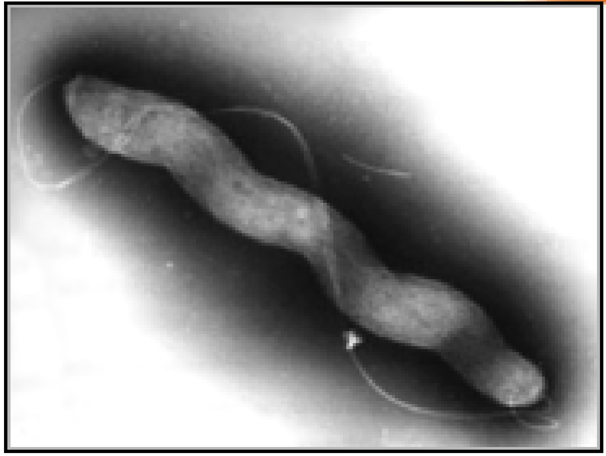


Using array-based expression profiling and comparative genomics to gain insights on bacterial virulence

Ed Taboada

Pathogen Genomics Group

NRC - Institute for Biological Sciences

C. jejuni is an important food pathogen

- Grows under micro-aerophilic conditions
  - Delays diagnosis
- Commensal organism in many domestic animals
  - Ubiquitous in commercial poultry
- Causes food poisoning in humans
  - leading cause of enteritis in the (1st) world
  - More cases than *E. coli* and *Salmonella* combined
- Secondary effects:
  - Auto-immune disorders (GBS, MFS)

Your Guide To Food Poisoning

When you sit down to enjoy a good meal, the first thing on your mind is not whether or not your food is going to send you to the ER. Find out which foods are prime food poisoning suspects.

Campylobacter jejuni

This organism is the most commonly identified cause of diarrhea in the world.

How it's spread: Raw or undercooked chicken, or its juices

Get Tons of Useful Stuff from AskMen.com

Dating & Love

Answering your questions: Do women kiss on the first date?

Great tips on dating: The top ten signs you're a bad kisser.

Understanding the pathogen…

- What are the molecules / proteins / pathways necessary for virulence ?
  - Surface structures (adhesion, invasion, host evasion)
  - Iron acquisition
  - Toxins, etc...
- How is the expression of virulence factors regulated ?
  - Temporal regulation
  - *In vivo* triggers
  - Interaction with the host

gene expression profiling

But....you can't express what you don't have

comparative genomic hybridization (CGH)

Part I

Cj virulence and gene expression profiling

Coordinate Regulation of Motility and Virulence

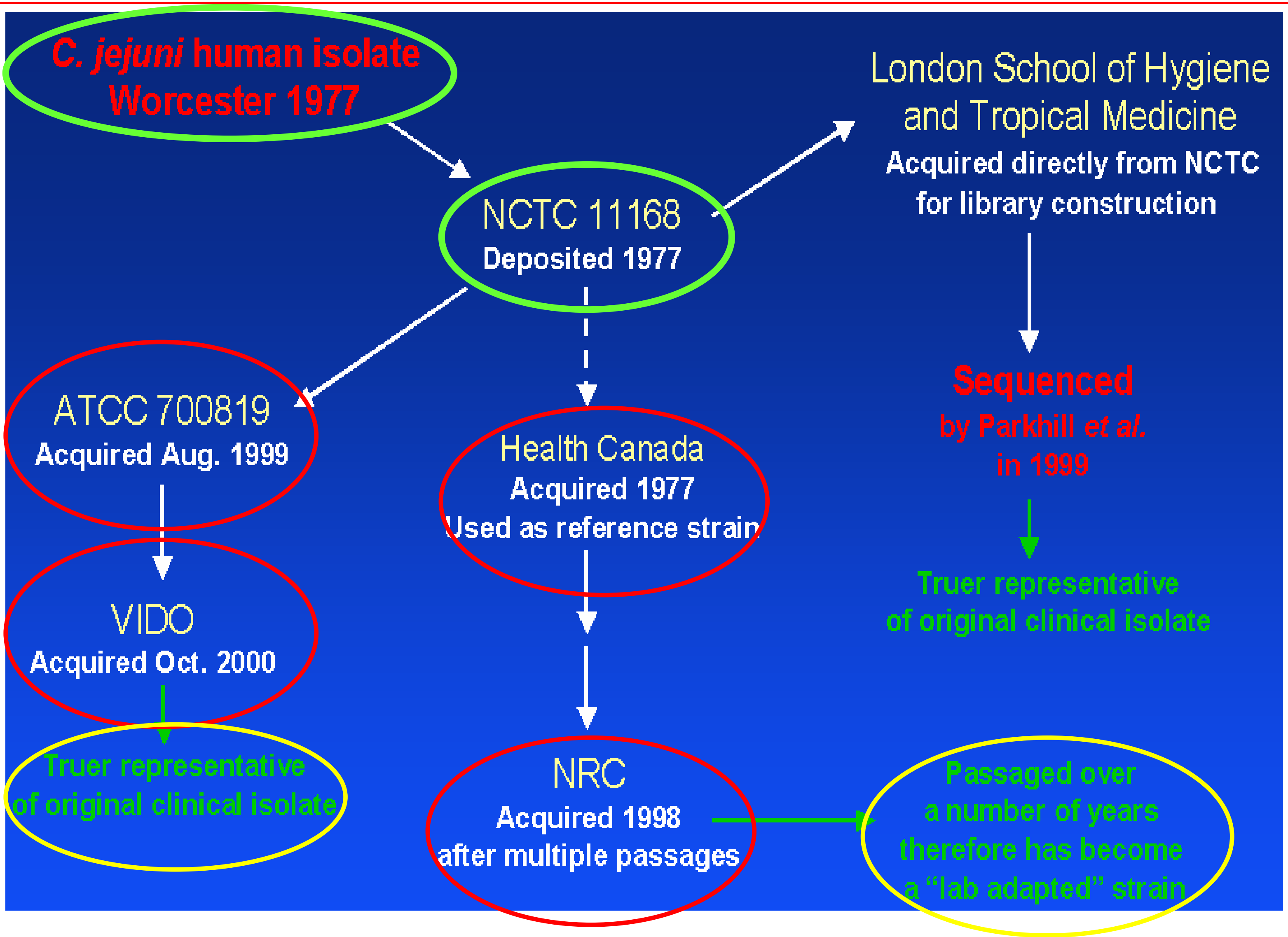
Carrillo *et al*., (2004) Journal of Biological Chemistry (in press)

Aim:

Comparison of gene expression profiles from the genome strain NCTC11168 (“VIDO”) vs. a lab-attenuated variant (“NRC”)

- Comparative Genomic Hybridization
- Gene-expression Profiling
- (Correlation with) Proteomics
- Mutant analysis

Tracing the lineage of two variants…



***C. jejuni* human isolate Worcester 1977**

NCTC 11168
Deposited 1977

London School of Hygiene and Tropical Medicine
Acquired directly from NCTC for library construction

Sequenced by Parkhill *et al.* in 1999

Truer representative of original clinical isolate

ATCC 700819
Acquired Aug. 1999

VIDO
Acquired Oct. 2000

Truer representative of original clinical isolate

Health Canada
Acquired 1977
Used as reference strain

NRC
Acquired 1998
after multiple passages

Passaged over a number of years therefore has become a “lab adapted” strain

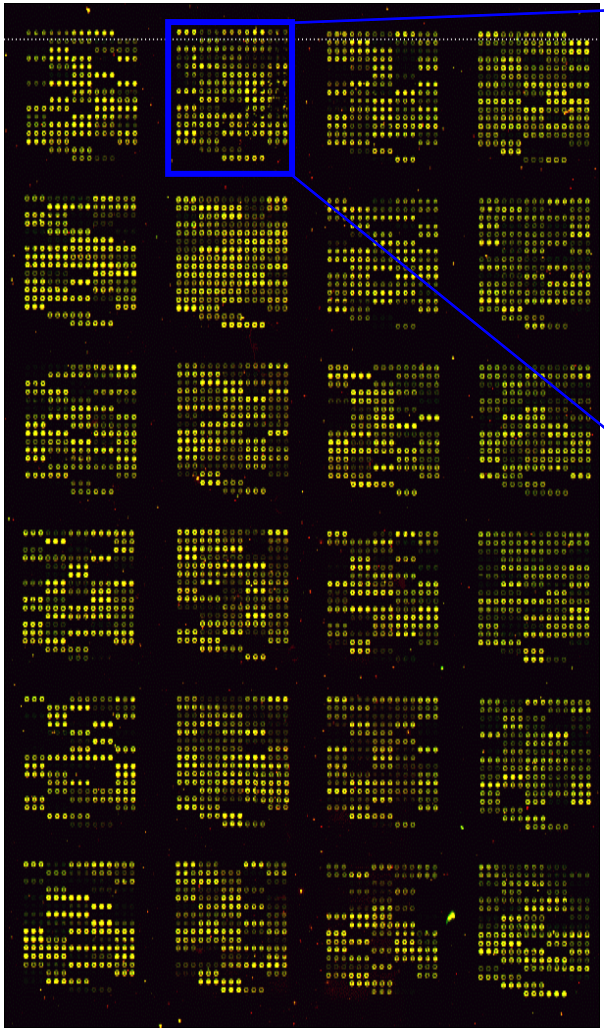
VIDO vs NRC: observed variations in phenotype

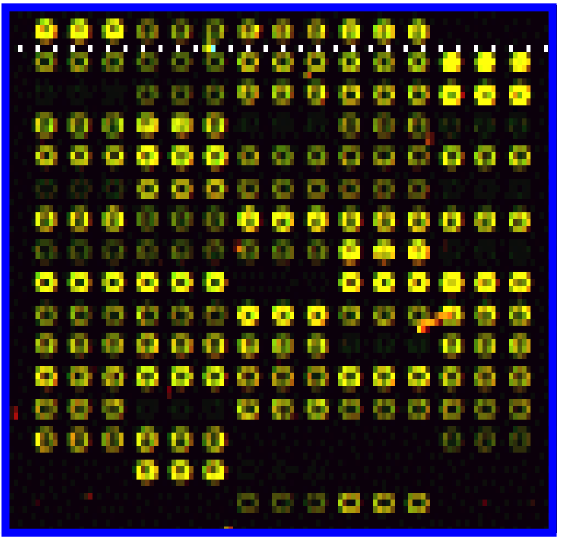
Feature	VIDO	NRC
Flagellation	++	+/-
Motility	++	+/-
Adherence to CaCo-2 cells	++	+
Dose requirement for Infection	10^5 cfu	10^8 cfu
Lateral Spread	Yes	No

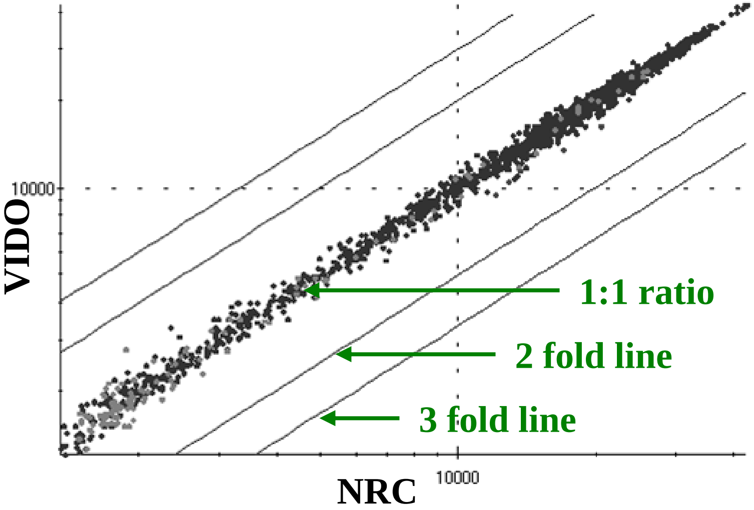
courtesy of Christine Szymanski, Brenda Allen *et al.*

NRC and VIDO variants differ in their “virulence characteristics”

Comparing the genomes of VIDO vs NRC

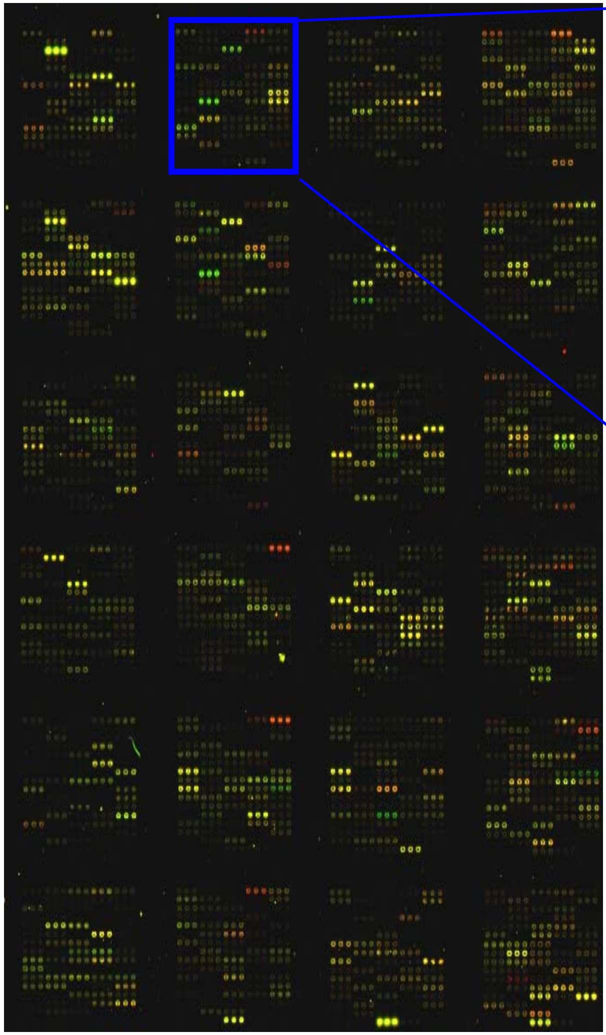


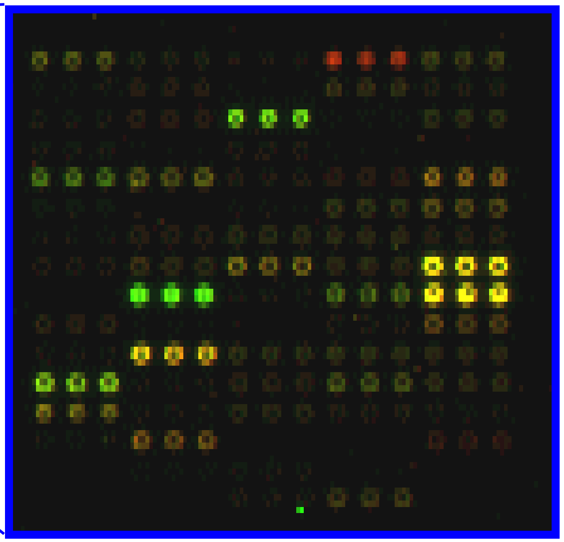


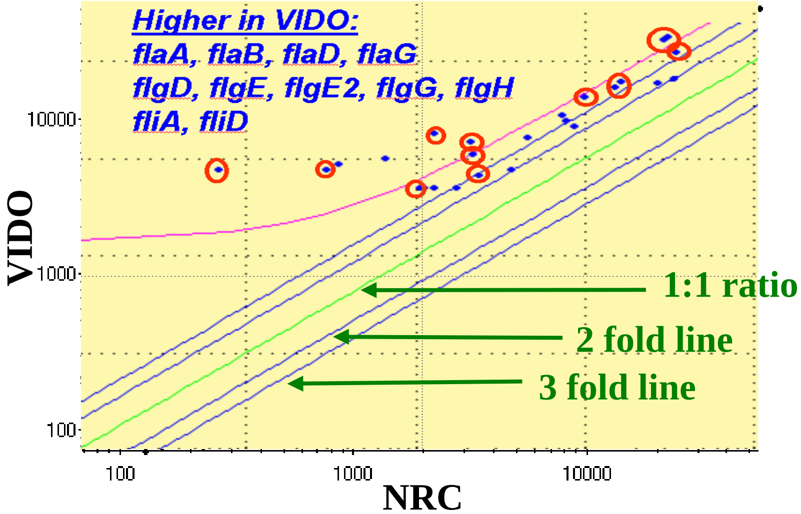


NRC and VIDO variants have same CGH profile (ie.“identical gene content”)

Comparing the transcriptomes of VIDO vs NRC

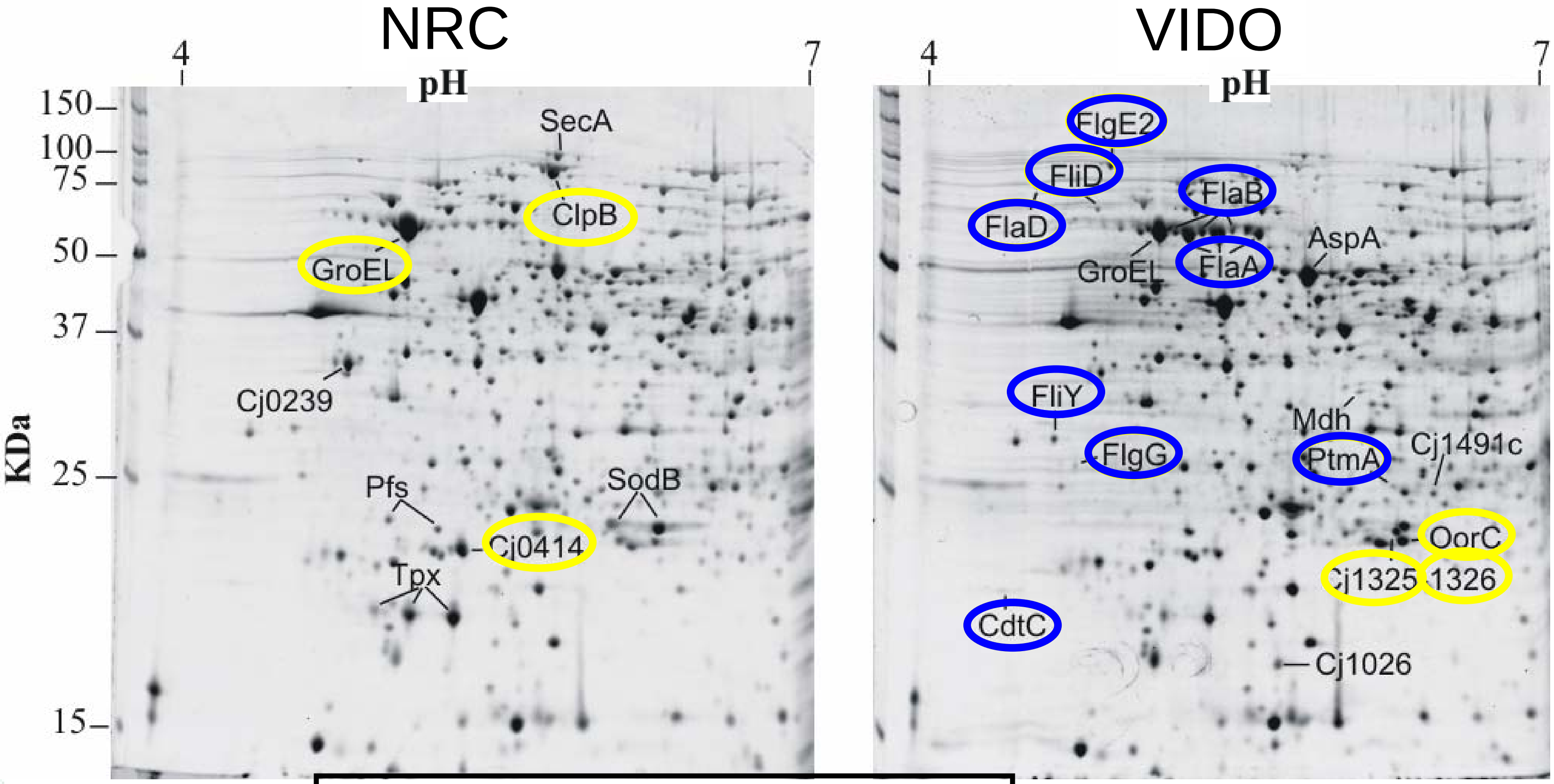






Differences are observed in expression of flagellar proteins in NRC and VIDO variants

Comparing the proteomes of VIDO vs. NRC



○ transcriptome/proteome “confirmation”

○ flagellar/virulence genes

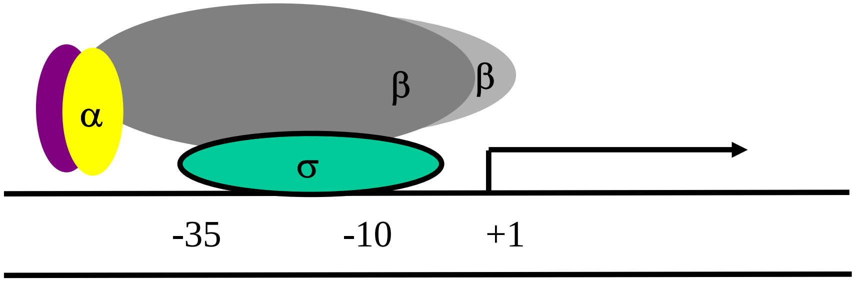
courtesy of Pat Lanthier & John Kelly

The proteome of the VIDO variant shows increased levels of:

- *flagellar proteins*
- *toxin protein cdtC*

Could virulence factors be co-regulated in VIDO ?

Bacterial Promoters:



Only 3 sigma factors identified in *C. jejuni* NCTC11168

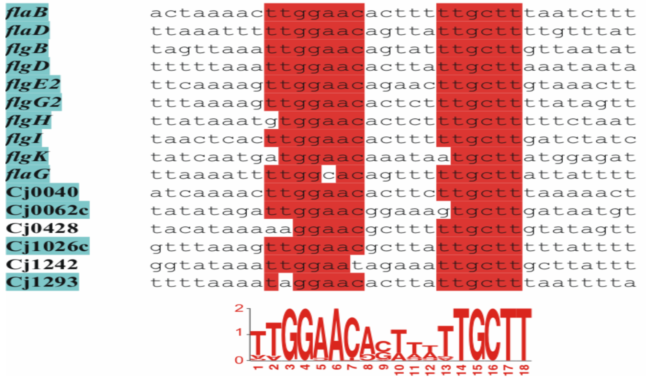
σ^{70}: rpoD

σ^{54}: rpoN

σ^{28}: fliA

} flagellar regulation

σ^{54} Promoters



flaB	actaaaacttggaacactttttgctttaatcttt
flaD	ttaaatttttggaacagttattgcttttgtttat
flgB	tagttaaattggaacagtatttgcttgttaatat
flgD	tttttaaattggaacacttattgcttaaataata
flgE2	ttcaaaagttggaacagaacttgcttgtaaactt
flgG2	tttaaaagttggaacactctttgcttttatagtt
flgH	ttataaatgtggaacactctttgcttttctaat
flgI	taactcacttggaacactttttgcttgatctatc
flgK	tatcaatgatggaacaaataatgcttatggagat
flaG	ttaaaatttggcacagttttttgcttattatttt
Cj0040	atcaaaacttggaacacttcttgctttaaaaact
Cj0062c	tatatagattggaacggaaagtgcttgataatgt
Cj0428	tacataaaaaggaacgctttttgcttgtatagtt
Cj1026c	gtttaaagttggaacgcttattgcttttatttt
Cj1242	ggtataaattggaatagaaattgcttgcttattt
Cj1293	ttttaaaataggaacacttattgctttaatttta

σ^{28} Promoters

flaA	aagtataaaatattttttttgattgcacgatatagcat
flaG	agtggttcataaatttataaatttgtcgatataagct
pseA	aatcataaaataaaatttaattataacgatatatata
Cj0045c	aattattatttttatttttcaaaaaatacgatataaaaa
Cj0391c	ccttataaatttttttaagtcaaaagtcgatataaatt
Cj0859c	tatagtatattttttttaaaataatgtcgatttagtgt
Cj0977	aaatataaattctttttatttttttgccgatattgaat
Cj1450	tatactaaaatattttttaaaaataaccgattttaat
Cj1464	aggggttaagtttttttaaatttaggtcgatatggttt
Cj1656c	aaatattaagttttaaaaatatcatccgatttaaata

VIDO variant shows increased levels of ***both*** *σ^{54}- and σ^{28}-regulated genes*

Mutagenesis studies on putative flagellar regulators

Several genes are known to have an effect on flagellar gene regulation

σ^{54} (rpoN)

σ^{28} (fliA)

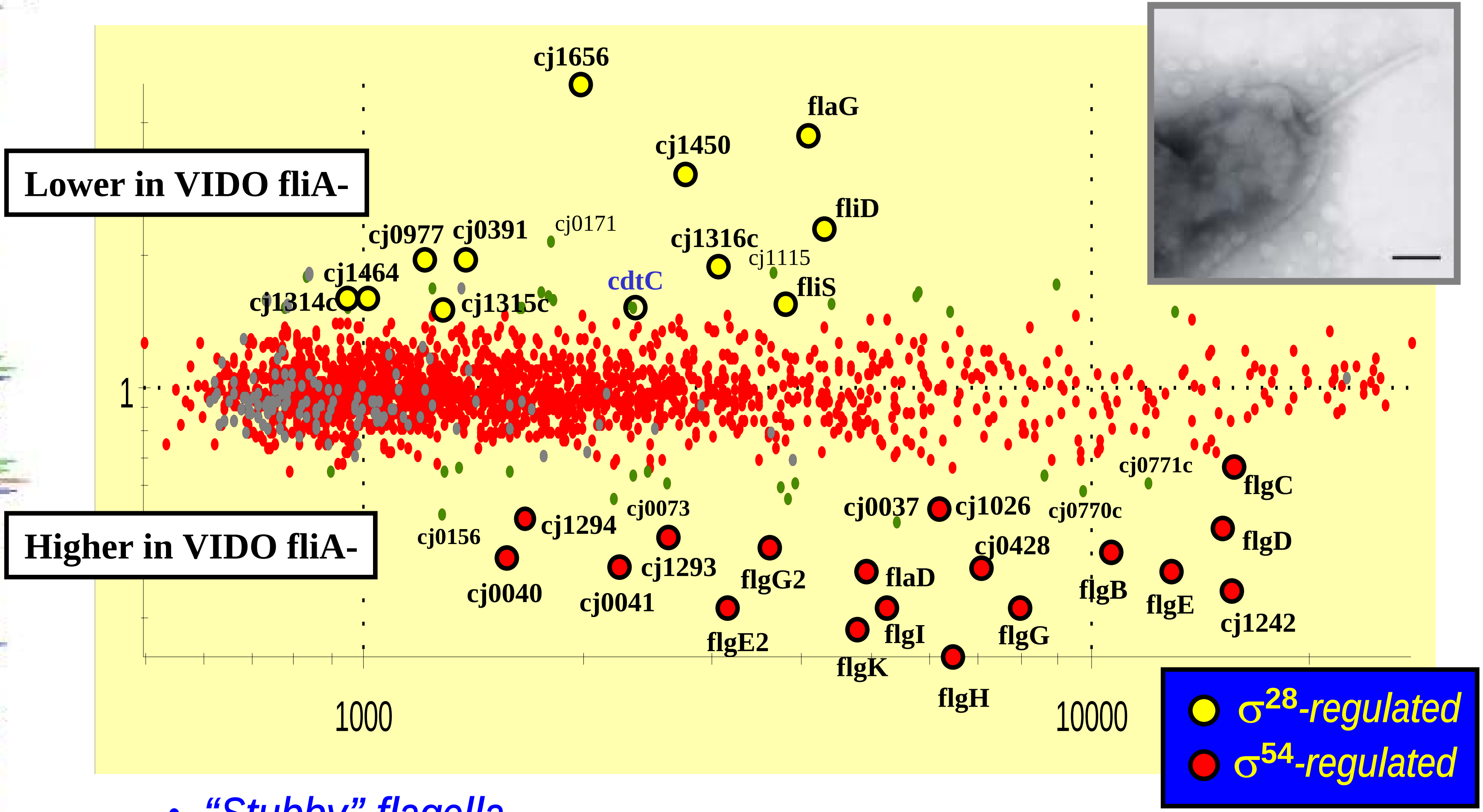
flgR (cj1024c)- σ^{54} activator

flhB

flhA

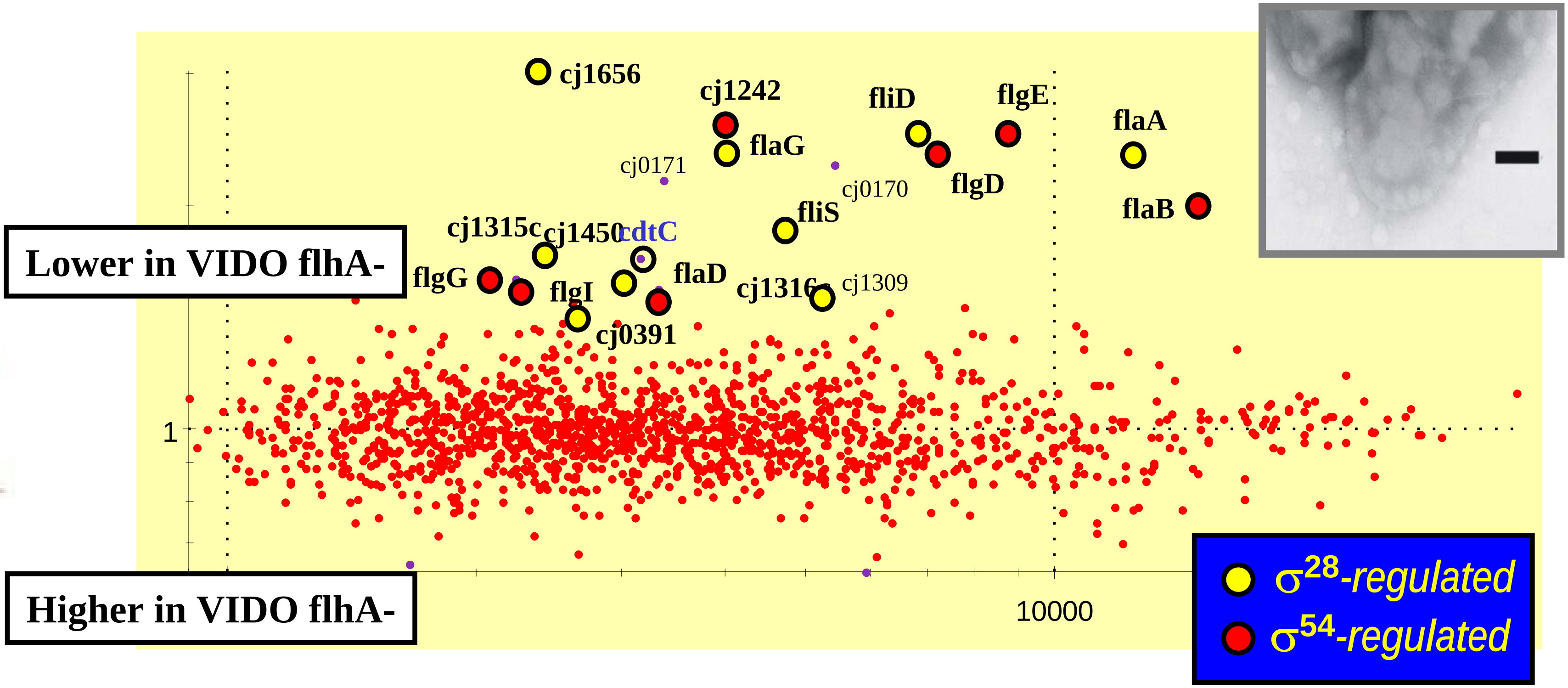
mutagenesis / expression profiling of mutants

Study of a σ^{28} (fliA) knock-out mutant in VIDO

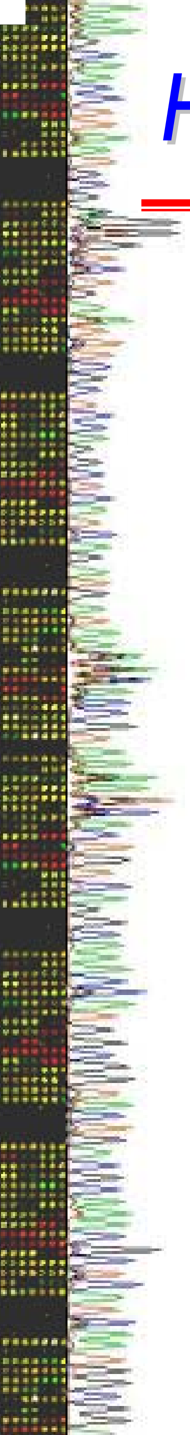


- *"Stubby" flagella*
- *Lower levels of σ^{28}-regulated genes*
- *Increased levels of σ^{54}-regulated genes*
  - *different from "NRC effect"*

Study of a flhA knock-out mutant in VIDO



- *No flagella*
- *lower transcript levels of <u>both</u> σ^{28}- and σ^{54}-regulated genes*
  - *similar to "NRC effect"*
  - FlhA *regulates the expression of the genes that are differentially expressed between NRC and VIDO variants*



How is FlhA involved in flagellar regulation ?

- FlhA is not a transcription factor
- FlhA is a structural protein of the nascent flagellum
    - involved in export of late flagellar proteins
- Impairment of flagellar protein export in *Pseudomonas aeruginosa*
    - downregulation of flagellar genes
    - accumulation of flagellar precursons / feedback inhibition ?

Part II

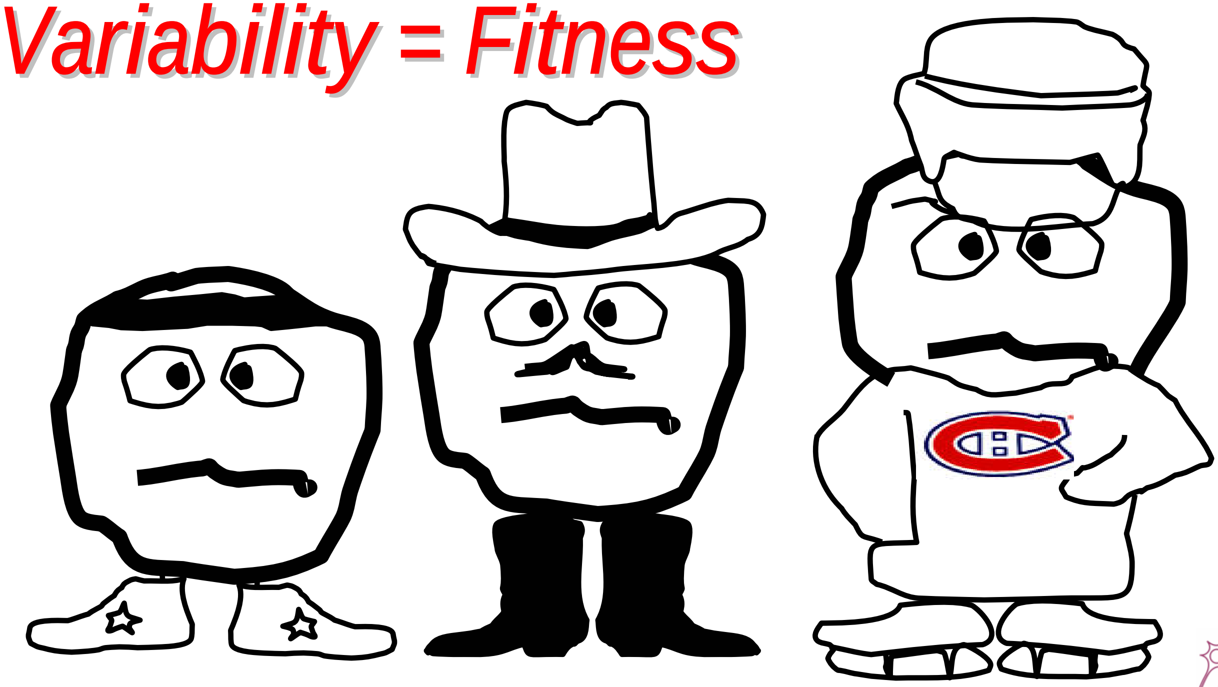
Cj virulence
and
Comparative Genomic Hybridization

Adaptability = Survival = Fitness

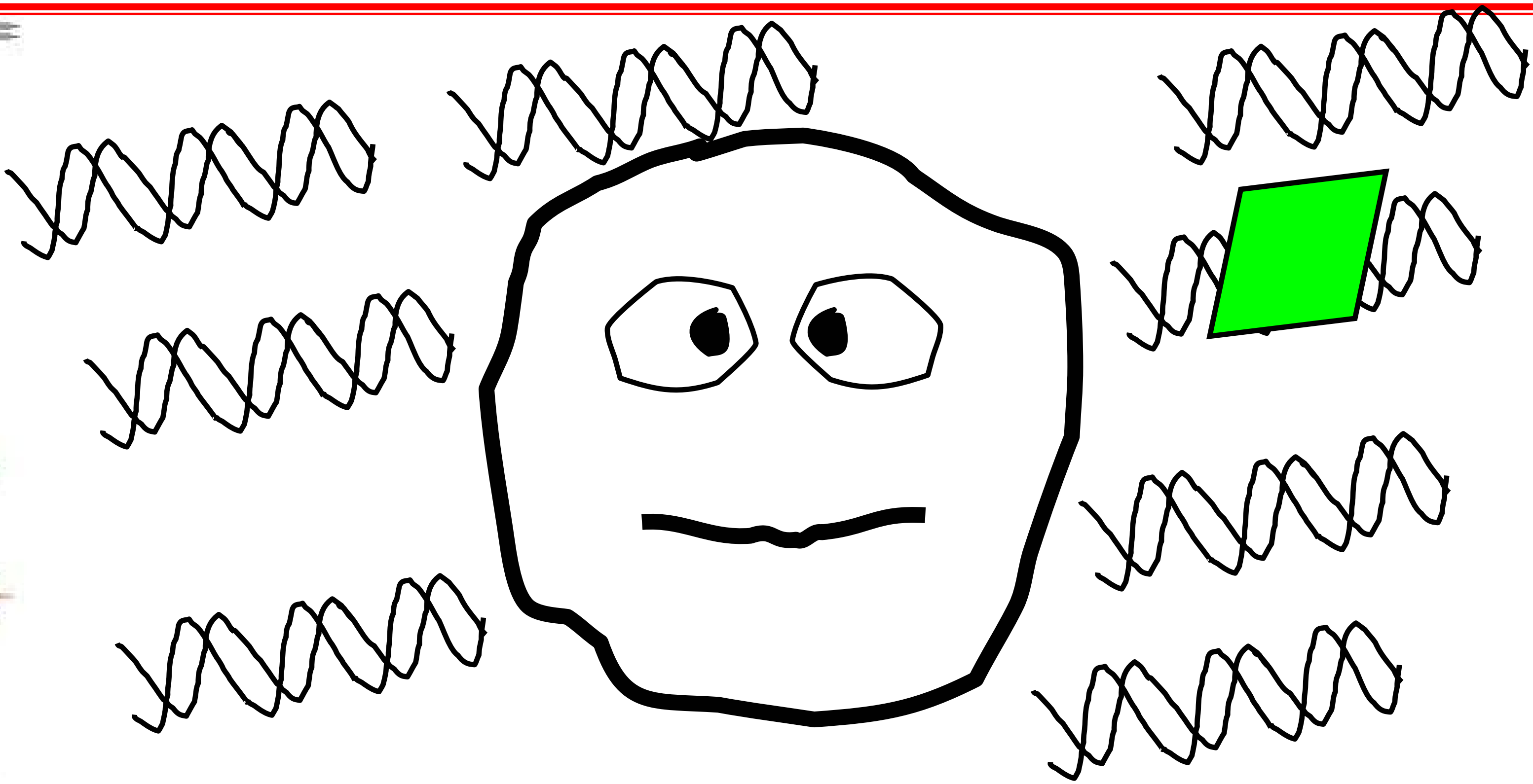
at the level of the individual:
-short term adaptability = differential gene expression

at the level of the species:
-long term adaptability = genetic variability

Variability = Fitness



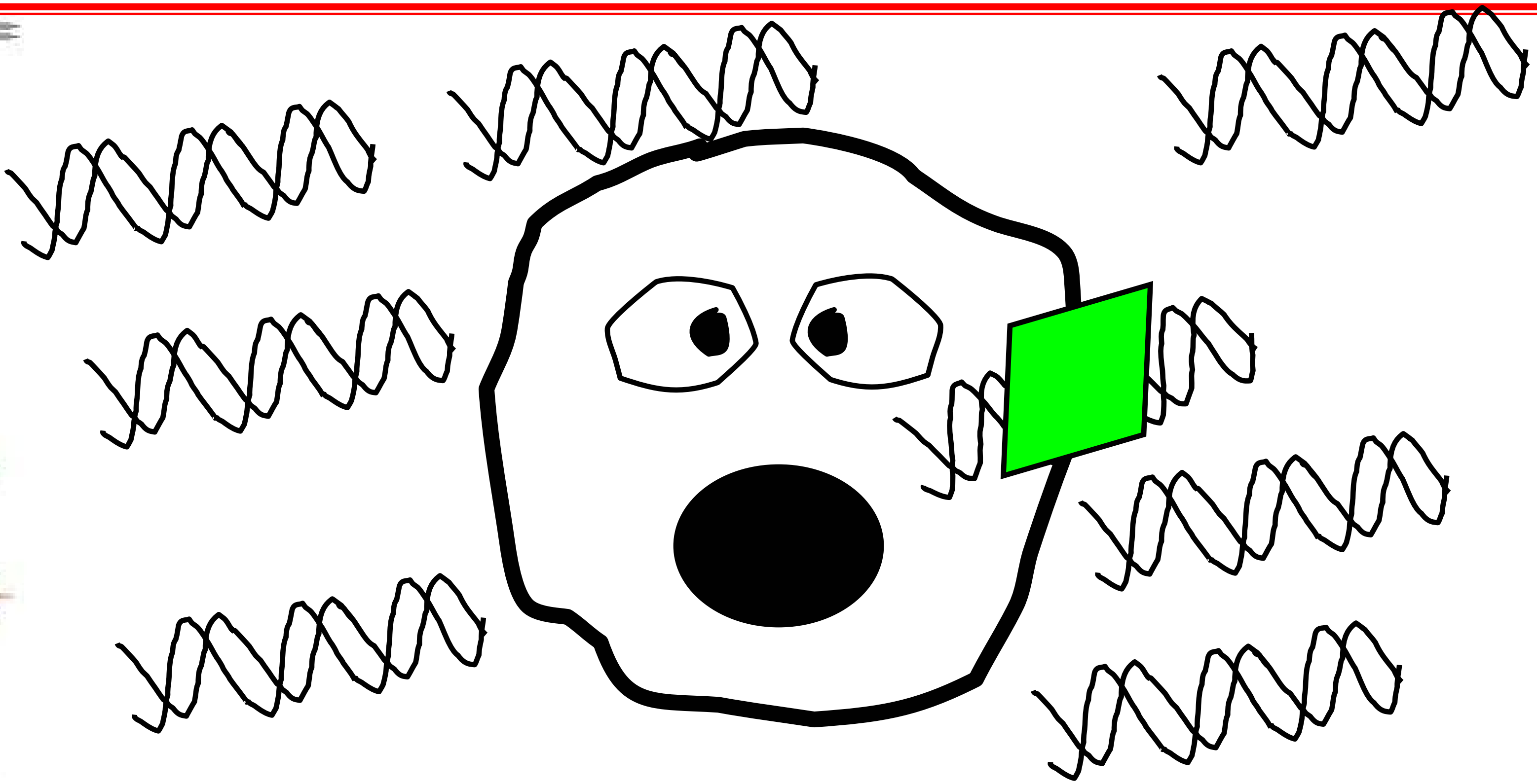
Bacteria love to take up (foreign) DNA...



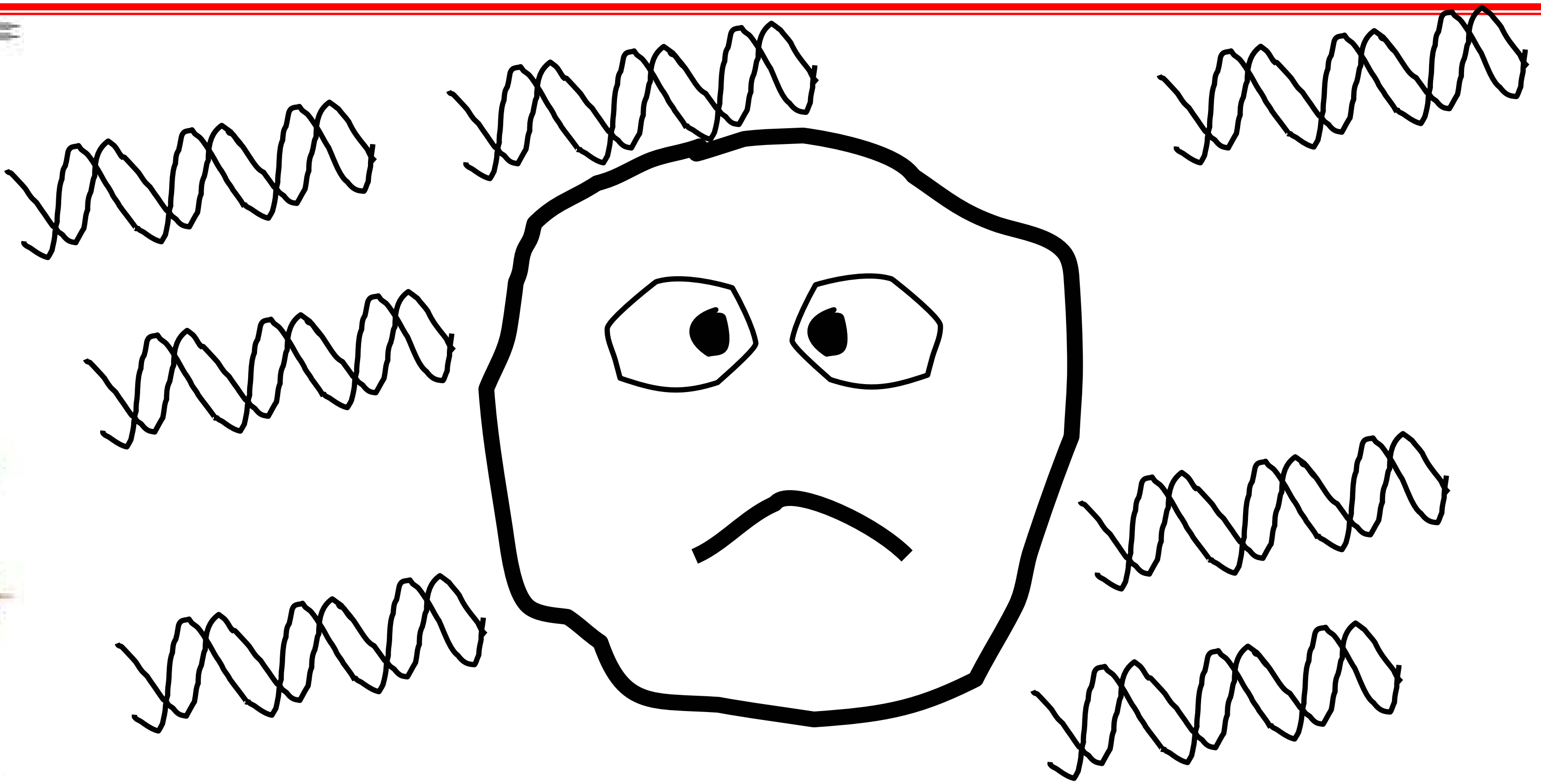
Bacteria live in “dilute DNA solutions”

- dead brothers and sisters
- other dead things
- mobile genetic elements

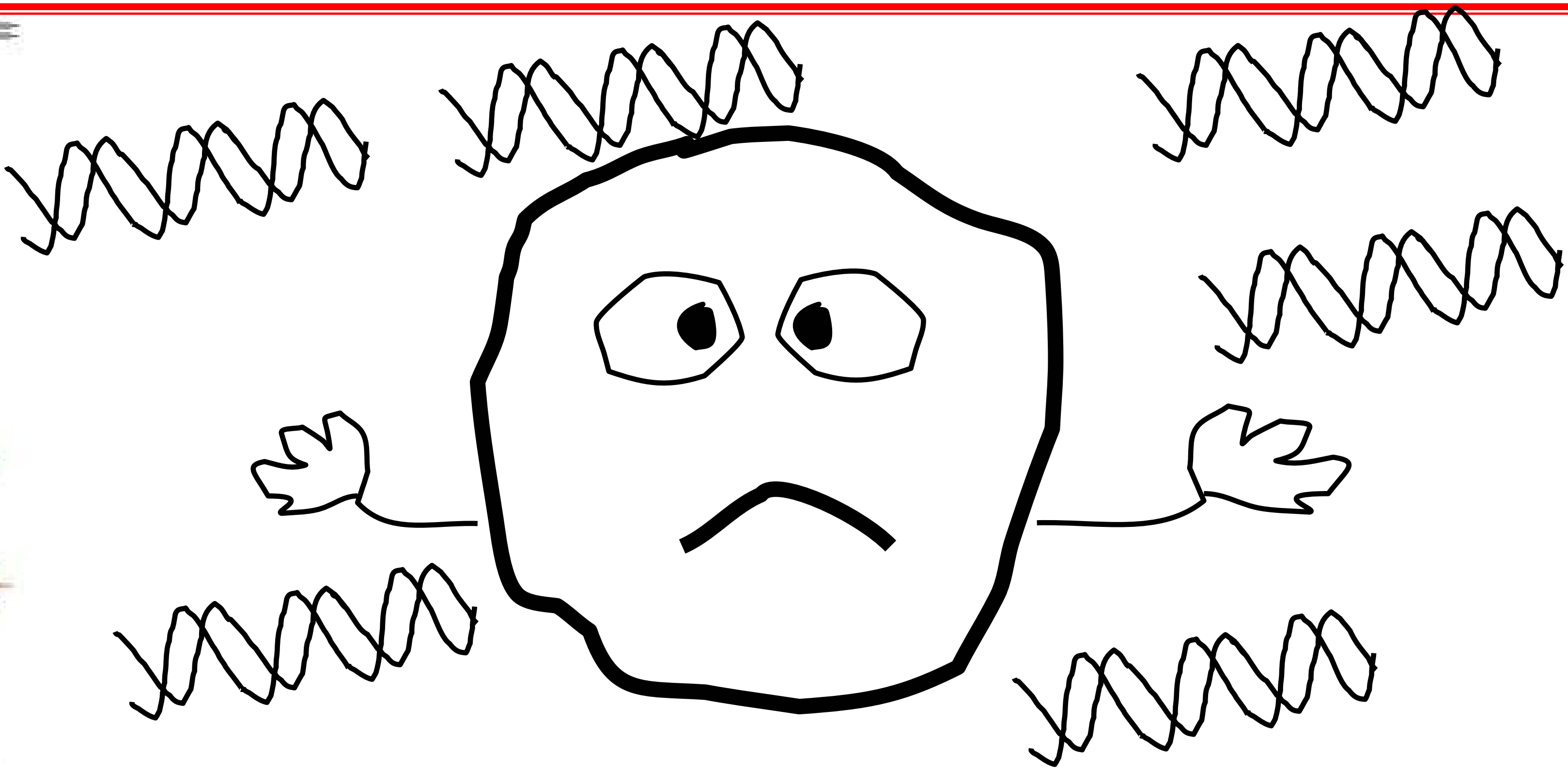
Bacteria love to take up (foreign) DNA...



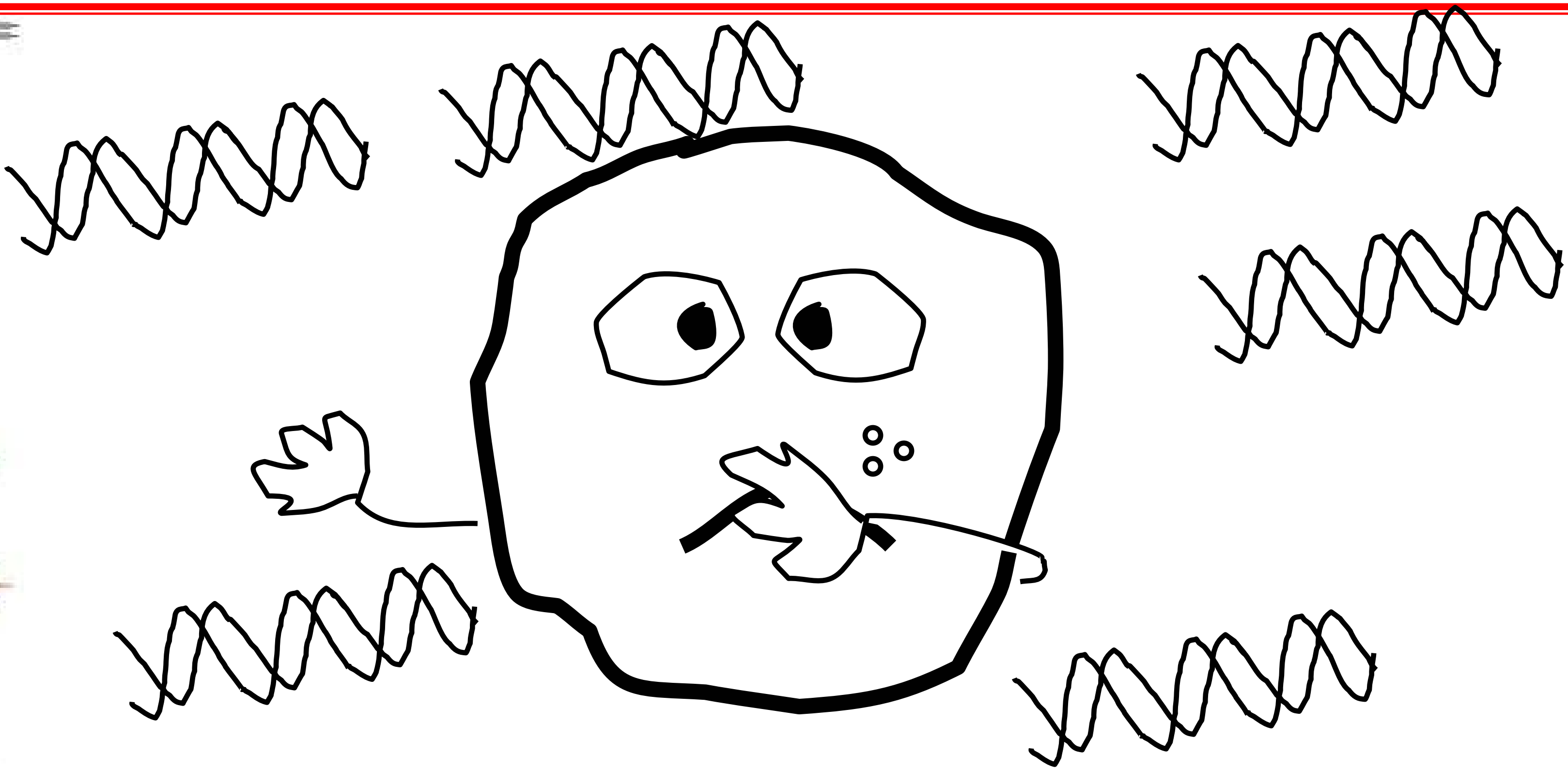
Bacteria love to take up (foreign) DNA...



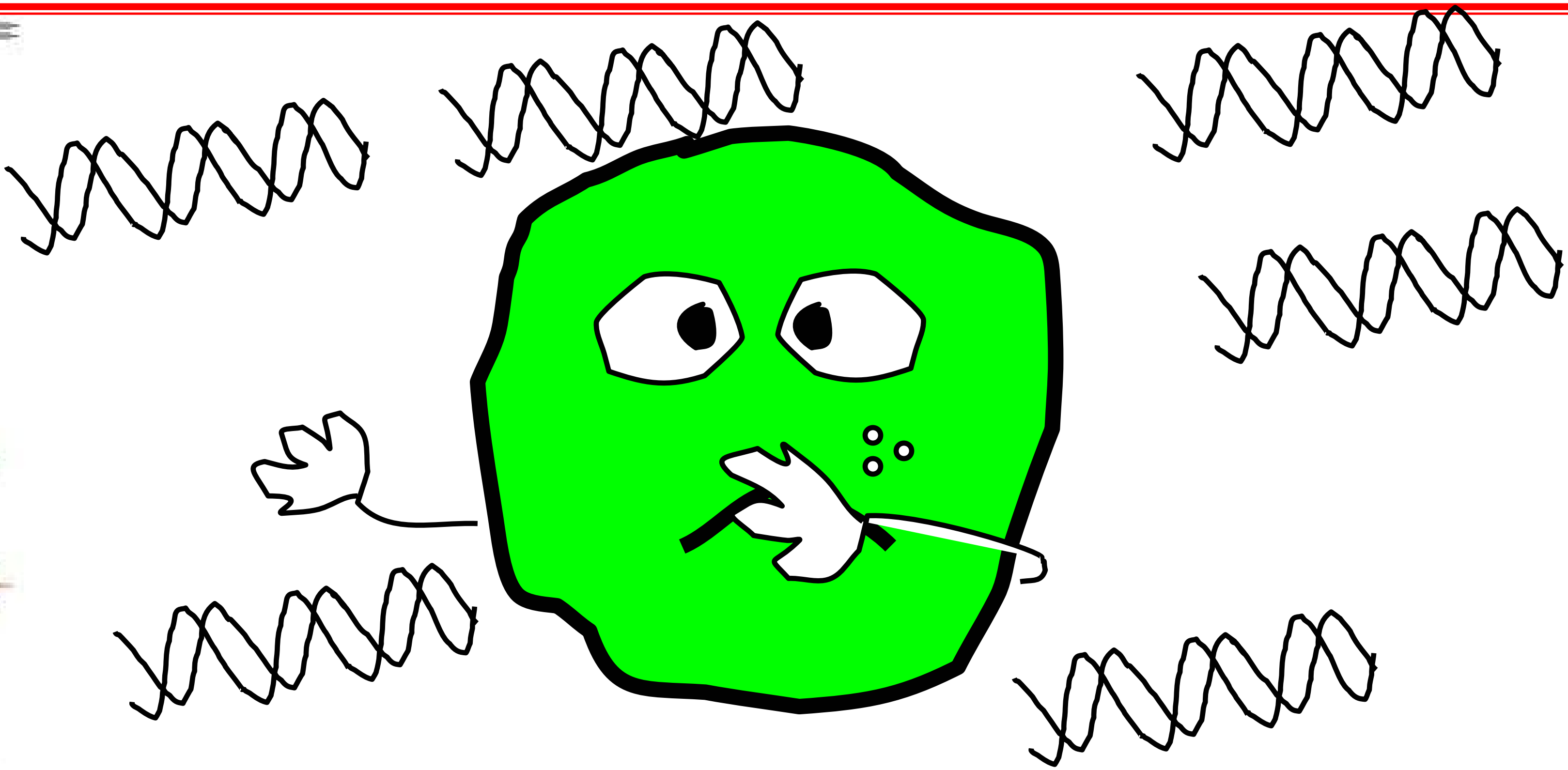
Bacteria love to take up (foreign) DNA...



Bacteria love to take up (foreign) DNA...



Bacteria love to take up (foreign) DNA...



- Lateral (Horizontal) Gene Transfer
- Very common in bacteria
- Detectable in short evolutionary time frames
- Bacterial strains of the same species can be <u>VERY</u> different

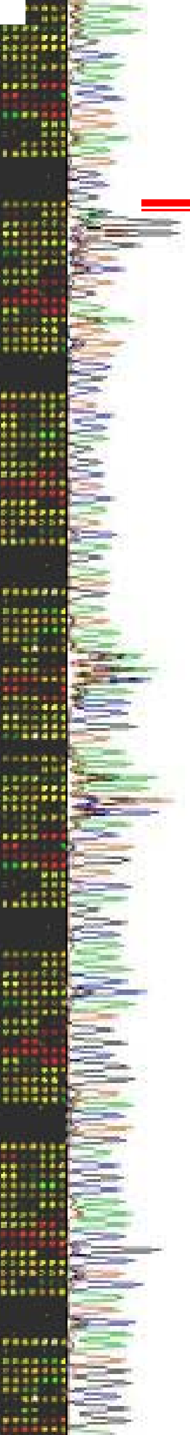
Comparative Genomics to the Rescue

The problem with comparative genomics…

- Ideally, performed with genome sequence data
    - need multiple strains sequenced

- Otherwise, use indirect methods:
    - directed sequencing of specific loci
    - indirect molecular methods
        - SSH, PFGE, AFLP, etc…

- Microarrays
    - Strengths: tool for whole-genome composition analysis
    - Limitations: can only test for the genes present in array



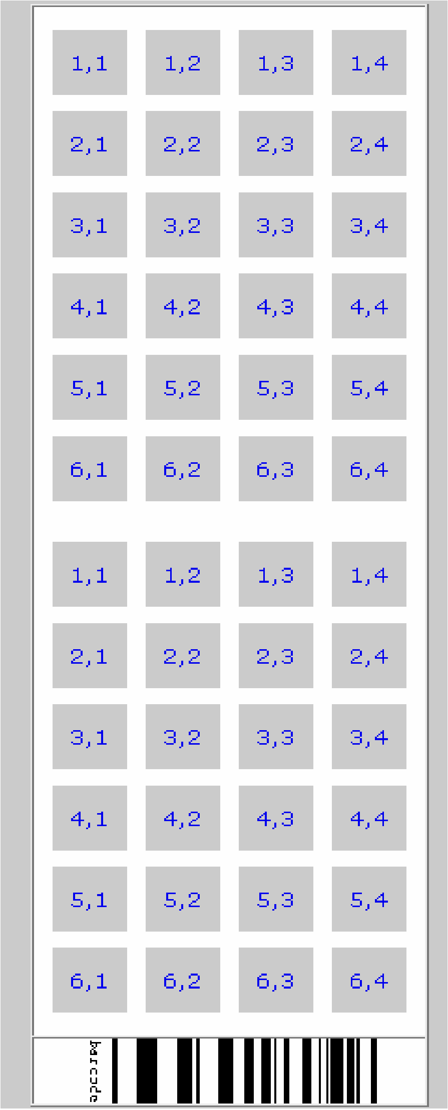
Comparative genomics of *C. jejuni*

The C. jejuni Genome Variability Survey

Aims:

- Quantify the extent of genetic variability across different strains using full-genome Cj microarray
- Define “hot spots” of genetic variability in the Cj genome
- Develop Array-CGH as a molecular typing method
- Correlation between gene content and phenotype

Designing a better array: Campychip v3.0



1,1	1,2	1,3	1,4
2,1	2,2	2,3	2,4
3,1	3,2	3,3	3,4
4,1	4,2	4,3	4,4
5,1	5,2	5,3	5,4
6,1	6,2	6,3	6,4

1,1	1,2	1,3	1,4
2,1	2,2	2,3	2,4
3,1	3,2	3,3	3,4
4,1	4,2	4,3	4,4
5,1	5,2	5,3	5,4
6,1	6,2	6,3	6,4

Better Coverage !

1867 Total Genes:

- *C. jejuni* NCTC 11168 backbone:
  - Genes with validated signal: 1570
  - Effective Coverage: 96%
- 297 genes derived from other strains
  - RM1221 ORFs: 188
  - 81-176 pVIR ORFs: 67
  - ORFs from other strains: 42

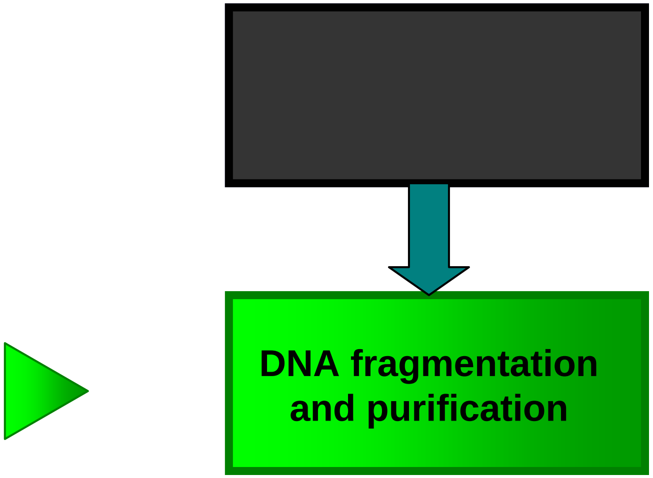
Comparative genomics using microarrays

Cell Harvest & DNA isolation

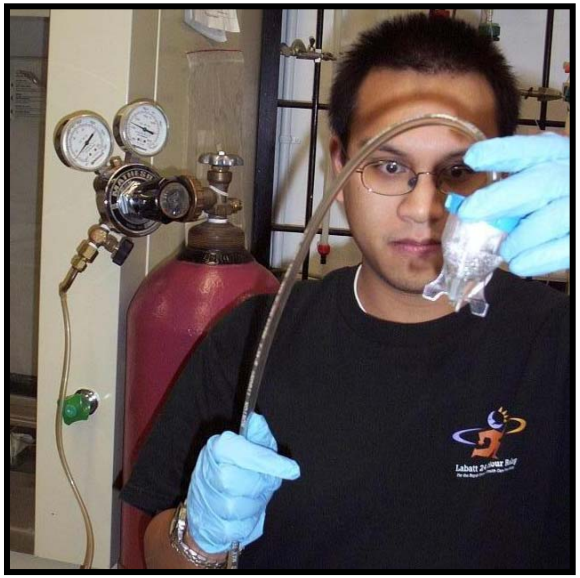
note: not actual Campy

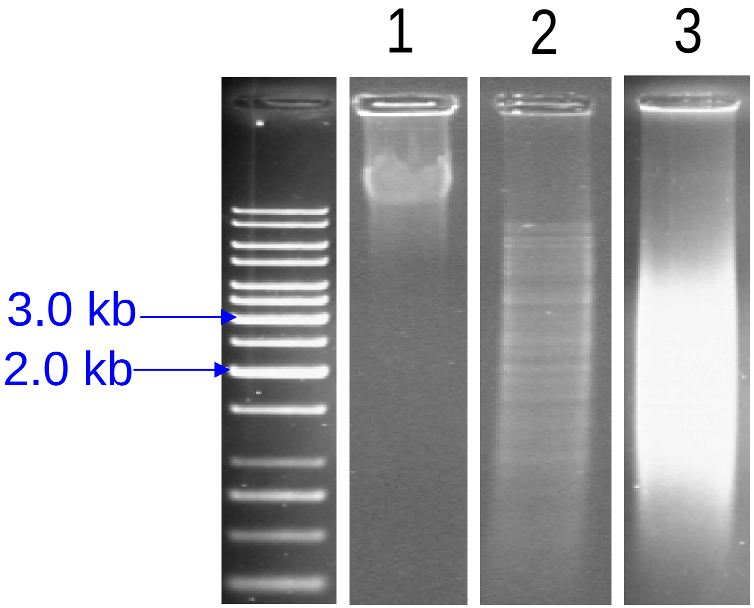
- Growth of *C. jejuni* on Mueller-Hinton plates
- Harvesting of cell pellets
- Lysis of cells and phenol extraction of DNA

Comparative genomics using microarrays



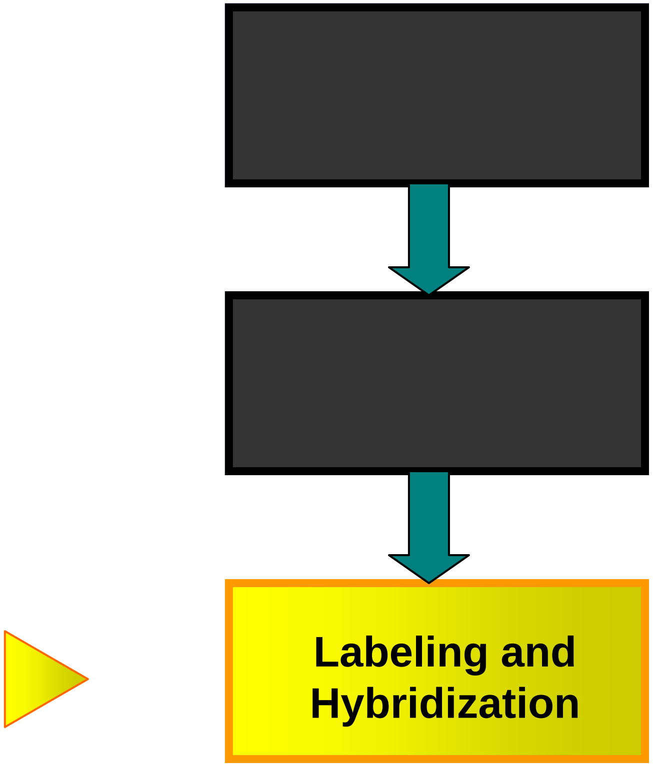
DNA fragmentation and purification



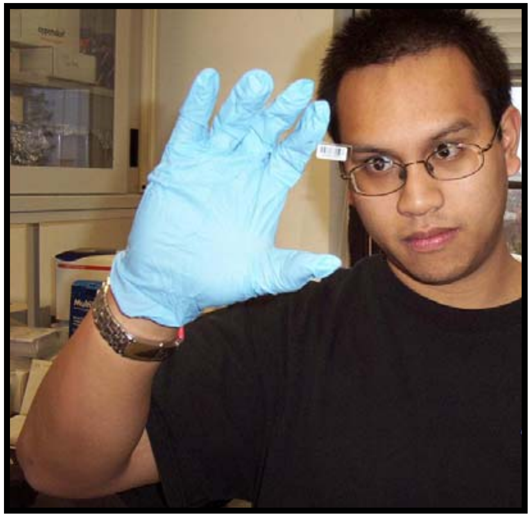


- Nebulization of DNA into linear fragments (~750 bp – 5.0 kb).
- Purification by ethanol precipitation

Comparative genomics using microarrays



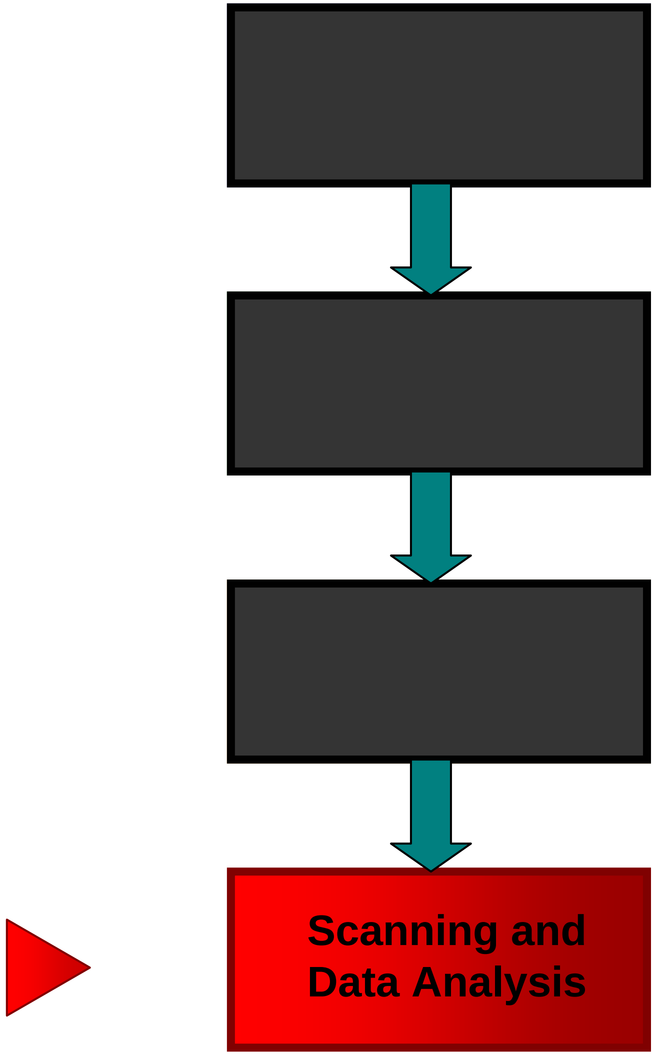
Labeling and Hybridization

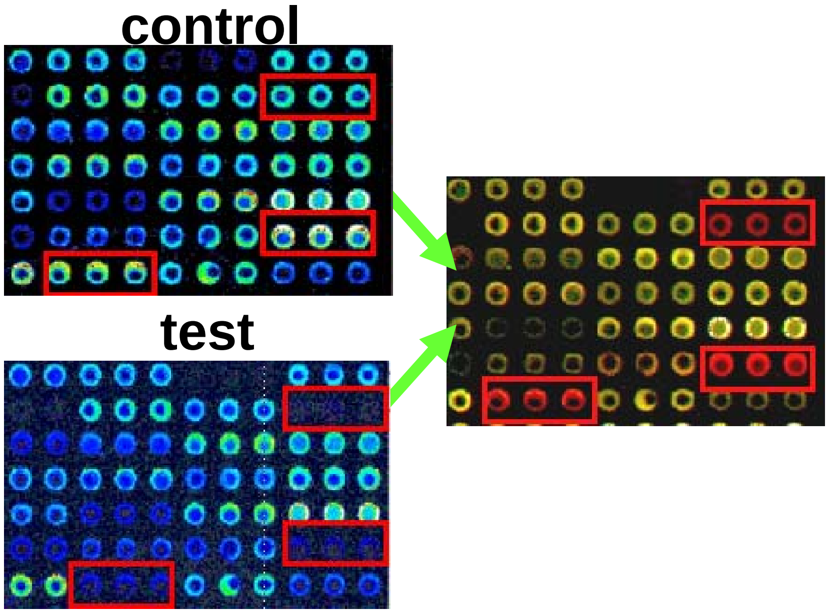


tester	control
Cy3	Cy5

- Direct chemical labeling of the tester and control DNAs with Cy3 and Cy5 dyes, followed by column purification
- Hybridization to microarray

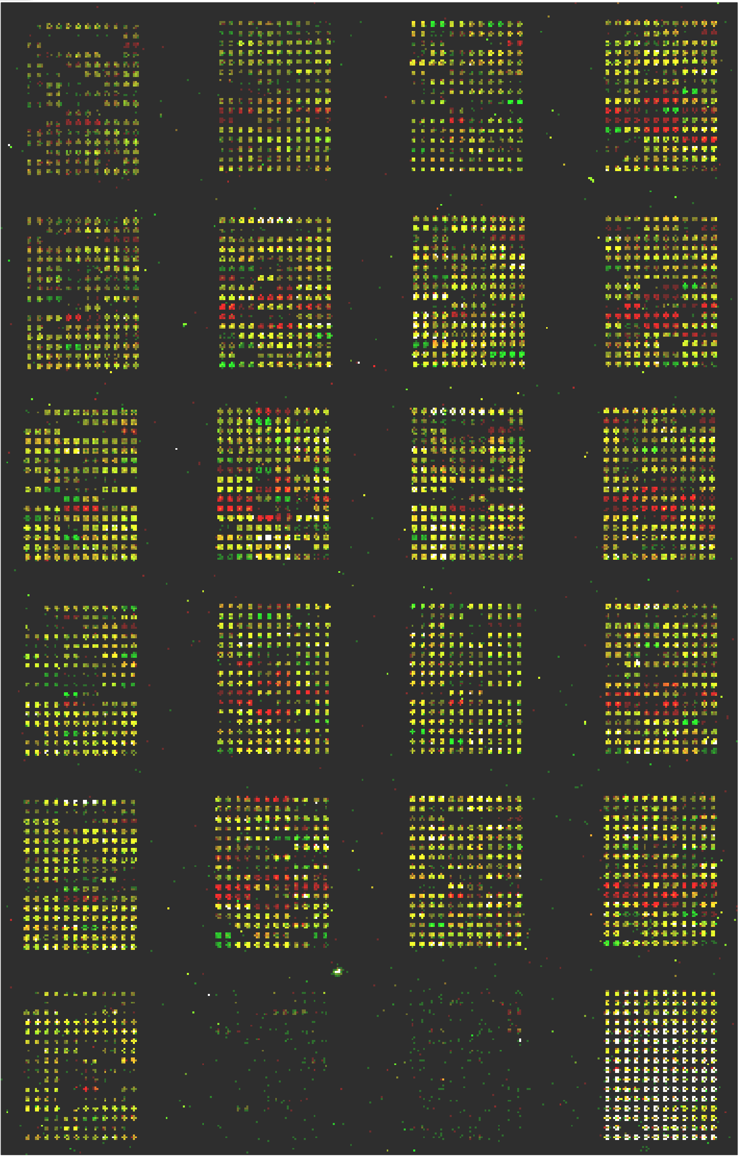
Comparative genomics using microarrays

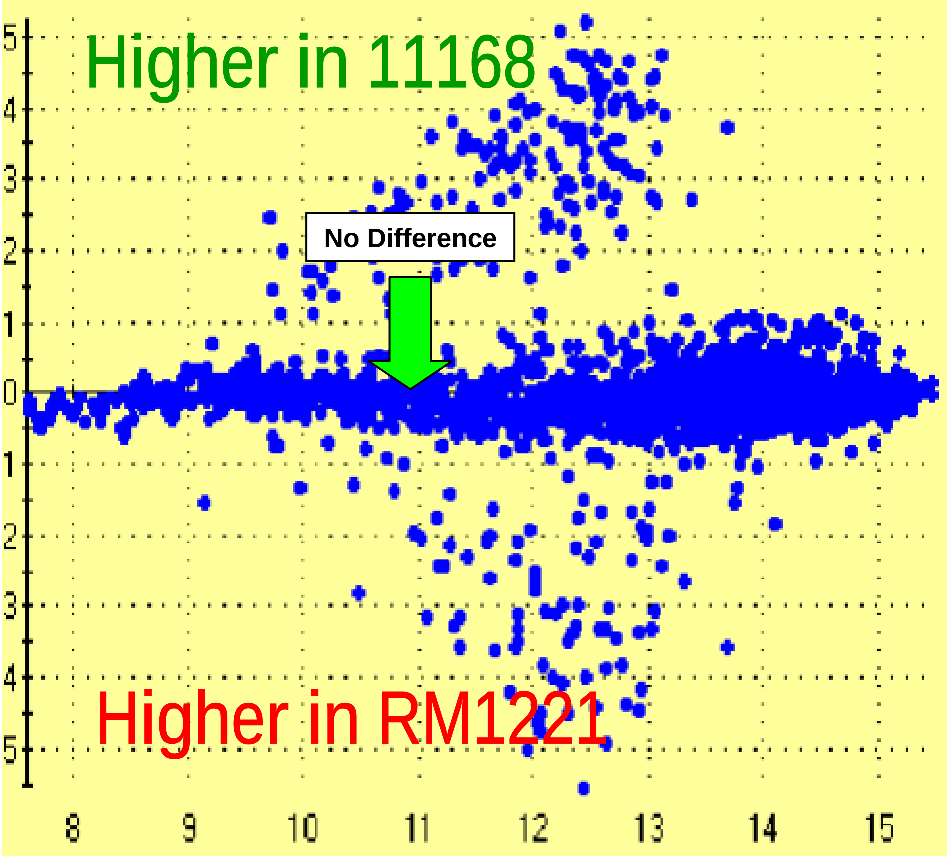

Scanning and Data Analysis



- Microarray scanning
- Analysis using ArrayPro software
  - Filtering of background and noise
  - Categorization of spots
- Data export into BASE followed by further analysis in a biological context

Microarray comparison of 11168 (Cy3) vs. RM1221 (Cy5)

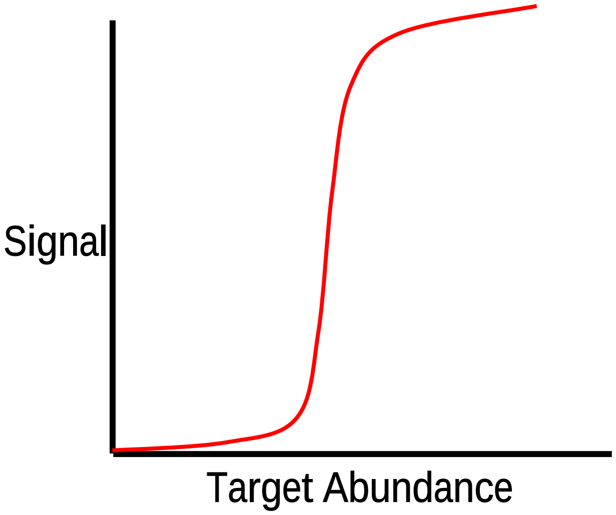




Comparative Genomic Hybridization vs. Transcript Profiling

Hybridization signal affected by :

- target abundance
- target / probe similarity

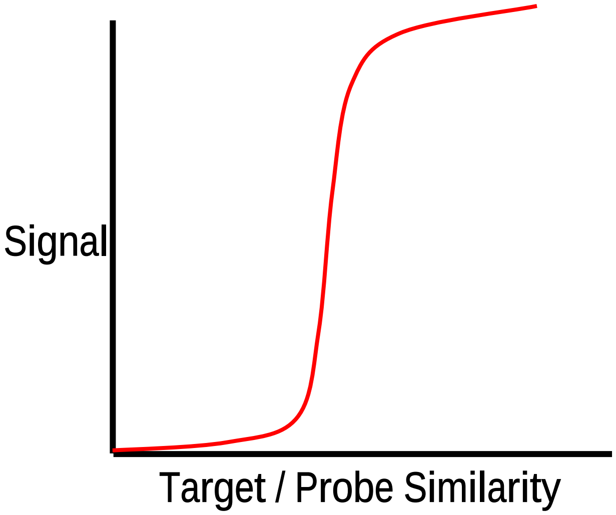


Transcript Profiling

- transcript abundance

CGH

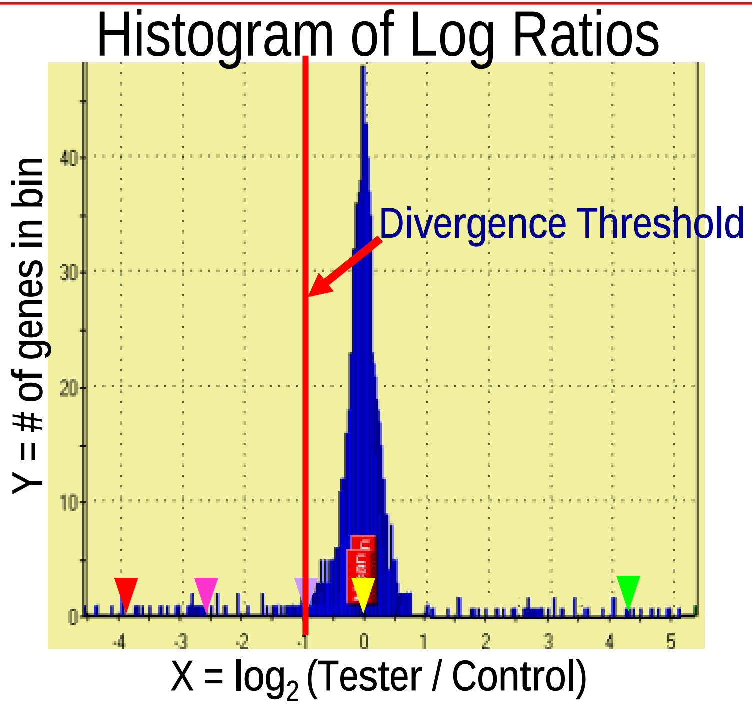
- gene copy number

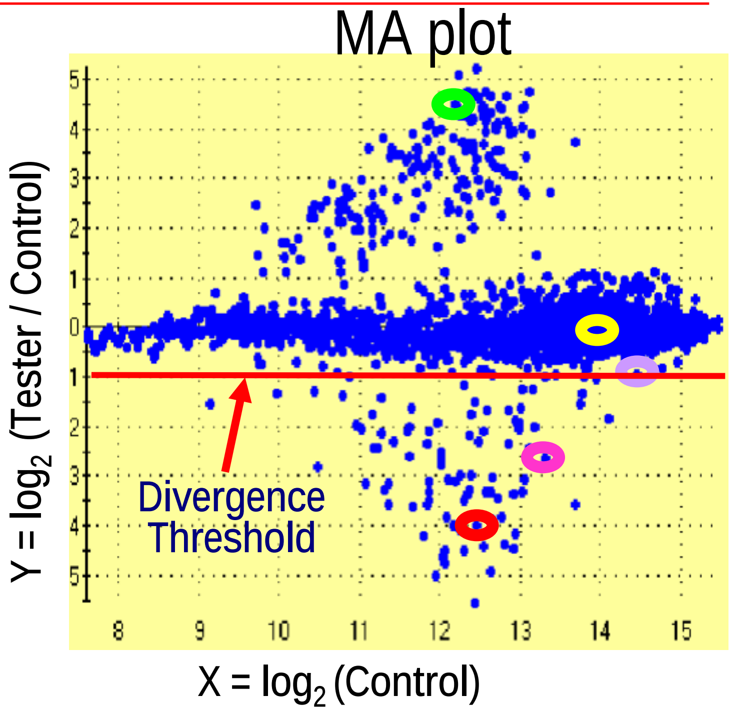


CGH

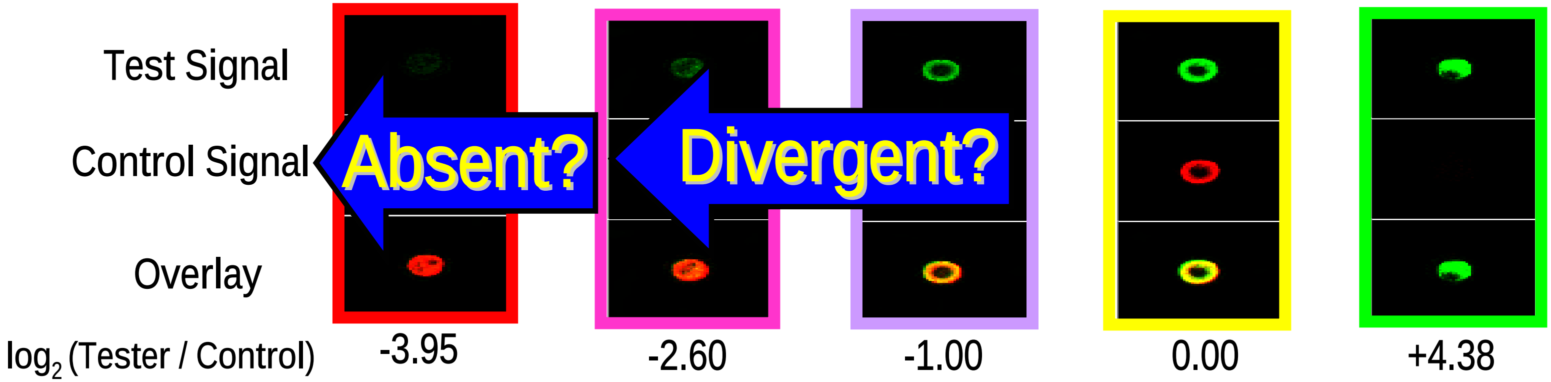
- gene divergence

Absent Genes vs. Divergent Genes ????


Histogram of Log Ratios

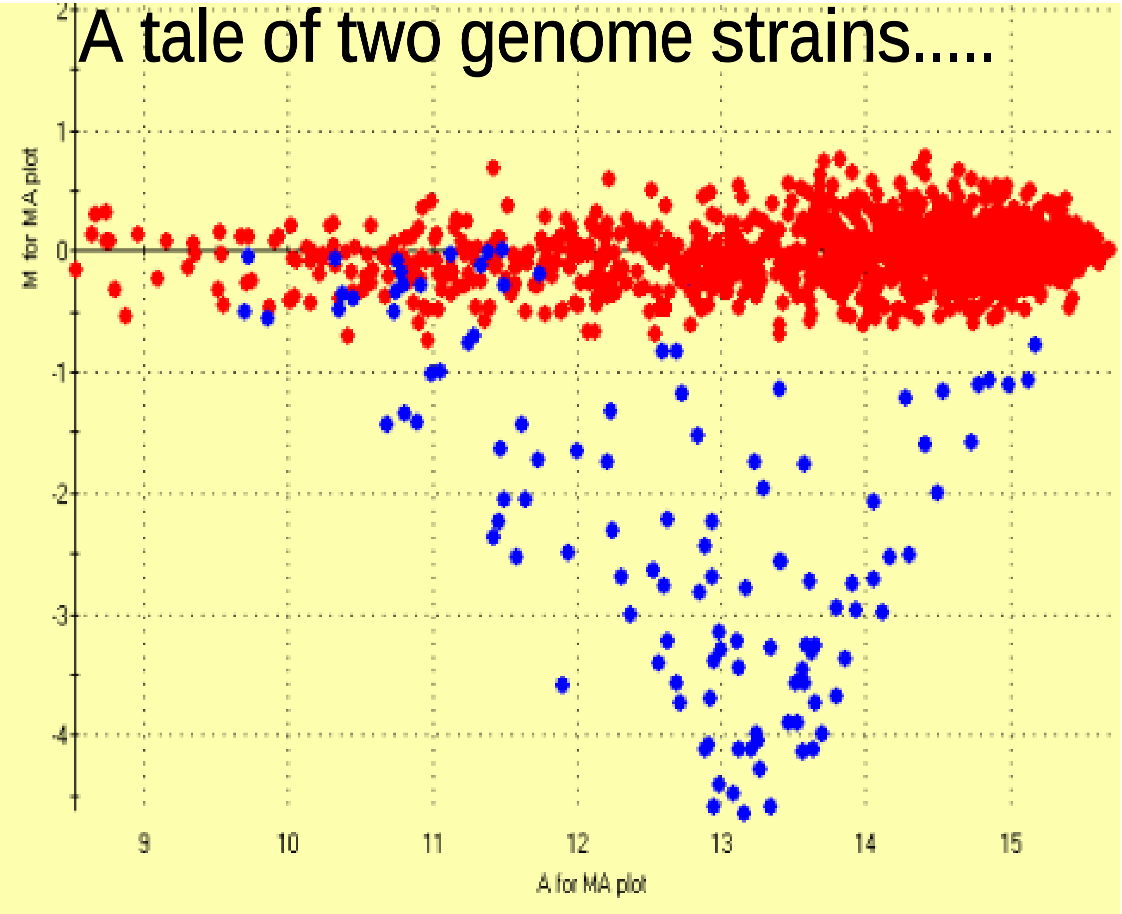

MA plot

- Selection of <u>arbitrary</u> cut-off as Threshold for "Divergent Genes"



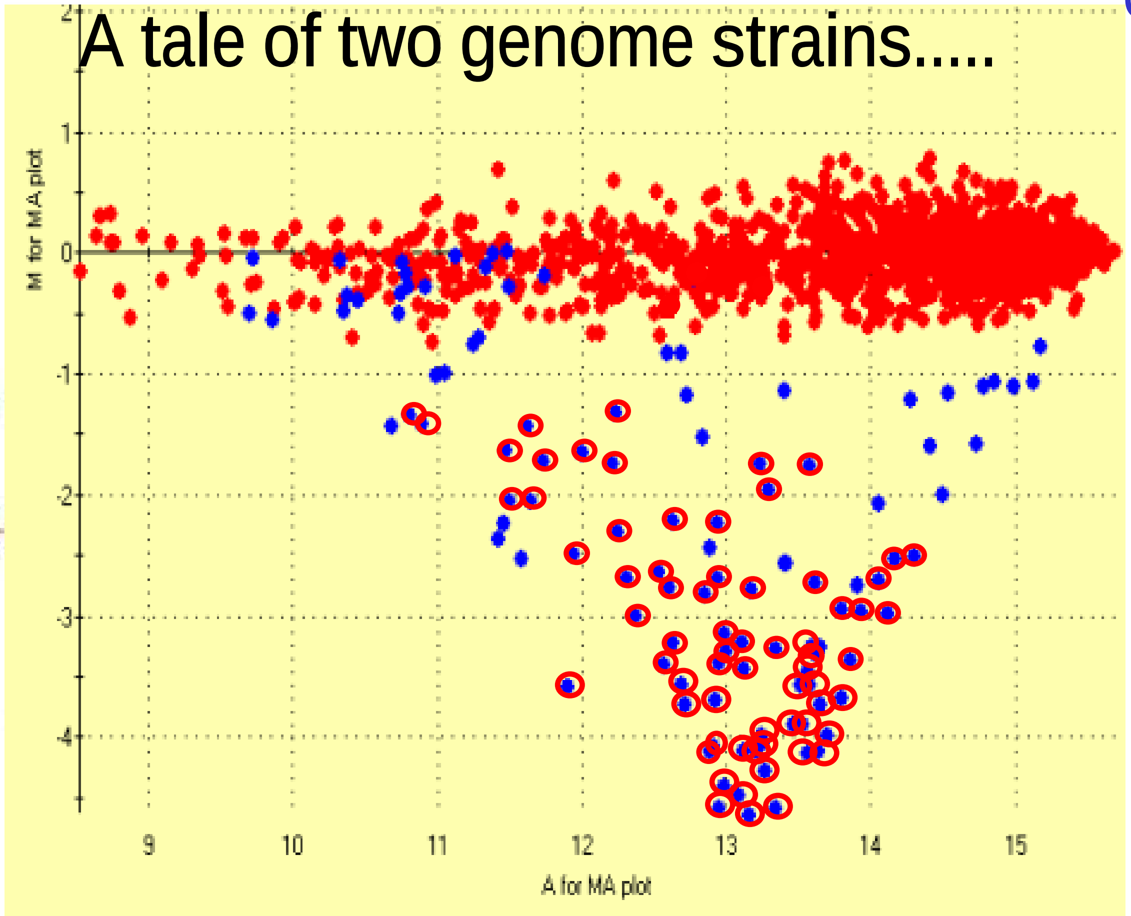
What is the cut-off for Absent vs. Divergent calls ?

C. jejuni NCTC11168
vs.
C. jejuni RM1221


A tale of two genome strains.....

● Predicted RM1221 “outliers”

What is the cut-off for Absent vs. Divergent calls ?



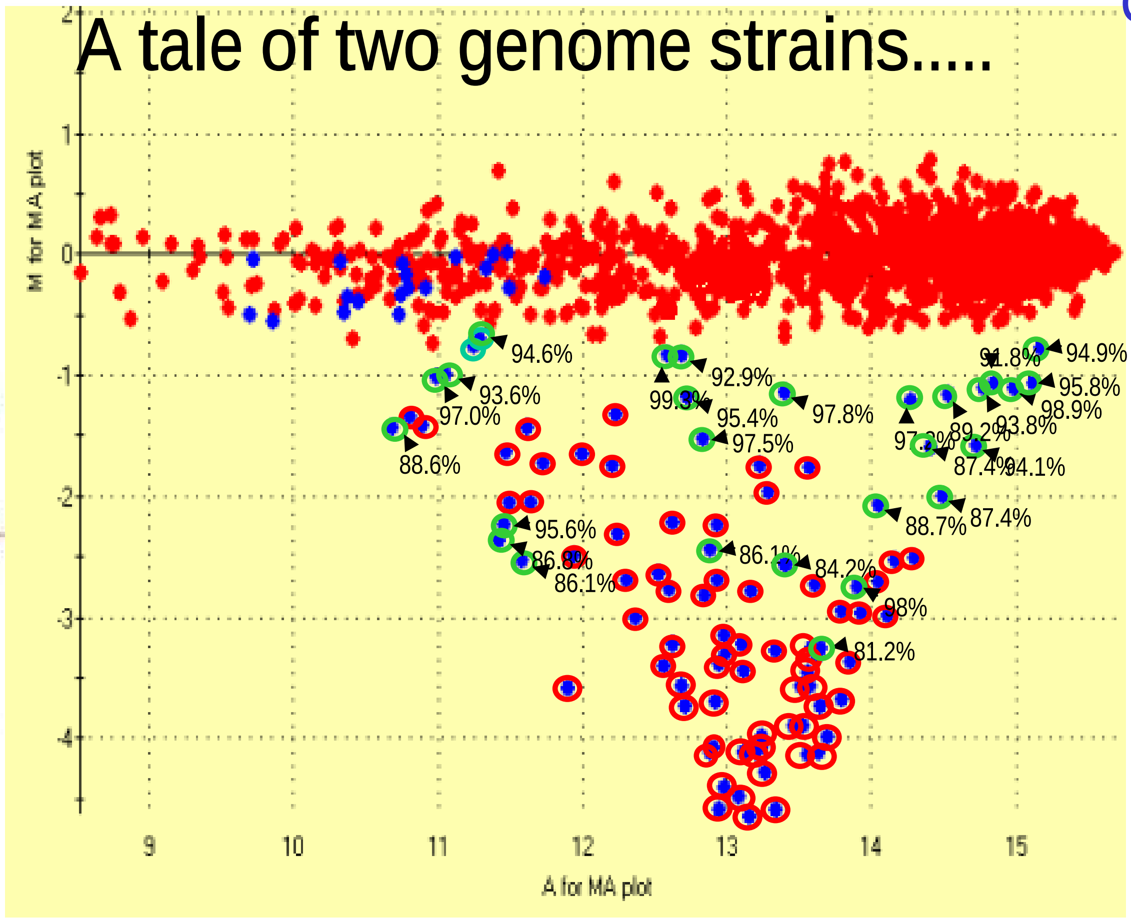
A tale of two genome strains.....

C. jejuni NCTC11168
vs.
C. jejuni RM1221

● Absent in RM1221

What is the cut-off for Absent vs. Divergent calls ?

A tale of two genome strains.....



C. jejuni NCTC11168
vs.
C. jejuni RM1221

~ 95% identity +

~ 90% identity

~ 85% identity

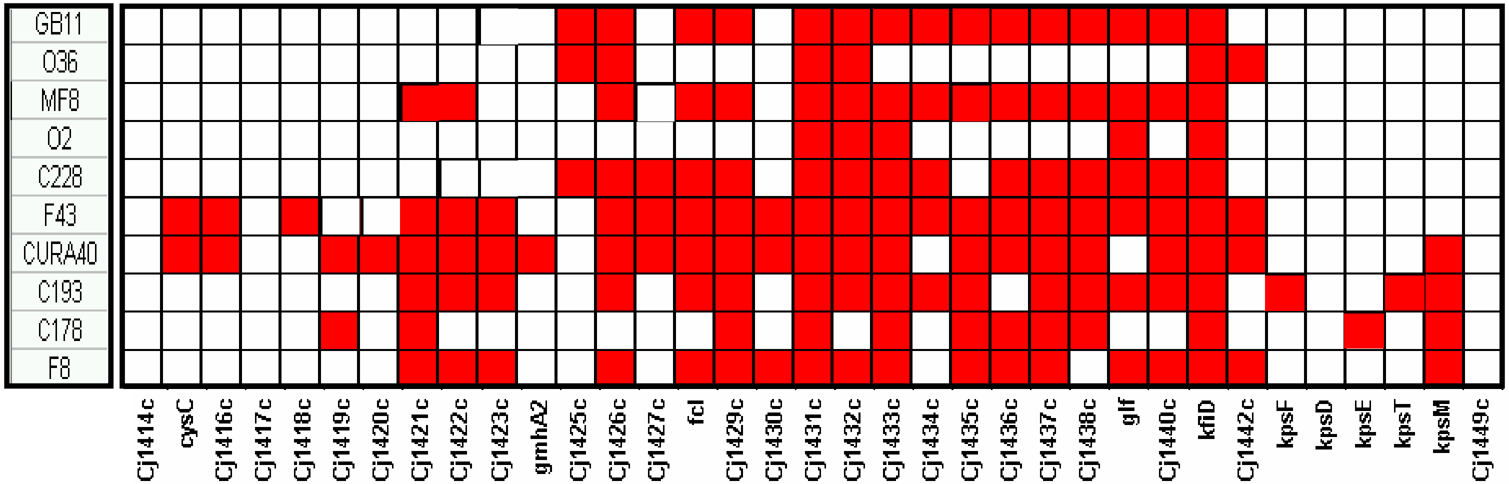
~ 80% identity or less

● Absent in RM1221

● Divergent in RM1221

there **is** a correlation between log ratio and sequence conservation

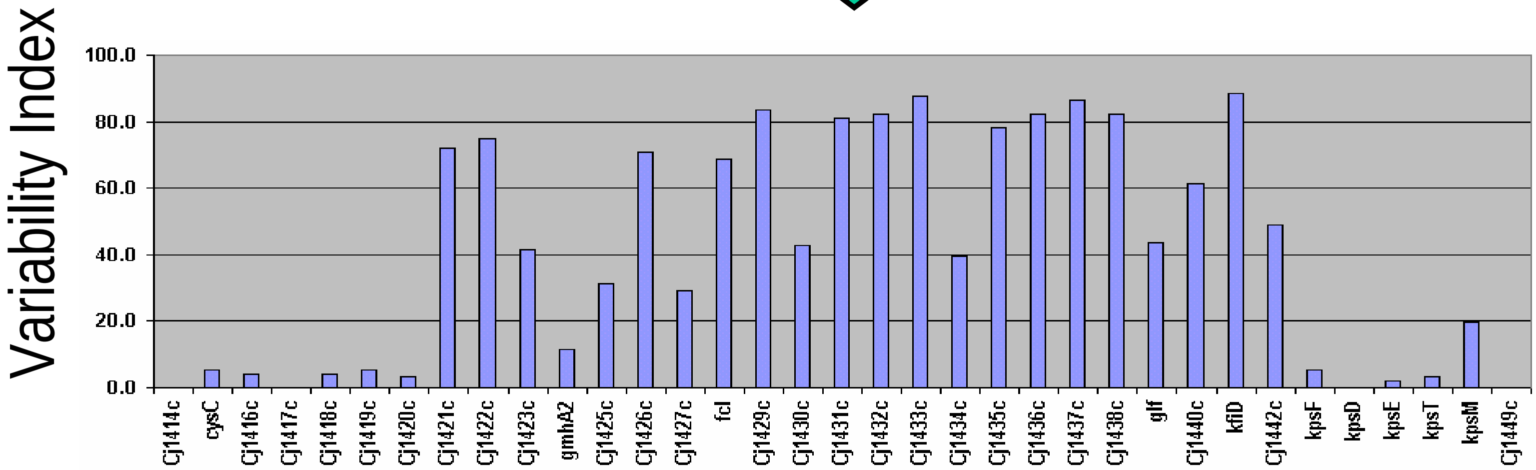
Tabulating genetic variability from CGH data



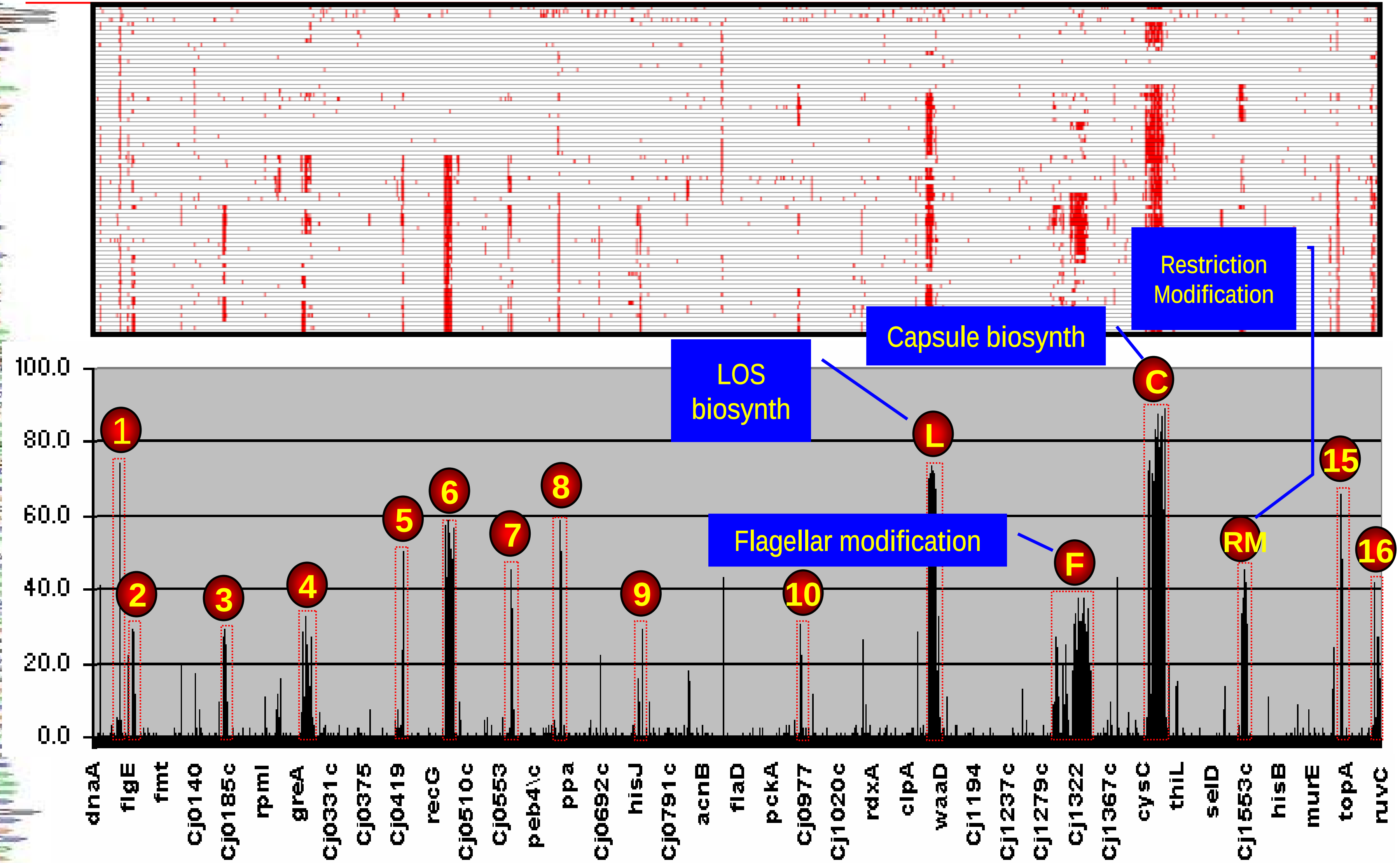
	Cj1414c	cysC	Cj1416c	Cj1417c	Cj1418c	Cj1419c	Cj1420c	Cj1421c	Cj1422c	Cj1423c	gmhA2	Cj1425c	Cj1426c	Cj1427c	fcl	Cj1429c	Cj1430c	Cj1431c	Cj1432c	Cj1433c	Cj1434c	Cj1435c	Cj1436c	Cj1437c	Cj1438c	glf	Cj1440c	kfiD	Cj1442c	kpsF	kpsD	kpsE	kpsT	kpsM	Cj1449c
GB11	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟥	🟥	⬜	🟥	🟥	⬜	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	⬜	⬜	⬜	⬜	⬜	⬜	⬜
O36	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟥	🟥	⬜	⬜	⬜	⬜	🟥	🟥	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟥	🟥	⬜	⬜	⬜	⬜	⬜	⬜
MF8	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟥	🟥	⬜	⬜	⬜	🟥	⬜	🟥	🟥	⬜	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	⬜	⬜	⬜	⬜	⬜	⬜	⬜
O2	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟥	🟥	🟥	⬜	⬜	⬜	⬜	⬜	🟥	⬜	🟥	⬜	⬜	⬜	⬜	⬜	⬜	⬜
C228	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟥	🟥	🟥	🟥	🟥	⬜	🟥	🟥	🟥	🟥	⬜	🟥	🟥	🟥	🟥	🟥	🟥	⬜	⬜	⬜	⬜	⬜	⬜	⬜
F43	⬜	🟥	🟥	⬜	🟥	⬜	⬜	🟥	🟥	🟥	⬜	⬜	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	⬜	⬜	⬜	⬜	⬜	⬜
CURA40	⬜	🟥	🟥	⬜	⬜	🟥	🟥	🟥	🟥	🟥	🟥	⬜	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	⬜	🟥	🟥	🟥	🟥	⬜	🟥	🟥	🟥	⬜	⬜	⬜	⬜	🟥	⬜
C193	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟥	🟥	🟥	⬜	⬜	🟥	⬜	🟥	🟥	⬜	🟥	🟥	🟥	🟥	🟥	⬜	🟥	🟥	🟥	🟥	🟥	⬜	🟥	⬜	⬜	🟥	🟥	⬜
C178	⬜	⬜	⬜	⬜	⬜	🟥	⬜	🟥	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟥	⬜	🟥	⬜	🟥	⬜	🟥	🟥	🟥	🟥	⬜	⬜	🟥	⬜	⬜	⬜	🟥	⬜	🟥	⬜
F8	⬜	⬜	⬜	⬜	⬜	⬜	⬜	🟥	🟥	🟥	⬜	⬜	🟥	⬜	🟥	🟥	🟥	🟥	🟥	🟥	⬜	🟥	🟥	🟥	⬜	🟥	🟥	🟥	🟥	⬜	⬜	⬜	⬜	🟥	⬜

🟥 “diverged”

⬜ “conserved”



Genetic variability hotspots among Cj strains



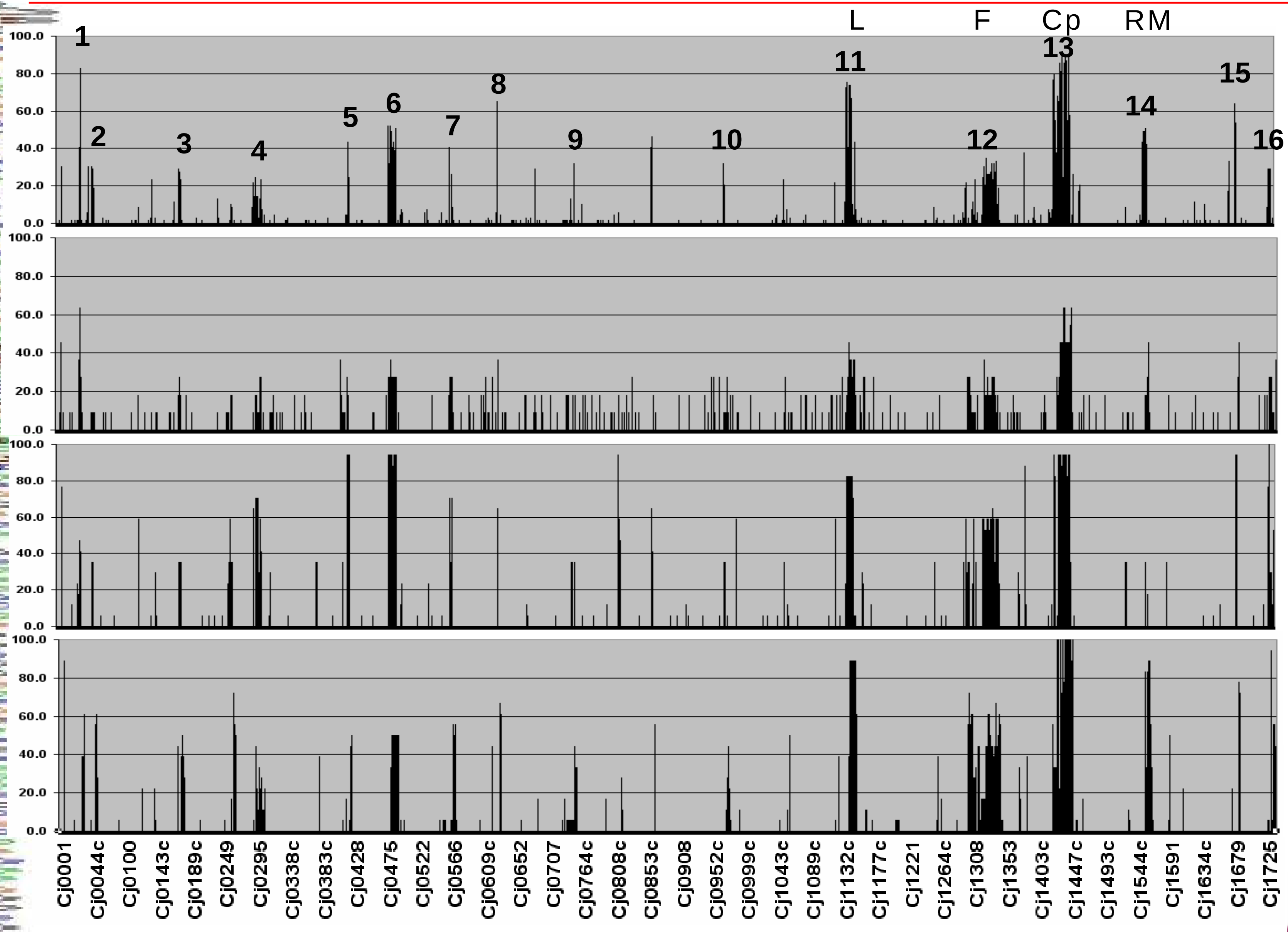
LOS biosynth

Capsule biosynth

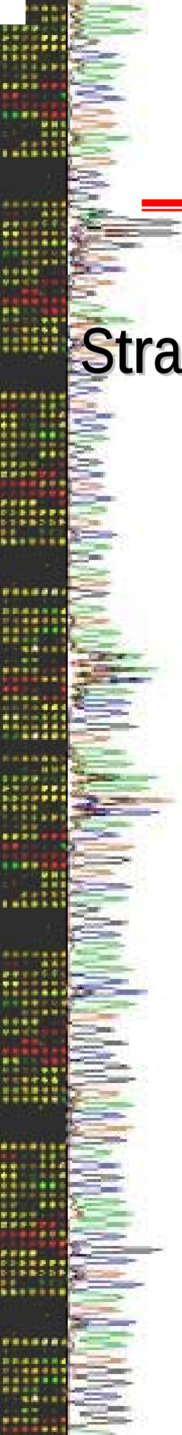
Restriction Modification

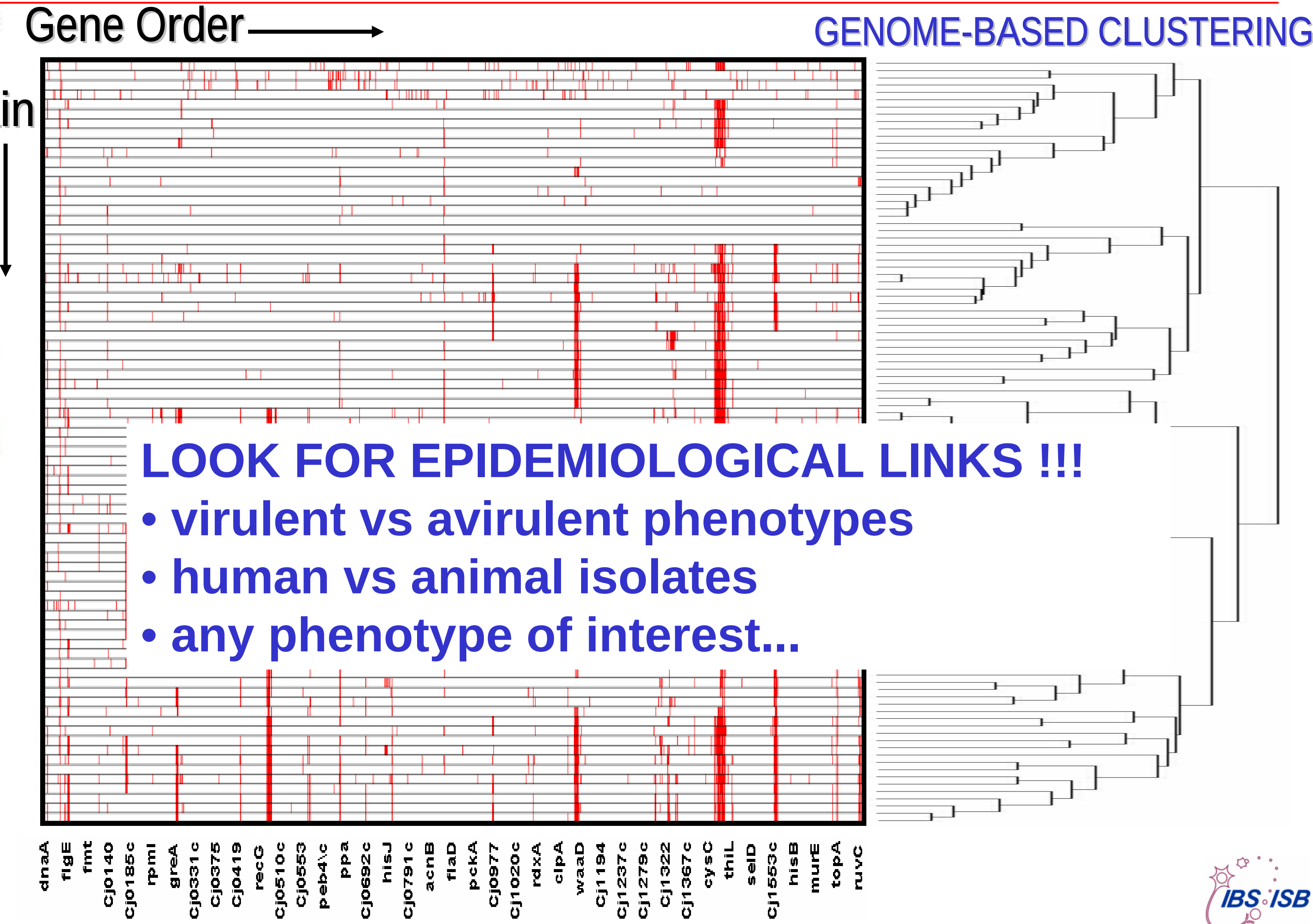
Flagellar modification

Meta-analysis of 97 *C. jejuni* strains



Genetic variability in C. jejuni

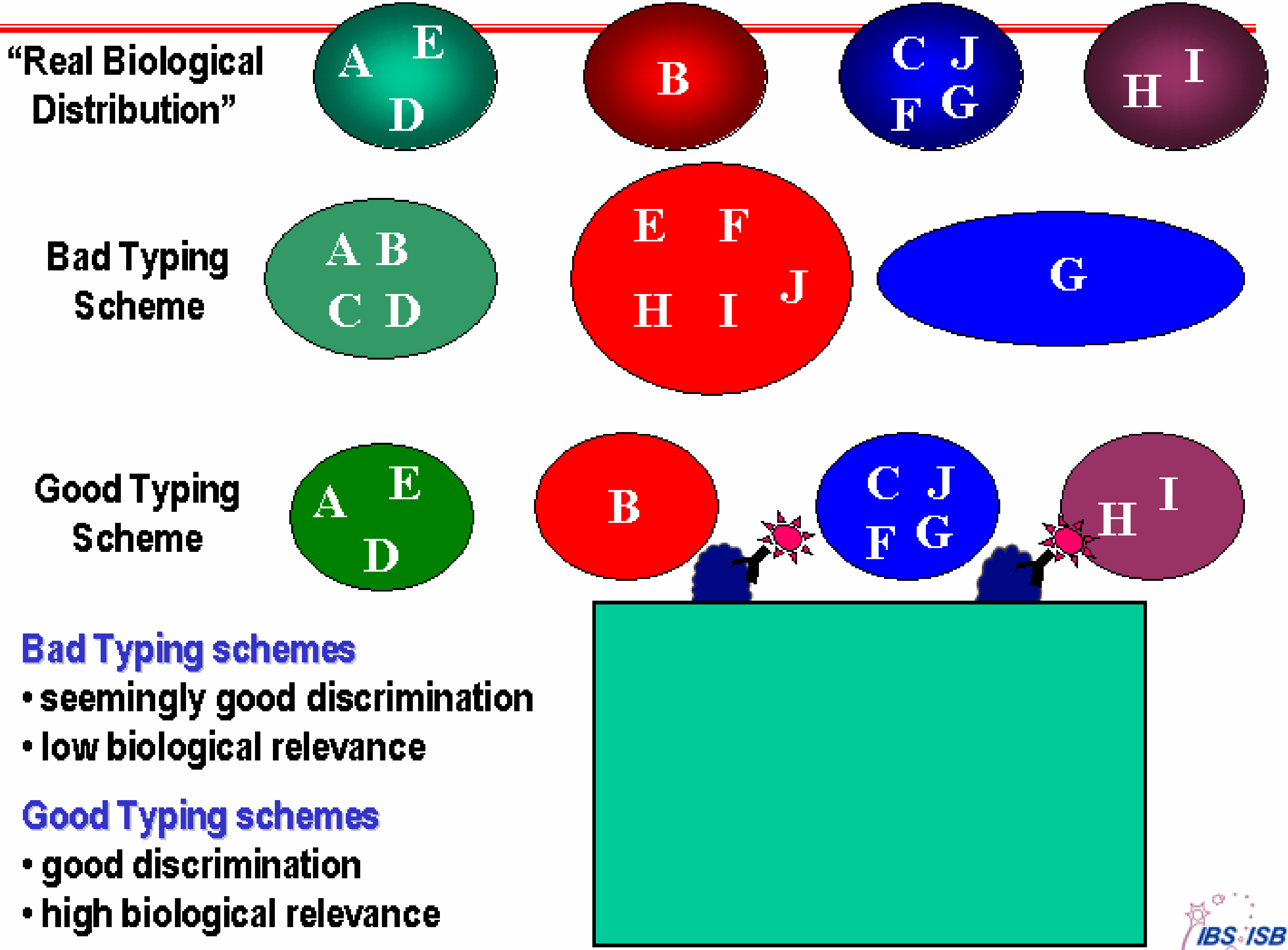




LOOK FOR EPIDEMIOLOGICAL LINKS !!!

- **virulent vs avirulent phenotypes**
- **human vs animal isolates**
- **any phenotype of interest...**

(The Problem with) Molecular Typing



"Real Biological Distribution"

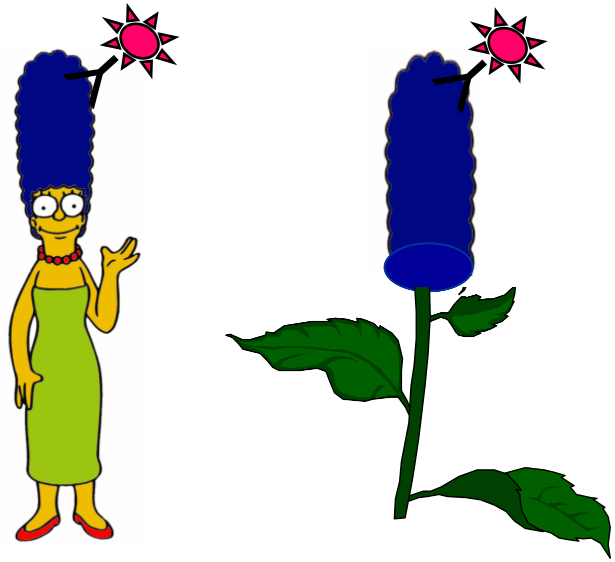
Bad Typing Scheme

Good Typing Scheme

Bad Typing schemes
- seemingly good discrimination
- low biological relevance

Good Typing schemes
- good discrimination
- high biological relevance

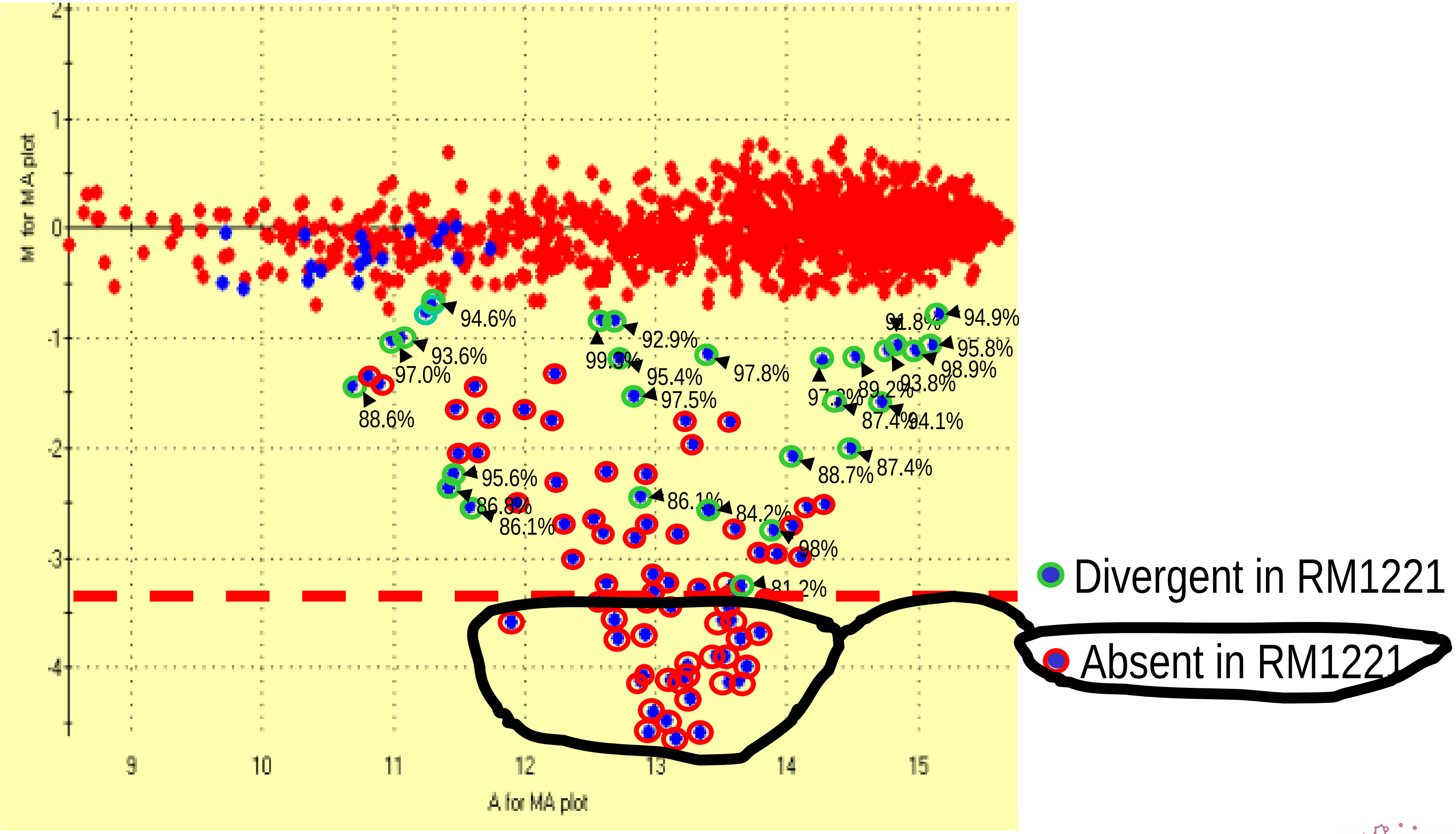
The Problem with Molecular Typing



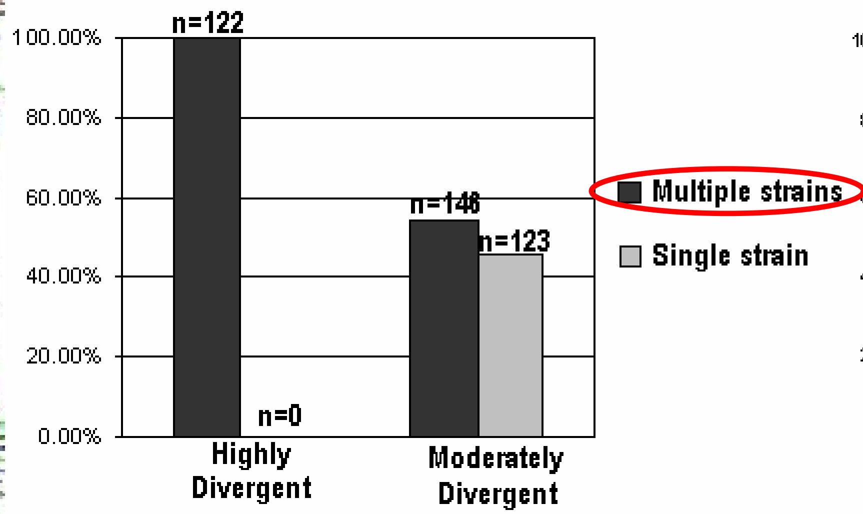
CGH-based typing ???

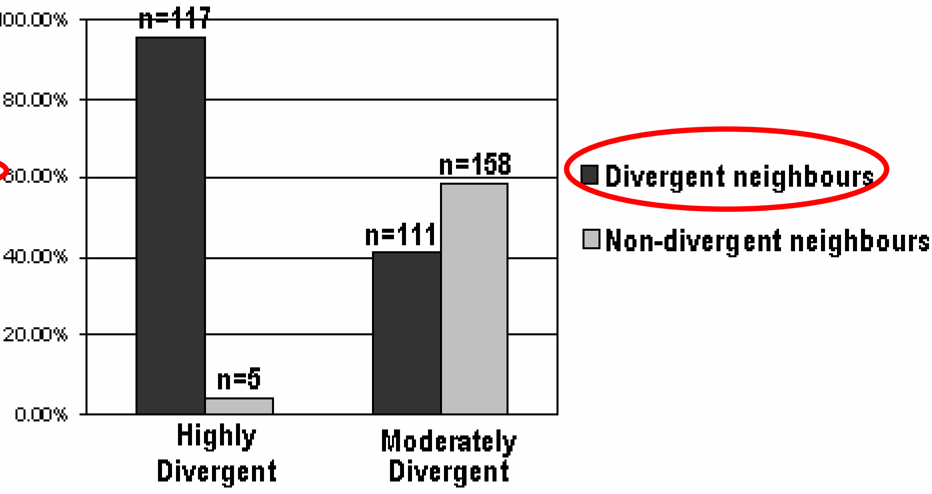
To be useful...

- good discrimination
- robust assay
- relevant



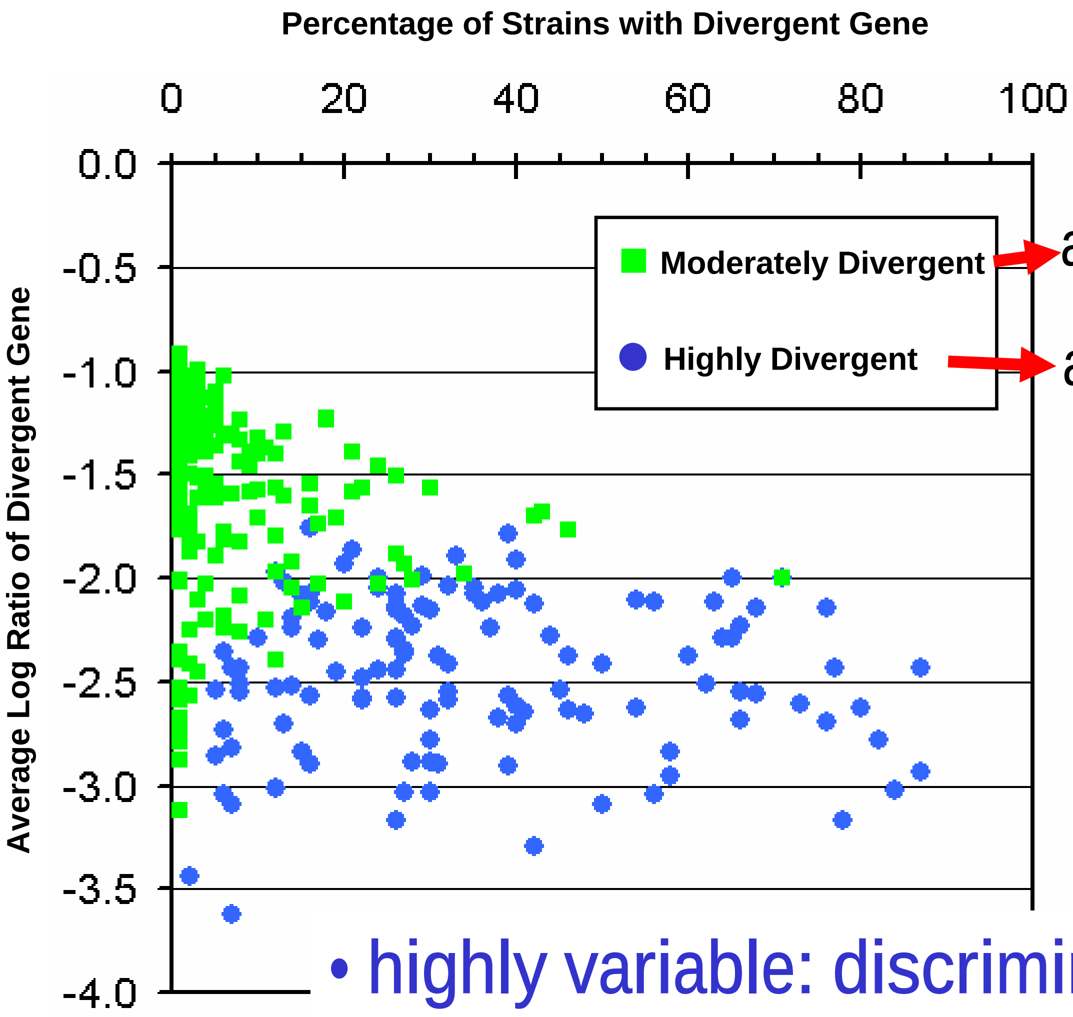
Filtering for Robust CGH data...





- highly divergent genes have high intra-species variability
- highly divergent genes tend to have divergent neighbours
  - robustness !!!!

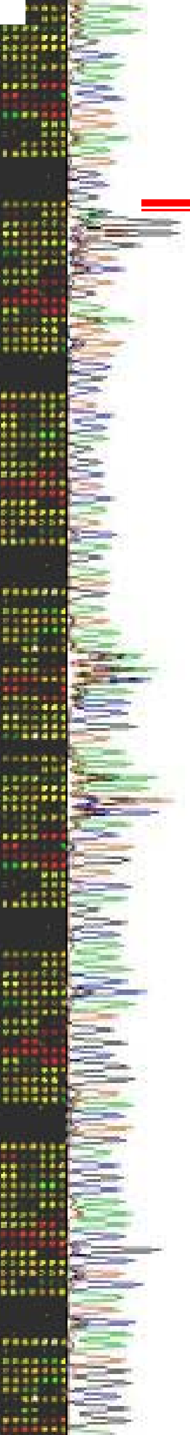
Good polymorphic markers for genotyping



avg # ~ 6

avg # ~ 35

- highly variable: discriminating
- highly divergent: robust
- (biologically) relevant



Microarrays and Studies on Virulence...

- We have developed a full genome microarray for *C. jejuni*

- Used microarray to study gene expression
  - expression profiles in virulent and avirulent strains
  - expression profiles of knock-out mutants

- Used microarray to study *C. jejuni* genetic variability
  - characterized genomic diversity in the species
    - mining data for association with virulence
  - hypervariable genes for molecular typing

Acknowledgements

- NRC:
  - IBS Pathogen Genomics: John Nash, Cathy Carrillo, Brian Agnew, Anne Bouevitch, Simon Foote, Oksana Mykytczuk, Wendy Findlay, Chris Luebbert, Michel Gilbert, Christine Szymanski, Warren Wakarchuk, Martin Young
  - Students: Mike Roberts, Su-Harn Hwang, Rey Acedillo
  - BRI Microarray Facility: Daniel Tessier *et. al.*
  - PBI Robotics Lab

- (future and past) Collaborators:
  - Health Canada: Jeff Farber *et. al.* (BMH), Kris Rahn *et al.* (LFZ), Cliff Clark *et al.* (NML)
  - VIDO: Andy Potter *et al.*
  - Erasmus University: Hubert Endtz *et al.*
  - Dokkyo University: Nobuhiro Suki *et al.*

  - NRC Genomics and Health Initiative
  - Human Frontier Science Program (Japan)
  - HC-LFZ (Guelph)