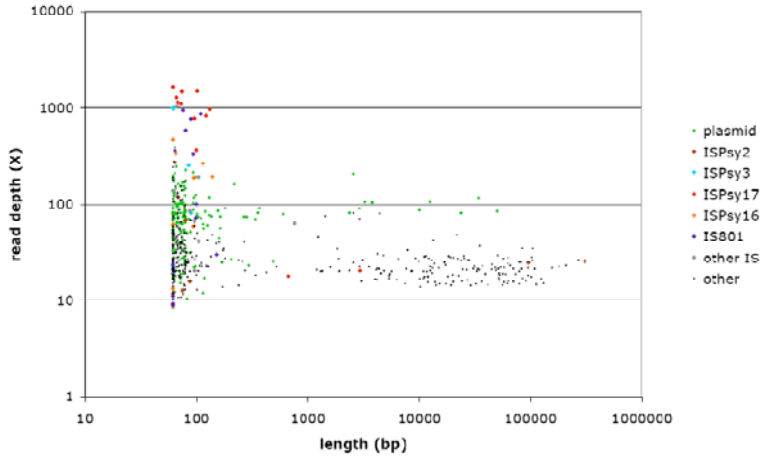


What evolutionary changes permitted or drove the diversification onto bean?

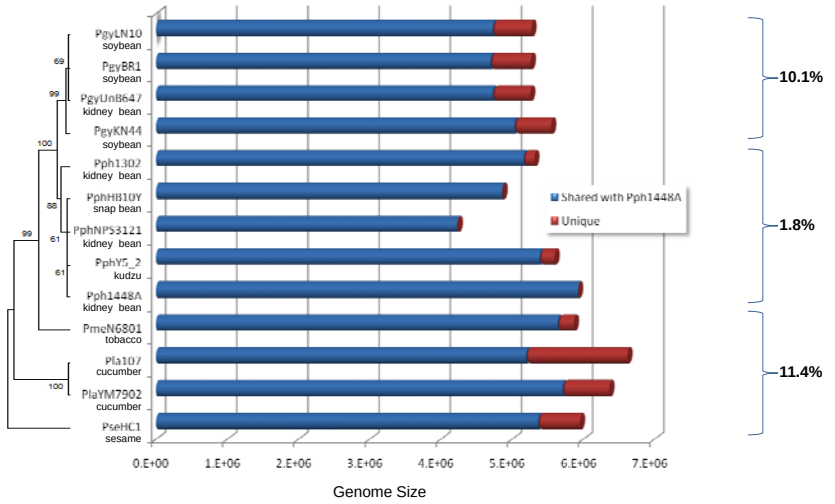
Genome Sequence Stats

Strain	Host	N Contigs	N50 (bp)	Max Contig (bp)
Pgy BR1	Soybean	1,492	43,010	140,733
PgyLN10	Soybean	1,548	38,674	145,485
PgyKN44	Soybean	1,387	33,571	139,678
PgyUnB647	Bean	1,154	38,237	140,574
Pph1302A	Bean	1,144	29,634	140,206
PphY5-2	Bean	738	41,333	278,663
PphNPS3121	Bean	736	42,628	155,374
PmeN6801	Tobacco	341	78,471	427,475
Pta6606	Tobacco	308	120,651	321,702
PseHC-1	Sesame	935	46,955	174,422
PmaES4326	Radish	675	57,962	298,778

Coverage of Pph1302A



Comparative Genomics of *P. syringae* phylogroup 3

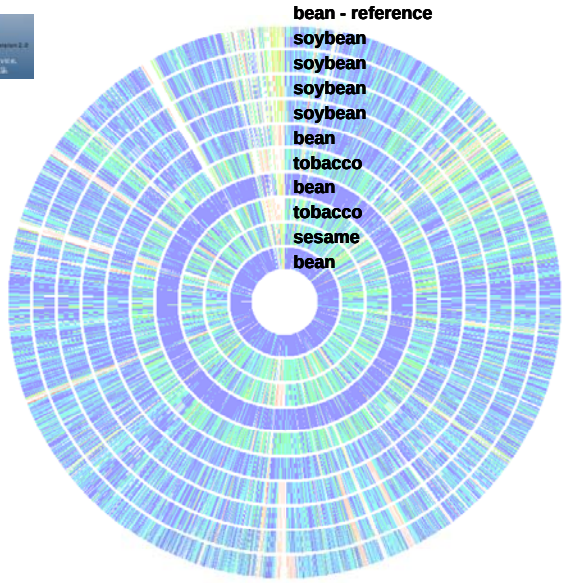


Comparative Genomics of *P. syringae* phylogroup 3

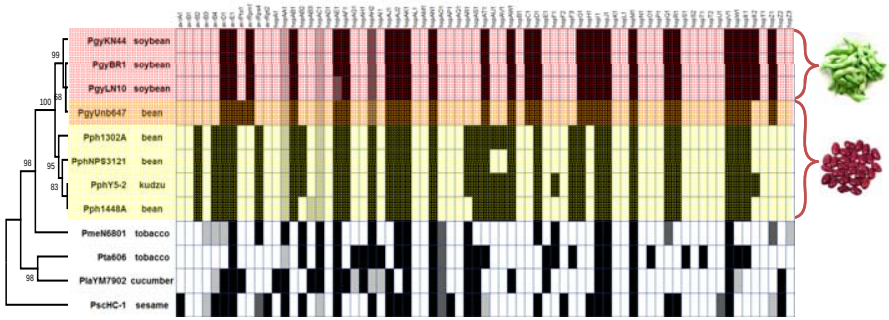
RAST Rapid Annotation using Subsystem Technology

Reference genome:
***P. syringae* pv. phaseolicola 1448A**

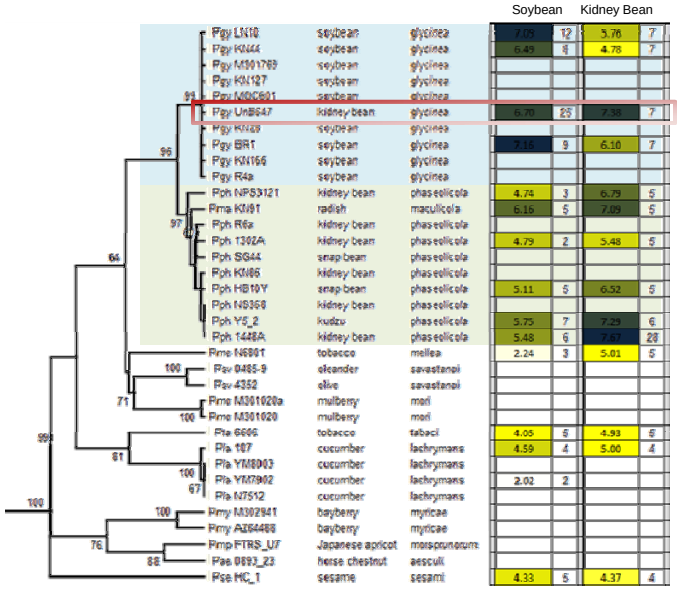
***P. syringae* pv. glycinea UnB647**
***P. syringae* pv. glycinea BR1**
***P. syringae* pv. glycinea KN44**
***P. syringae* pv. glycinea LN10**
***P. syringae* pv. phaseolicola Y5-2**
***P. syringae* pv. mellea N6801**
***P. syringae* pv. phaseolicola NPS3121**
***P. syringae* pv. tabaci 6606**
***P. syringae* pv. sesami HC-1**
***P. syringae* pv. phaseolicola 1302A**
(from outermost to innermost circles)



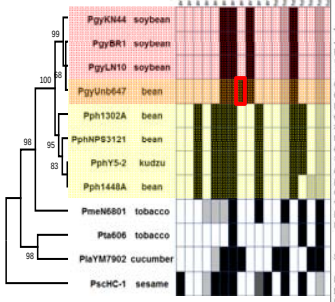
Type III Effector Profiles



Host Specific Growth



Type III Effector Profiles



Current Biology 18, 74–80, January 8, 2008 ©2008 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2007.12.020

Report

Pseudomonas syringae Effector AvrPto Blocks Innate Immunity by Targeting Receptor Kinases

Tingting Xiang,[1,2,5] Na Zong,[2,5] Yan Zou,[2,3,5] Yong Wu,[2] Jie Zhang,[2] Weimao Xing,[2] Yan Li,[2] Xiaoyan Tang,[4] Lihuang Zhu,[2] Jijie Chai,[2] and Jian-Min Zhou[2,*]
[1]State Key Laboratory of Plant Physiology and Biochemistry
College of Biological Sciences
China Agricultural University
Beijing 100094
China
[2]National Institute of Biological Sciences
Beijing 102206
China
[3]Institute of Genetics
Chinese Academy of Sciences
Beijing 100864
China
[4]Department of Plant Pathology
Kansas State University
Manhattan, Kansas 66506

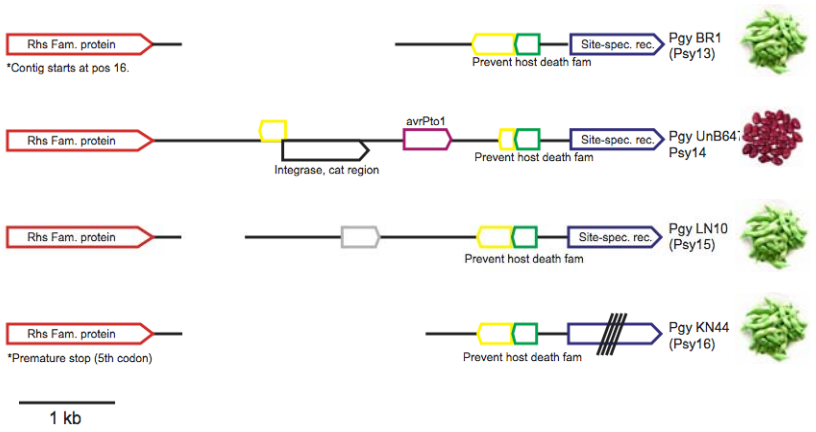
Summary

Plants use receptor kinases, such as FLS2 and EFR, to perceive bacterial pathogens and initiate innate immunity. This immunity is often suppressed by bacterial effectors, allowing pathogen propagation. To counteract, plants have evolved disease resistance genes that detect the bacterial effectors and reinstate resistance. The *Pseudomonas syringae* effector AvrPto promotes infection in susceptible plants but triggers resistance in plants carrying the protein kinase Pto and the associated resistance protein Prf. Here we show that AvrPto binds receptor kinases, including *Arabidopsis* FLS2 and EFR and tomato LeFLS2, to block plant immune responses in the plant cell. The ability to target receptor kinases is required for the virulence function of AvrPto in plants. The FLS2-AvrPto interaction and Pto-AvrPto interaction appear to share similar sequence requirements, and Pto competes with FLS2 for AvrPto binding. The results suggest that the mechanism by which AvrPto recognizes virulence targets is linked to the evolution of Pto, which, in association with Prf, recognizes the bacterium and triggers strong resistance.

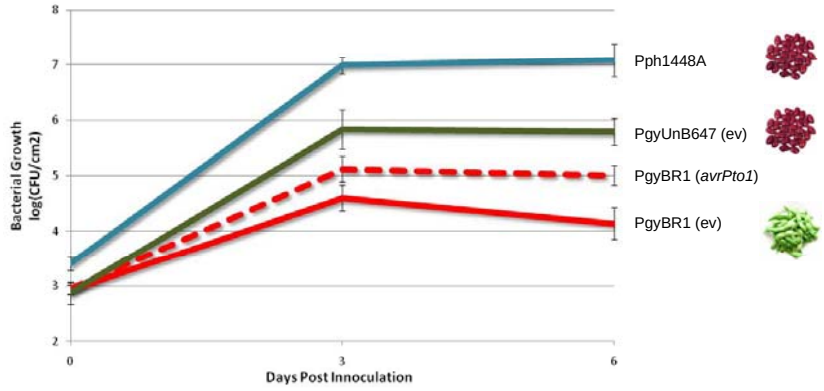
effector proteins that are delivered into plant cells through the type III secretion system to inhibit the PAMP-triggered immunity (PTI) [3]. Plants, in turn, have adapted to effector-mediated virulence by evolving resistance proteins that directly or indirectly detect the effectors inside the plant cell and trigger strong immune responses (effector-triggered immunity; ETI) [3] that often give rise to complete resistance to the invading pathogen. However, little is known about virulence targets of the effectors and how effector virulence activities have shaped the evolution of resistance proteins. In tomato plants, the resistance protein Pto directly interacts with the *Pseudomonas syringae* effector AvrPto inside the plant cell to activate ETI [4]. Interestingly, Pto resembles the cytoplasmic kinase domain of receptor-like kinases but lacks the extracellular leucine-rich-repeat domain [5]. Pto constitutively interacts with Prf, a typical resistance protein that contains a nucleotide-binding domain and a leucine-rich-repeat domain and is crucial for the activation of ETI [6]. In plants lacking Pto or Prf, AvrPto inhibits PTI and enhances bacterial virulence [7]. A recent report showed that AvrPto acts upstream of the mitogen-activated protein kinase (MAPK) cascade to inhibit the PTI signaling pathway, but the target(s) remains to be identified [8]. AvrPto is plasma-membrane-localized through myristic-acid modification at the N terminus [9], but its biochemical function is unknown.

To understand the biochemical function of AvrPto, we have solved the crystal structure of the Pto-AvrPto complex [10]. Structural and biochemical analyses demonstrated that AvrPto is an inhibitor of the Pto kinase. AvrPto occupies the P+1 loop and blocks the access of Pto to its substrates. We hypothesize that the kinase-inhibition ability is intrinsic to the virulence function of AvrPto, namely the interference with PTI by blocking PRRs. Two well-known PRRs in plants are FLS2 and EFR, which perceive bacterial flagellar peptide flg22 and EF-Tu peptide elf26, respectively [11, 12]. To determine whether FLS2 and EFR kinases might be structurally related to Pto, we searched Protein Data Bank (PDB) with the primary sequences. Among the known structures of serine/threonine kinases, IRAK-4 is most closely related to Pto, FLS2, and EFR kinases, particularly in the P+1 loop (Figure S1 available online).

avrPto1

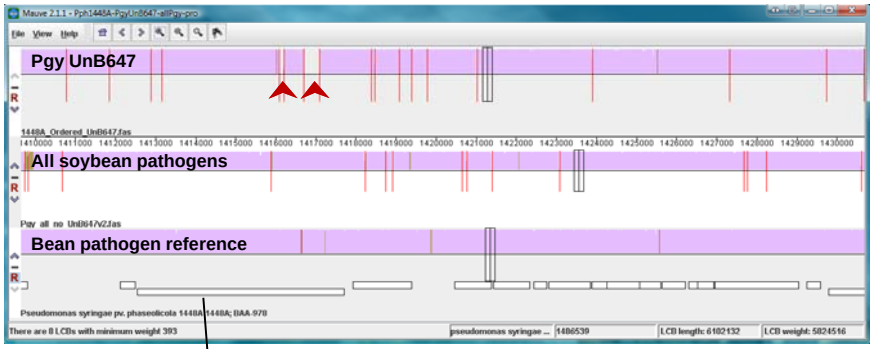


AvrPto1 as Bean Specificity Factor



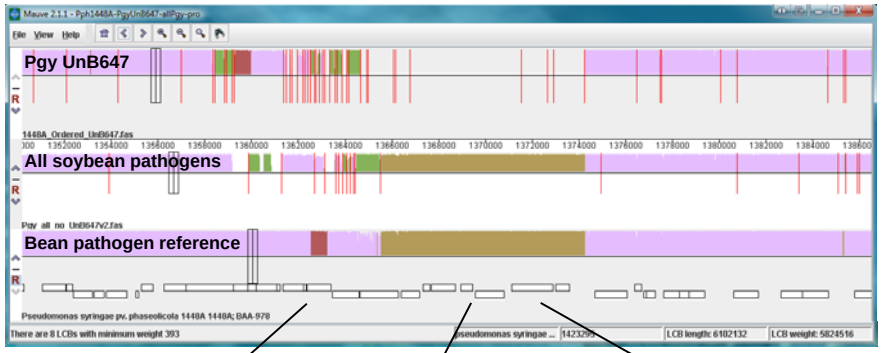
Kidney Bean vs. Soybean Pathogens

Pph1448A reference genome vs. PgyUnB647 & pooled Pgy genomic data
type III effector

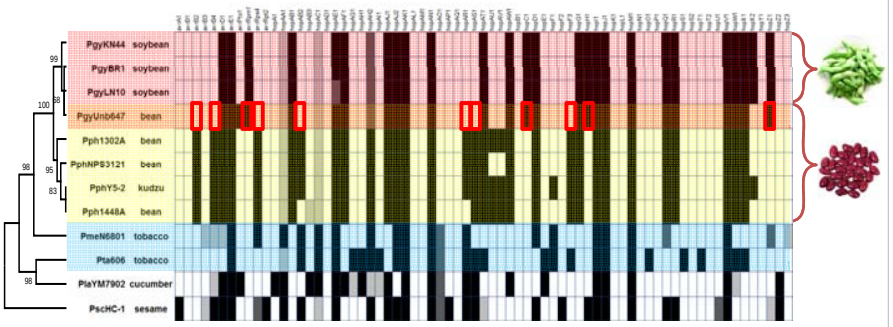


Kidney Bean vs. Soybean Pathogens

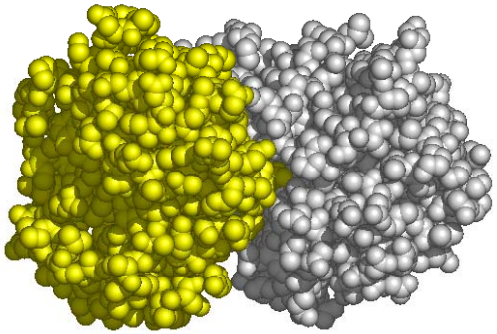
Pph1448A reference genome vs. PgyUnB647 & pooled Pgy genomic data
strain-specific genes



Type III Effector Profiles

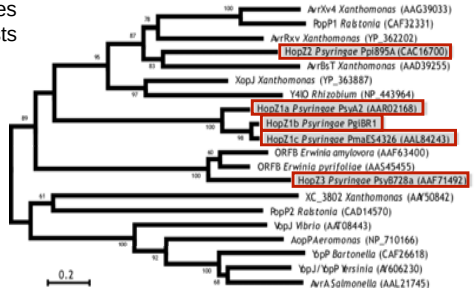


Massively Parallel Analysis of Type III Effector Interactions



Pseudomonas syringae HopZ Family

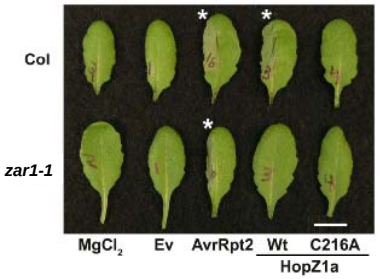
- Member of the broadly distributed YopJ Effector family
- Widespread throughout the *P. syringae* species complex
  - Diversified by both mutation accumulation and horizontal gene transfer
- Three major allele classes in *P. syringae*
  - Different alleles result in different host-specific interactions
  - Alleles are under strong positive selection
  - Some alleles promote bacterial growth *in planta*
  - The oldest allele (HopZ1a) induces an immune response in most hosts
- HopZ1a is recognized by the *Arabidopsis thaliana* ZAR1 CC-NBS-LRR resistance protein
  - Screen ARTIC R-gene collection
  - Positional cloning

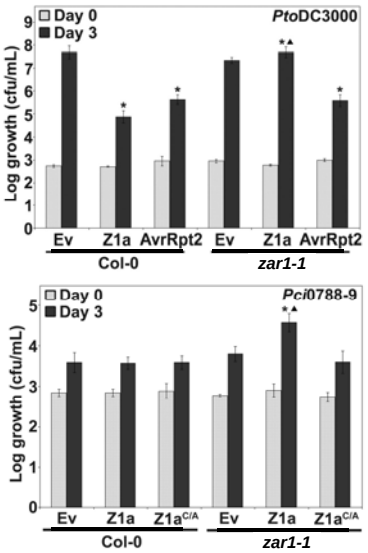


Arabidopsis R Gene T-DNA Insertion Collection (ARTIC)

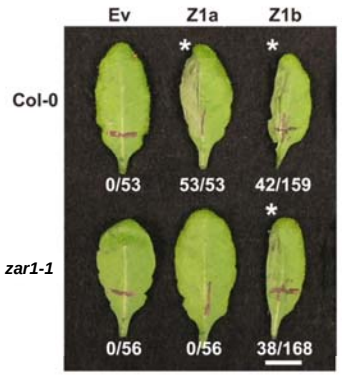
- Collection of publically available T-DNA insertion lines for all predicted *A. thaliana* Col-0 *R* genes
- Preference given to T-DNA insertion lines with a high confidence insertion in gene of interest, and preferably in an exon near the beginning of the gene.
- T-DNA insertions available for 166 / 170 predicted *R* genes.
  - Homozygous
    - 118 Salk, 13 Sail ,1 WiscDsLox
  - Heterozygous
    - 17 Salk, 7 Sail

HopZ1a Interaction with ZAR1



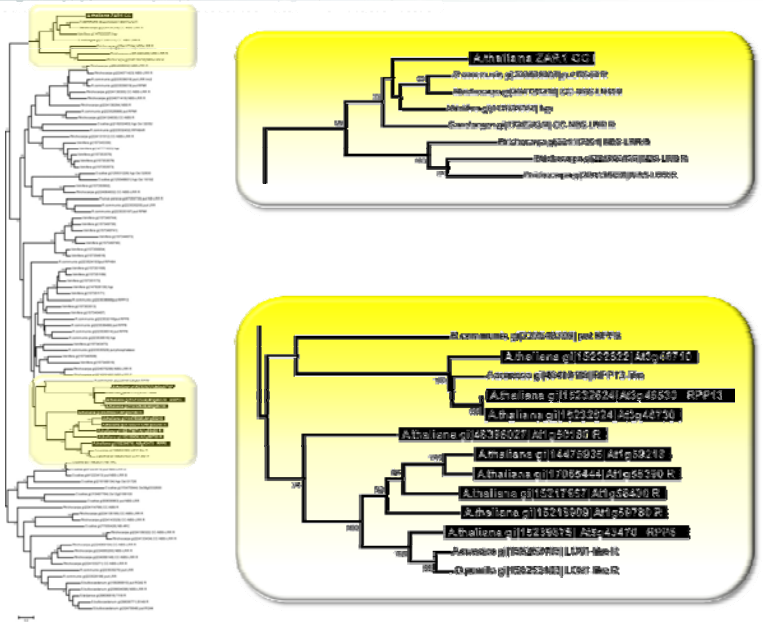


ZAR1 Does Not Recognize HopZ1b



HopZ1a – HopZ1b
75% nucleotide identity
72% amino acid identity

ZAR1 Coiled Coil Domain Evolutionary Relationships



Defense Signaling Pathways

TIR-NBS-LRR R Proteins

RPP2, RPP4, RPP5, RPP21, RPS4

CC-NBS-LRR R Proteins

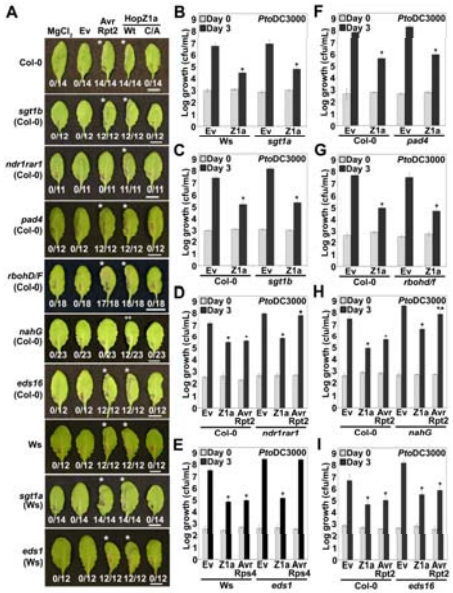
RPS2, RPM1, RPS5

EDS1, PAD4

NDR1

HR

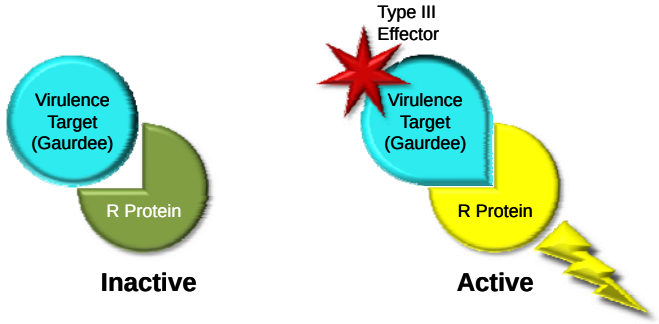
Immune responses



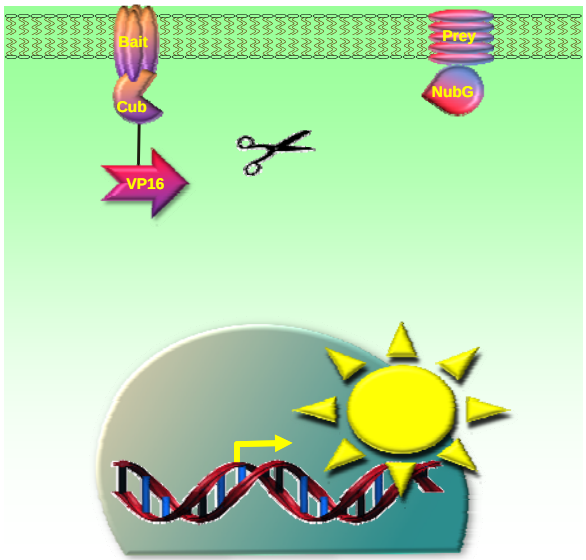
Interactor Screen

Massively Parallel Split Ubiquitin Yeast 2-Hybrid Screen

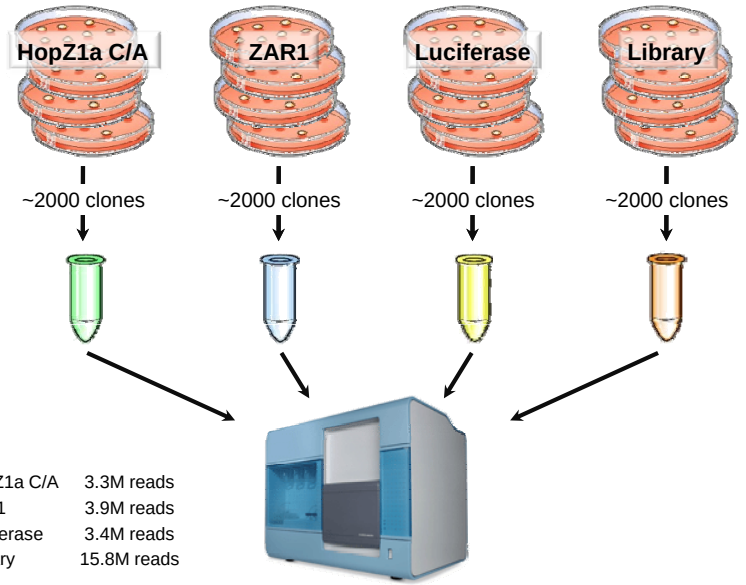
- Virulence targets of HopZ1a
- Guardees ZAR1



Split Ubiquitin Yeast 2-Hybrid Screen



Massively Parallel Split Ubiquitin Yeast 2-Hybrid Interactor Screen

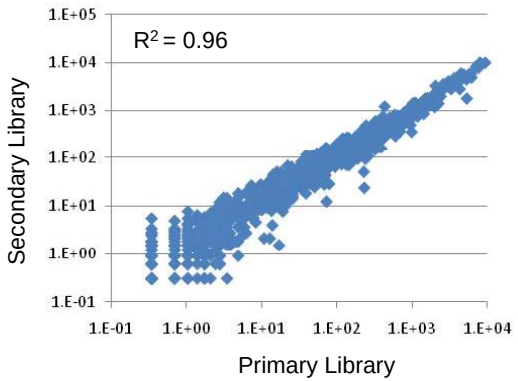


HopZ1a C/A	3.3M reads
ZAR1	3.9M reads
Luciferase	3.4M reads
Library	15.8M reads

Interaction Data (hits / million)

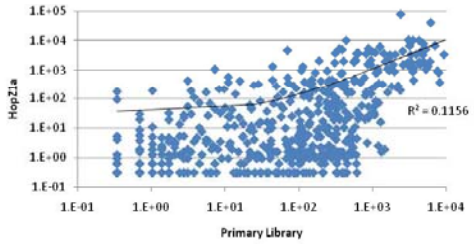
Description	HopZ1a	ZAR1	Library	Luciferase
NPL4 family protein	76,328.7	2,727.3	2,410.8	14,592.9
APG1 (ALBINO OR PALE GREEN MUTANT 1); methyltransferase	39,720.2	6,587.0	6,162.3	6,717.7
calmodulin-binding protein	10,475.6	4,492.5	3,093.3	365.7
pectate lyase family protein	10,158.5	8,432.4	3,557.5	982.4
protein kinase, putative	10,059.8	13,779.9	4,283.1	4,877.8
FLA16 (FASCICLIN-LIKE ARABINOGALACTAN PROTEIN 16	**10,017.1**	**0.0**	**438.3**	**73.7**
ATGPAT4/GPAT4 1	8,606.7	2,539.8	3,266.1	707.4
tudor domain-containing protein / nuclease family protein	7,868.0	4,020.6	4,877.9	728.1
nucleic acid binding	6,731.4	8,609.6	7,053.0	4,450.1
ERD1 (EARLY RESPONSIVE TO DEHYDRATION 1)	6,596.7	9,864.9	3,186.6	2,199.4
PAL2 (phenylalanine ammonia-lyase 2); phenylalanine ammonia-lyase	6,116.1	1,033.1	1,264.7	708.6
oxidoreductase, 2OG-Fe(II) oxygenase family protein	**5,346.0**	**4.1**	**876.3**	**80.1**
ADT4 arogenate dehydratase/ prephenate dehydratase	5,202.5	258.6	354.4	25.6
DIN4 (DARK INDUCIBLE 4)	5,139.0	8,812.6	5,288.0	6,188.6
MIF4G domain-containing protein / MA3 domain-containing protein	4,929.3	2,002.9	386.7	17.5
peptide methionine sulfoxide reductase, putative	**4,650.0**	**0.3**	**69.8**	**21.0**
PRP4 (PROLINE-RICH PROTEIN 4)	4,376.7	3,428.4	2,576.9	4,571.0

Primary & Secondary Library

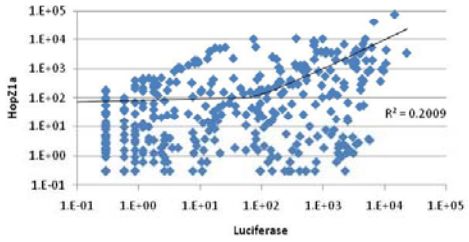


Bait – Library Association

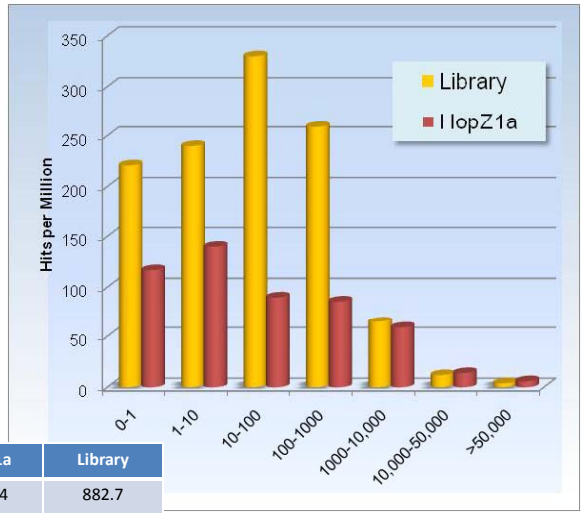
HopZ1a



Luciferase



Hit Frequency



	HopZ1a	Library
Mean	1949.4	882.7
Median	9.7	23.7
t-test	P=0.004	

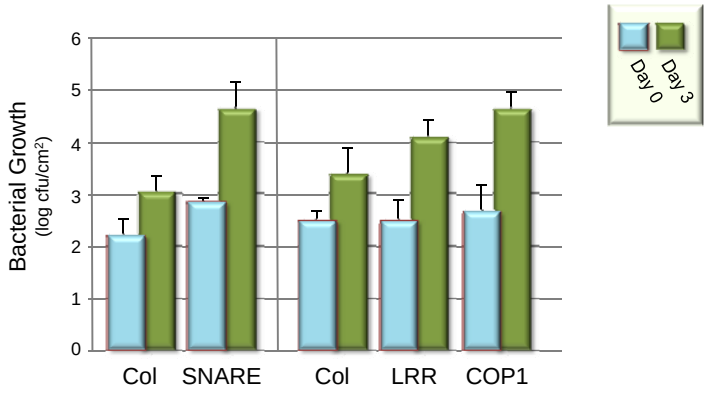
Putative Interactors

Putative Interactor	Value		Hits (per million)			
	HopZ1a C/A	ZAR1	HopZ1a C/A	ZAR1	Luc	Lib
SNARE-associated protein	**0.99**	**0.96**	429.3	155.7	1.8	2.5
Leucine-rich repeat protein	**0.98**	**0.00**	170.6	9.9	1.2	0.6
COP1-interacting protein	**0.96**	**0.00**	378	0	1.8	12.0

value = (Bait – Luciferase) / (Bait + Luciferase + Library)

HopZ1a C/A	3.3M reads
ZAR1	3.9M reads
Luciferase	3.4M reads
Library	15.8M reads

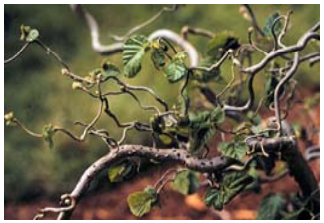
Virulence Affects of Interactor Knockouts



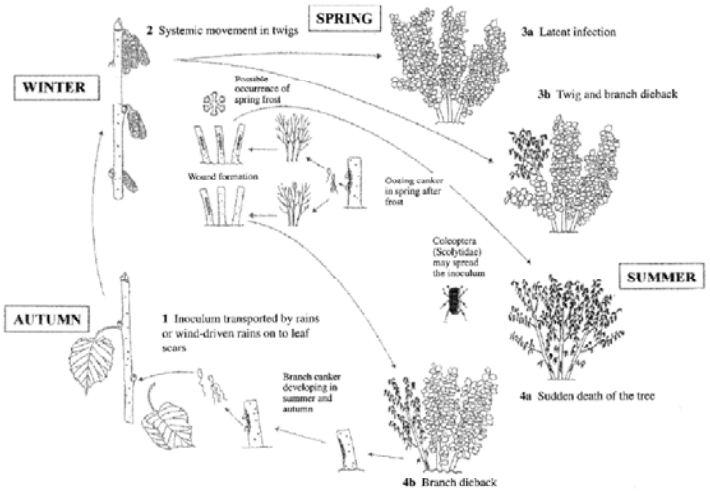
Growth assays
- Mildly virulent PmaM6ΔE
- Homozygous T-DNA knockout lines

Three putative interactors that positively regulate plant immunity

P. syringae pv. avellanae & Hazelnut Blight



P. syringae pv. avellanae & Hazelnut Blight



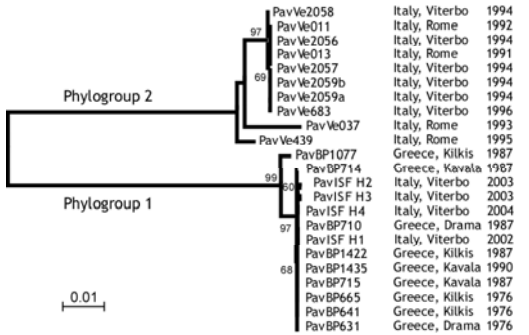
P. syringae pv. avellanae & Hazelnut Blight

Table 1. Pav strains used in this study

Strain*	Origin	Year of isolation
BPIC 631	Greece/Drama	1976
BPIC 641	Greece/Kilkis	1976
BPIC 665	Greece/Kilkis	1976
BPIC 710	Greece/Drama	1987
BPIC 714	Greece/Kavala	1987
BPIC 715	Greece/Kavala	1987
BPIC 1077	Greece/Kilkis	1987
BPIC 1422	Greece/Kilkis	1987
BPIC 1435	Greece/Kavala	1990
ISPaVe 013	Italy/Latium/Rome	1991
ISPaVe 011	Italy/Latium/Rome	1992
ISPaVe 037	Italy/Latium/Rome	1993
ISPaVe 439	Italy/Latium/Rome	1995
ISPaVe 2056	Italy/Latium/Viterbo	1994
ISPaVe 2057	Italy/Latium/Viterbo	1994
ISPaVe 2058	Italy/Latium/Viterbo	1994
ISPaVe 2059	Italy/Latium/Viterbo	1994
ISPaVe 683	Italy/Latium/Viterbo	1996
ISF H1	Italy/Latium/Viterbo	2002
ISF H2	Italy/Latium/Viterbo	2003
ISF H3	Italy/Latium/Viterbo	2003
ISF H4	Italy/Latium/Viterbo	2004

*All strains were isolated from diseased tissue and confirmed to be pathogenic on hazelnut.

P. syringae pv. avellanae & Hazelnut Blight



P. syringae pv. avellanae & Hazelnut Blight