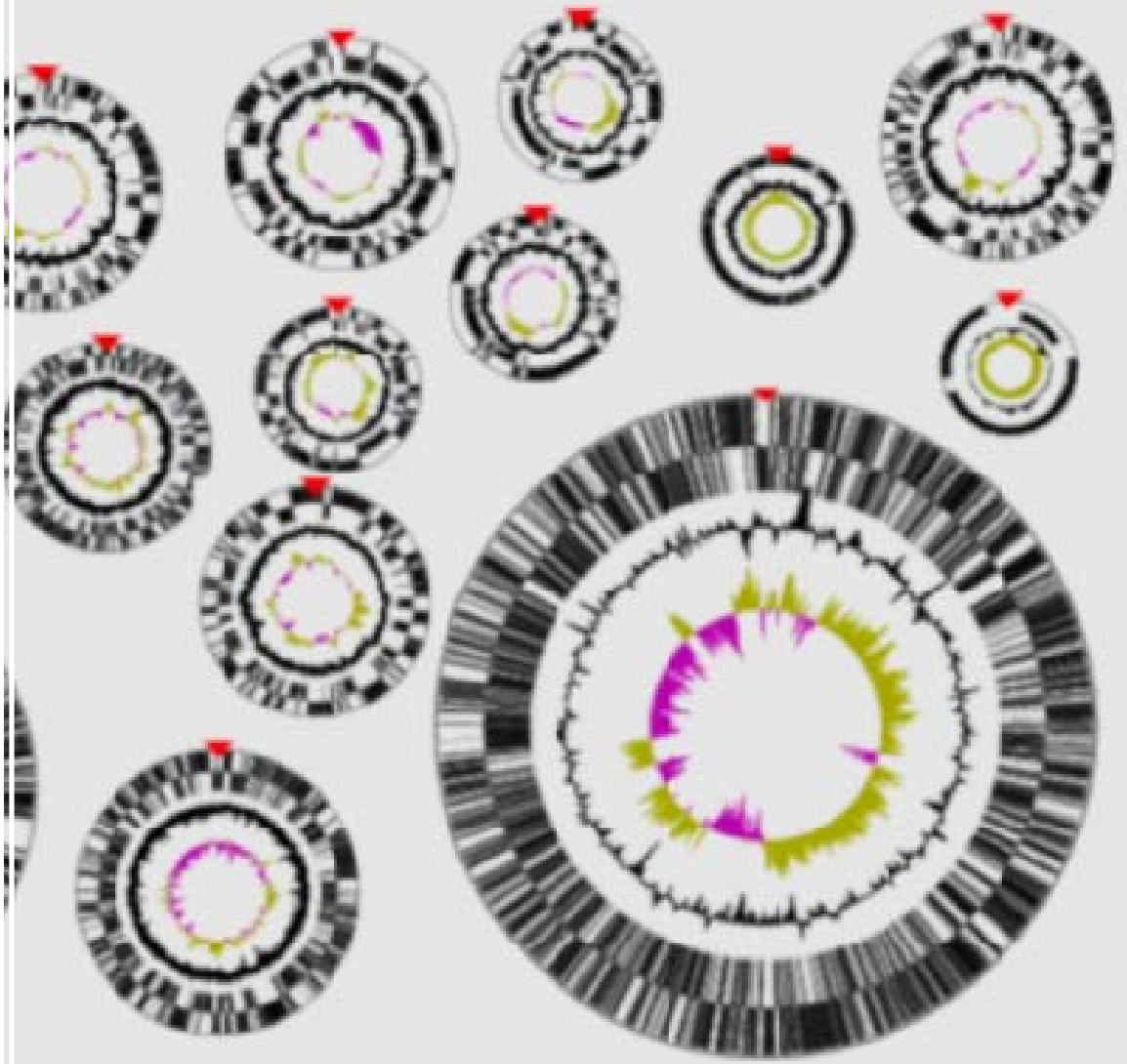
KASETSART UNIVERSITY FACULTY OF FISHERIES
คณะประมง มหาวิทยาลัยเกษตรศาสตร์

Extracting Biologically Meaningful Patterns from Bacterial Genome Data

Pattanapon Kayansamruaj, PhD

Department of Aquaculture, Faculty of Fisheries, Kasetsart University



Overview

- Impact of Next-Generation sequencing to science
- Generic workflow in bacterial genome analysis
- High-throughput screening of bacterial genomes
- Comparative genomics and Phylogenomics
- Getting published

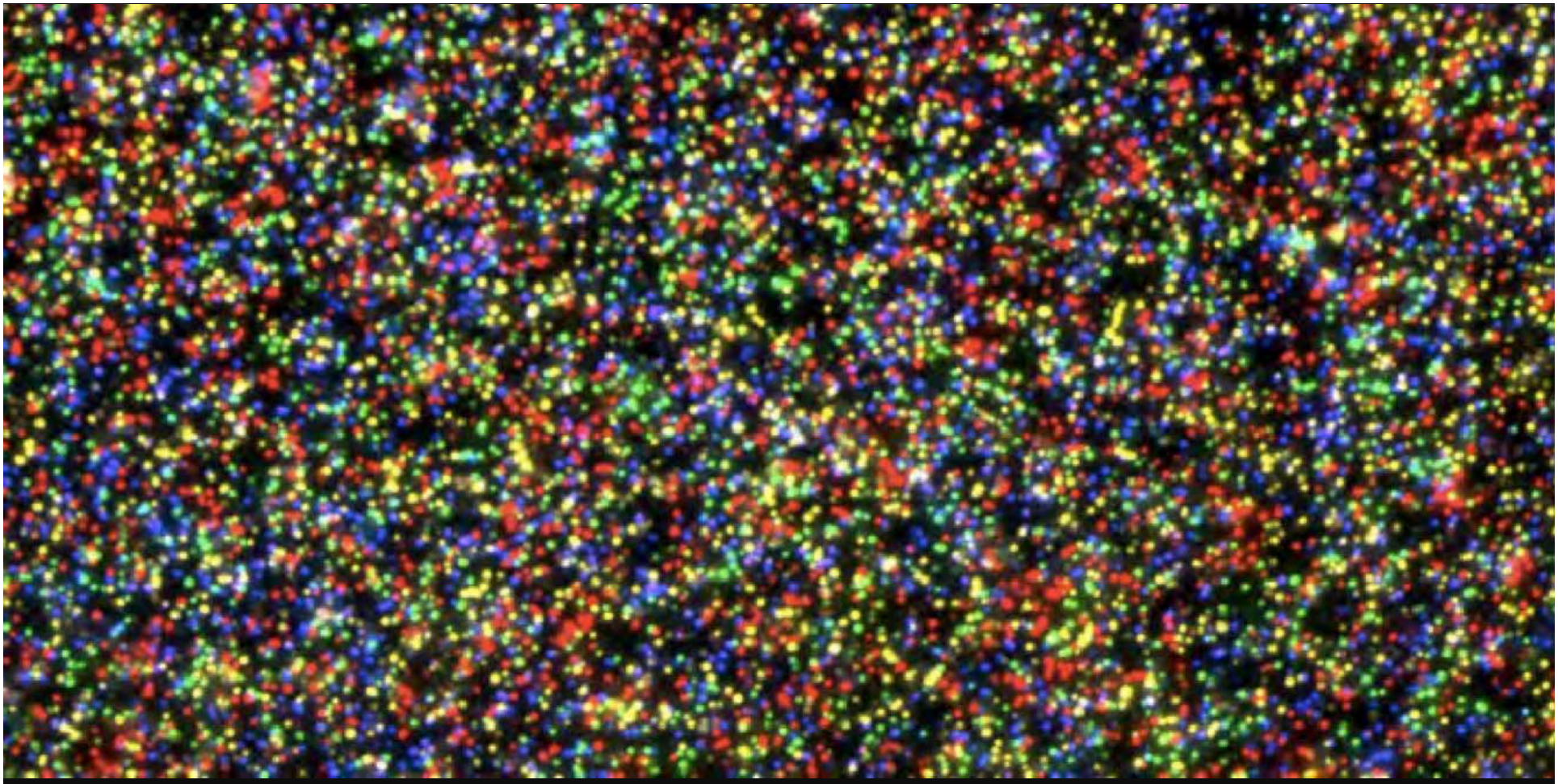
frontiers in Microbiology

One reviewer's comment to a manuscript conducting comparative genomics study

"The work is **very descriptive** and **does not have a real hypothesis** that can be tested…"

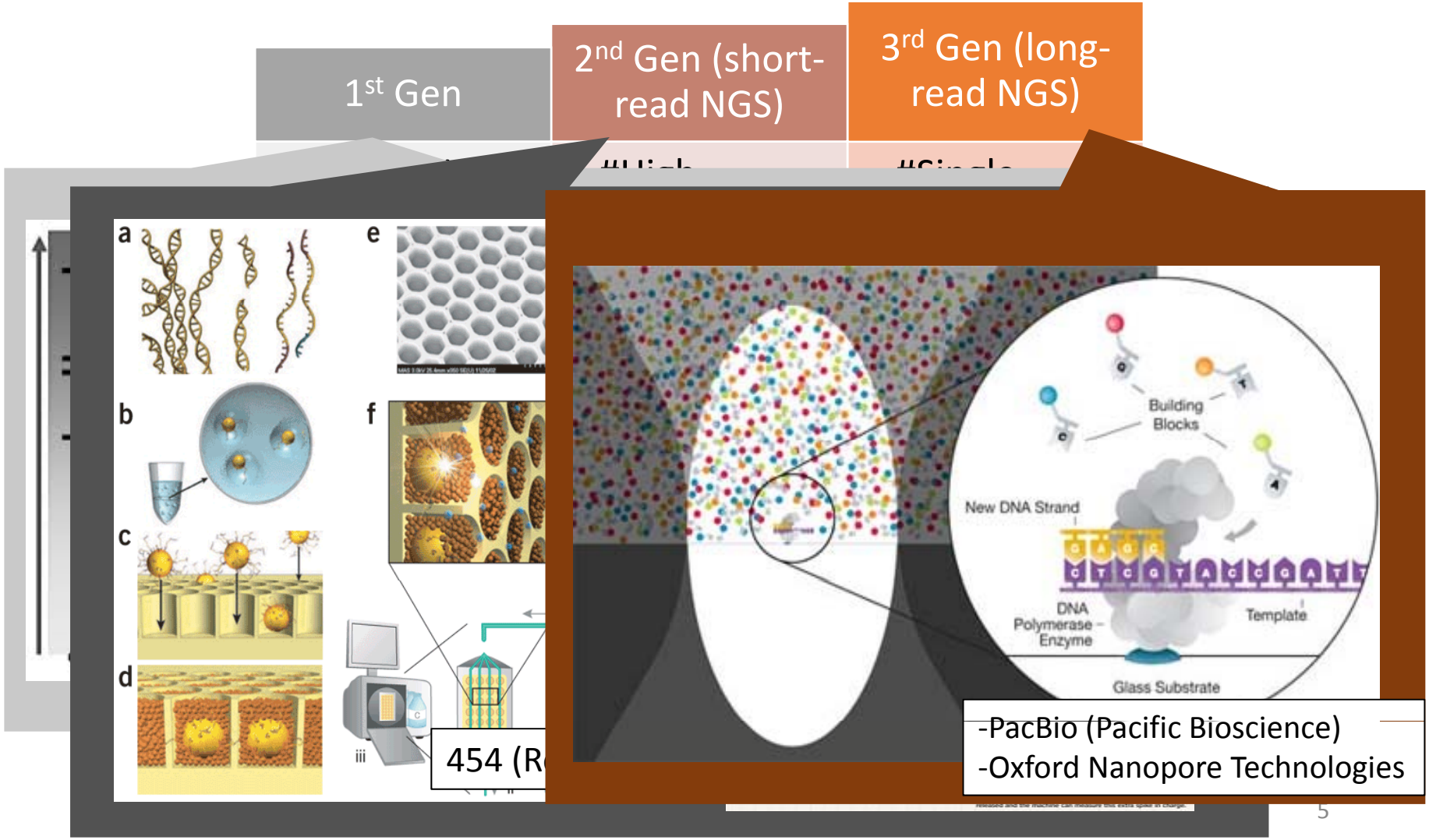
And the authors replied…

"…thought this manuscript **has no hypothesis** to be tested and that this observation somehow diminished the value of the study. We can simply point out that for "genome" analyses **that is universally true**. Such studies are **often hypothesis generating** and that perhaps is one of **their primary purposes**."



Impact of Next-Generation sequencing to science

What is Next-Generation sequencing?



NGS Platform available in the market

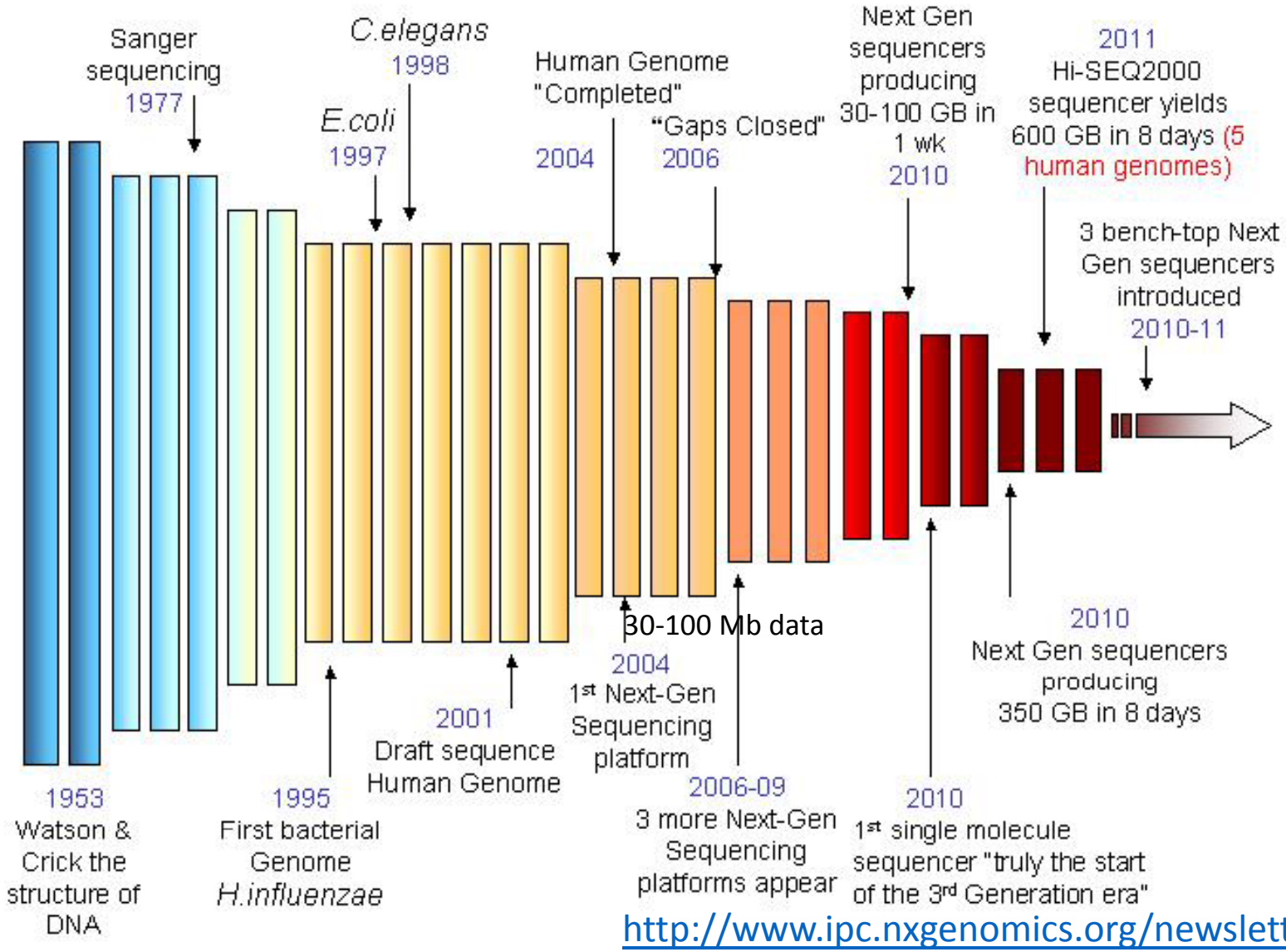
Short-read NGS

- 454 (Roche)
- SOLiD (Thermo Fisher)
- Complete Genomics (BGI)
- Illumina
- Ion Torrent (Thermo Fisher)

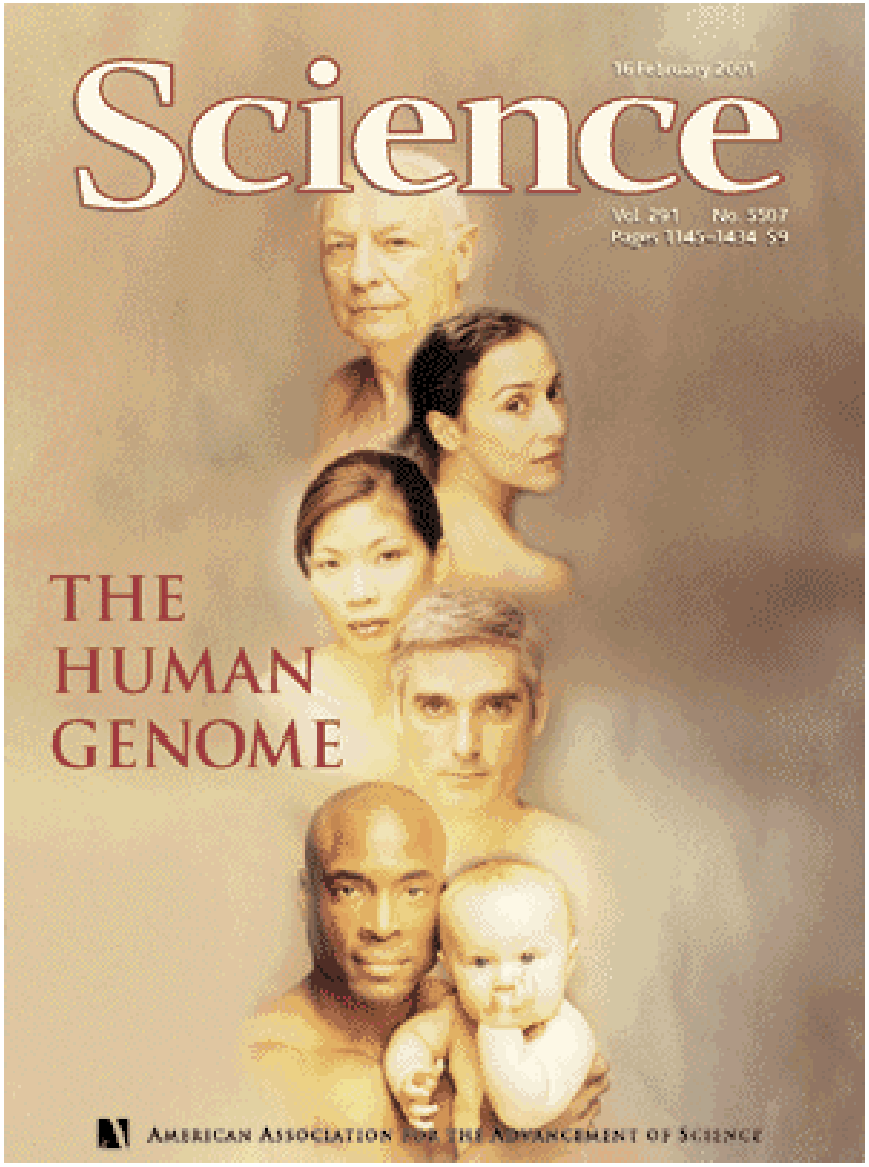
Long-read NGS

- PacBio (Pacific Biosciences)
- Oxford Nanopore Technologies

Sequencing technology timeline



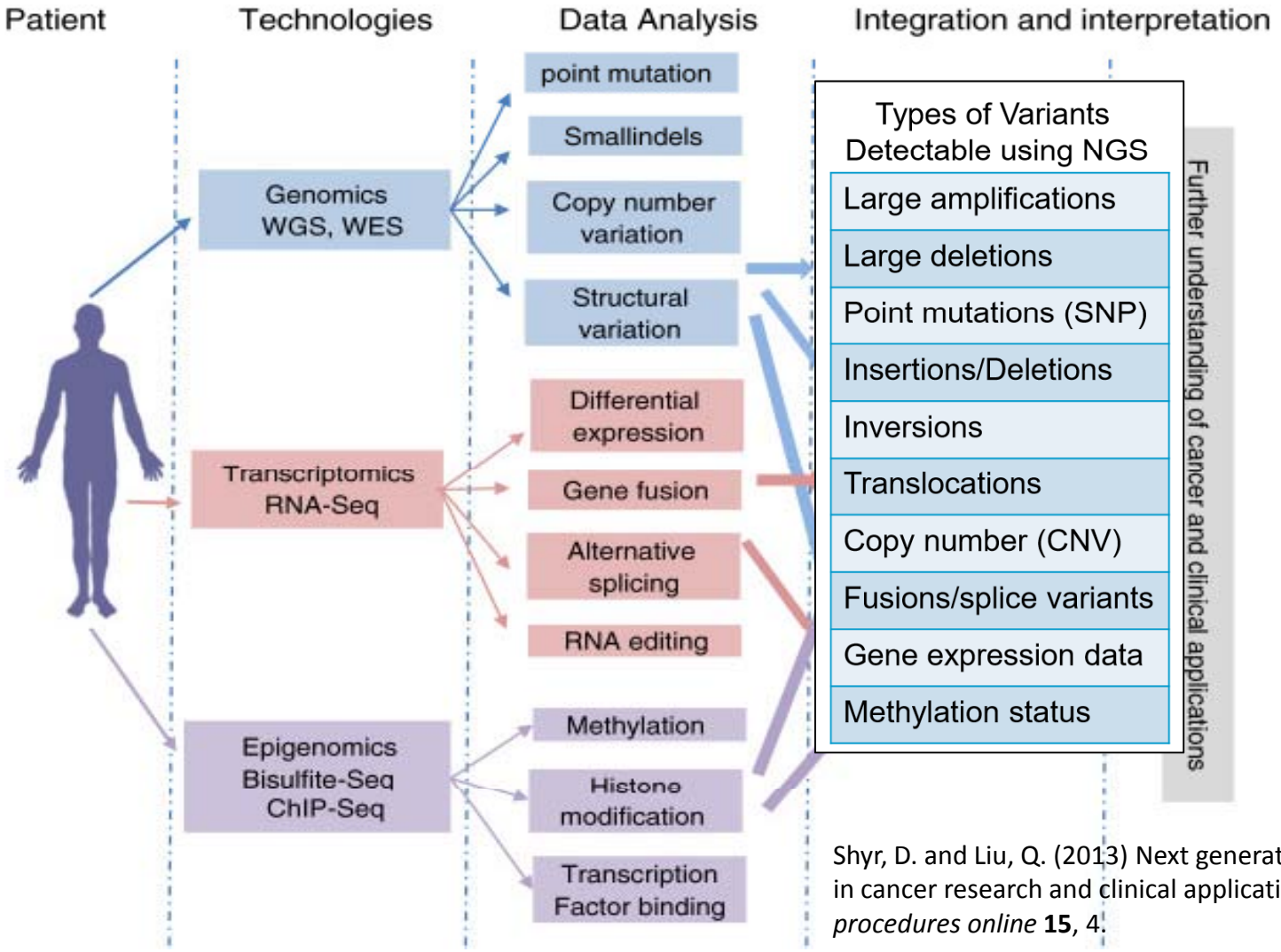
http://www.ipc.nxgenomics.org/newsletter.htm



- The first milestone for high-throughput analysis
- Initiated ~1990
- First draft of human genome was published in 2001 (92% of the genome)
- BAC clone and genome walking technique
- Multiple bioinformatic tools have been invented and polished to hasten analytical process
  - Sequence assembly
  - Genome annotation (gene finding)
  - Distributed computing

- Nowadays, the cost of sequencing a human genome down to around **US$1,000** (finished within a week)

Application for basic and clinical research



Shyr, D. and Liu, Q. (2013) Next generation sequencing in cancer research and clinical application. *Biological procedures online* **15**, 4.

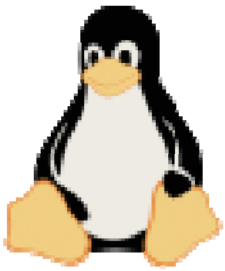
Generic workflow in bacterial genome analysis

Working with NGS data

Is there a perfect OS for NGS analysis

Windows

Mac



Linux

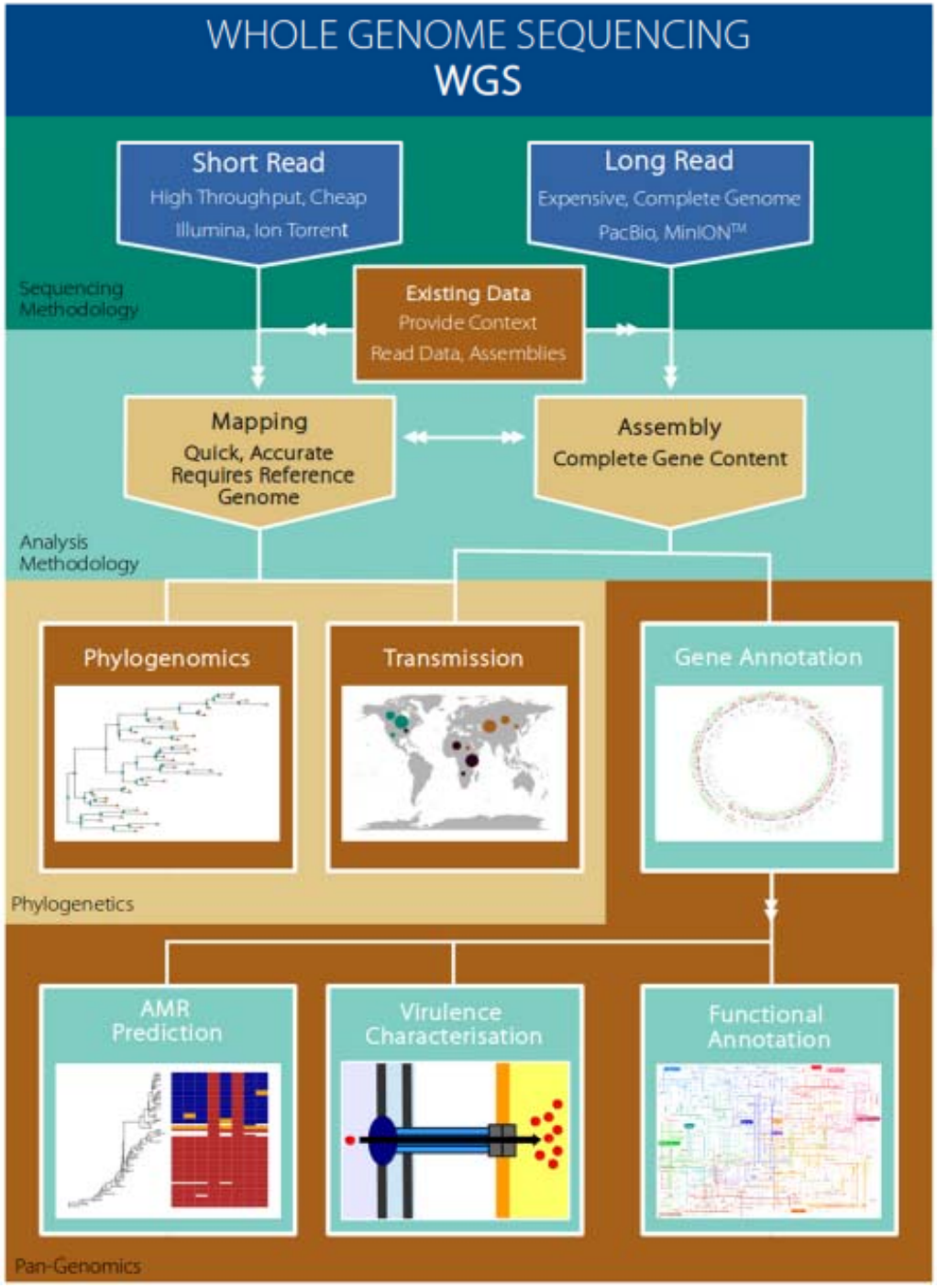
https://my.vmware.com/web/vmware/downloads

- *Program compatibility*: **Linux > MacOS >>>>>>> Windows**
- Most of NGS data analysis packages run on Unix-based OS (Linux, Mac)
- For Windows user …
  - Dual installation of Linux (Ubuntu is recommended)
  - Use virtual machine program (VMware) → limited computing performance

Working with NGS data

Operate computational analysis by…

Local computing	***Computer networking (server)***	***Web-based analysis (via web browser)***
• Most versatile, a lot of experimentation • Time-consuming, not applicable for large dataset	• High performance computing • Limited programs to use • *Authorization?*	• Fit for basic users • OS independent • Does not require hi-end desktop • Large variety of application and computational power



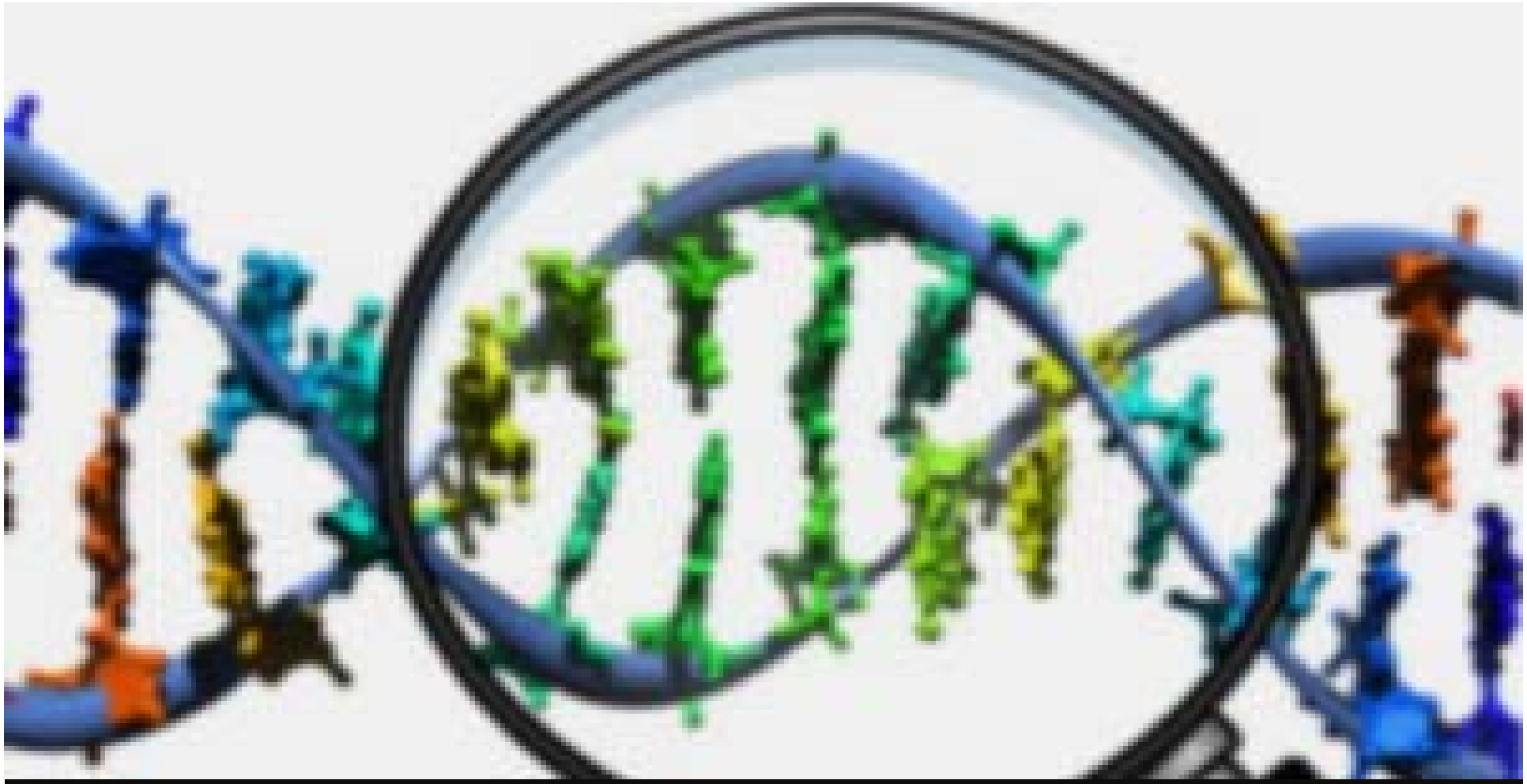
Pre-assembly analysis

- **Input**: Raw reads (.fastq)
- **Output** : Contigs, scaffolds, complete genome (.fasta)

Post-assembly analysis

- Data mining, Data interpreting, Visualization
- **Input**: .fasta
- **Output** : Biologically meaningful result

Bayliss, S.C., et al. (2017) The Promise of Whole Genome Pathogen Sequencing for the Molecular Epidemiology of Emerging Aquaculture Pathogens. *Frontiers in microbiology* **8**, 121.



High-throughput screening of bacterial genomes

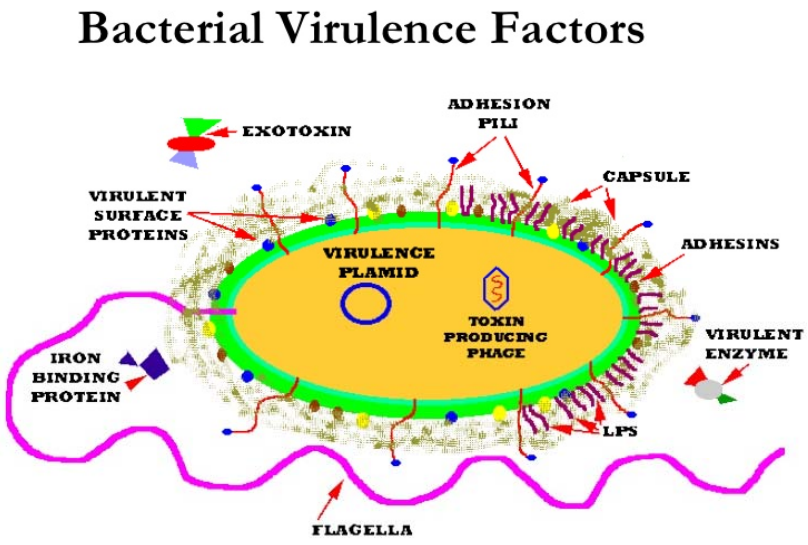
Identification gene of interest from bacterial genome

- Important elements in bacterial genome
  - Virulence factors
  - Antibiotic resistance genes
  - Genomic island
  - Bacteriophage
  - CRISPR
  - Prokaryotic Regulatory Proteins

Public databases enlisting important bacterial elements

Virulence factor

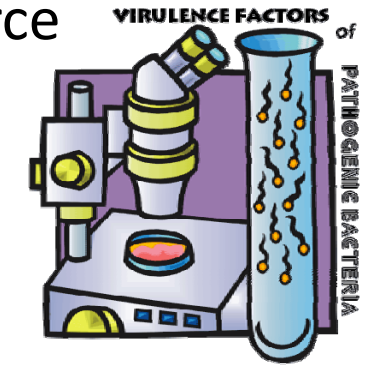
- **Virulence factors** refer to the properties (i.e., gene products) that enable a microorganism to establish itself on or within a host of a particular species and enhance its potential to cause disease.
- *e.g.* bacterial toxins, surface proteins mediated attachment, cell surface carbohydrates and proteins that protect a bacterium, and hydrolytic enzymes



Public databases enlisting important bacterial elements

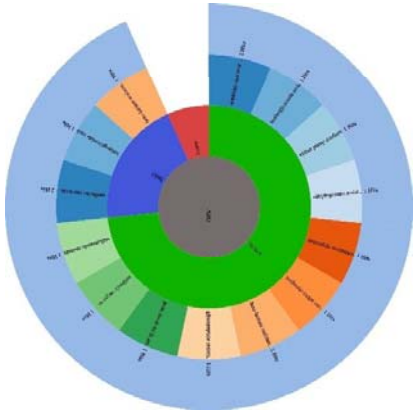
Virulence factor

- Virulence Factor Database (**VFDB**) → online resource for curating information about virulence factors of bacterial pathogens (*http://www.mgc.ac.cn/VFs/*)



Antibiotic resistance genes

- The Comprehensive Antibiotic Resistance Database (**CARD**): http://arpcard.mcmaster.ca/
- Antibiotic Resistance Genes Database (**ARDB**): https://ardb.cbcb.umd.edu/



Public databases enlisting important bacterial elements

Genomic island

- IslandViewer(*www.pathogenomics.sfu.ca/islandviewer/*)
  - Curated database for GIs present in bacterial genomes
  - Predict GI from submitted genome using integrated tools

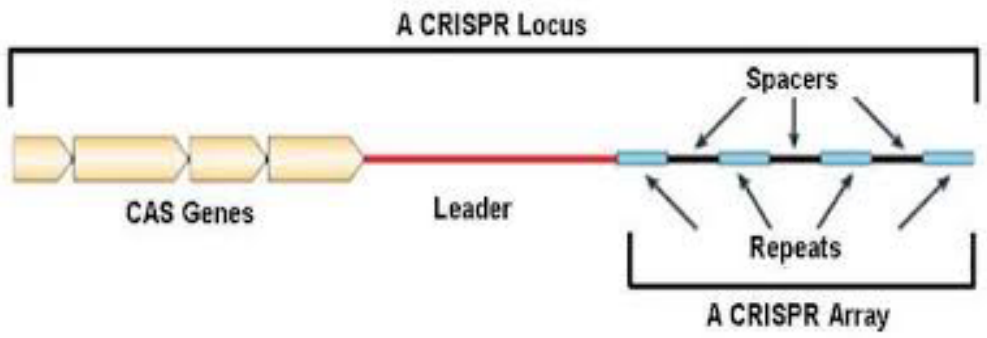
Bacteriophage

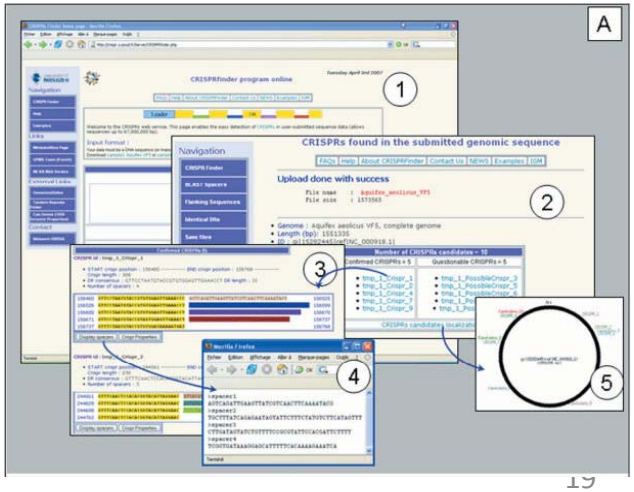
- Database
  - PhageDB (http://phagesdb.org/)
- Prediction tool
  - PHASTER (PHAge Search Tool - http://phaster.ca/
  - → web server for the rapid identification and annotation of prophage sequences within bacterial genomes and plasmids.

Public databases enlisting important bacterial elements

CRISPR

- Clustered Regularly Interspaced Short Palindromic Repeats
- A prokaryotic immune system that confers resistance to foreign genetic elements
- CRISPRFinder (http://crispr.i2bc.paris-saclay.fr/) → automated tool for identification of CRISPR array present in bacterial genome

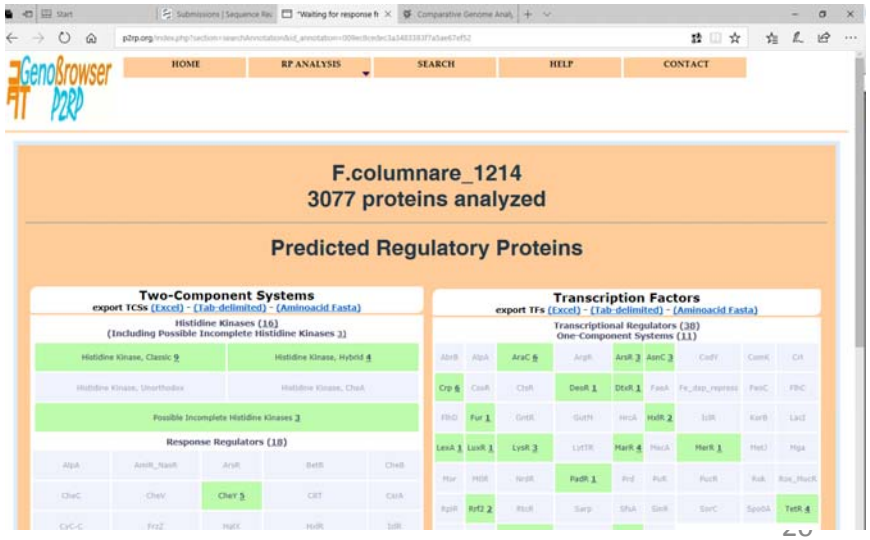




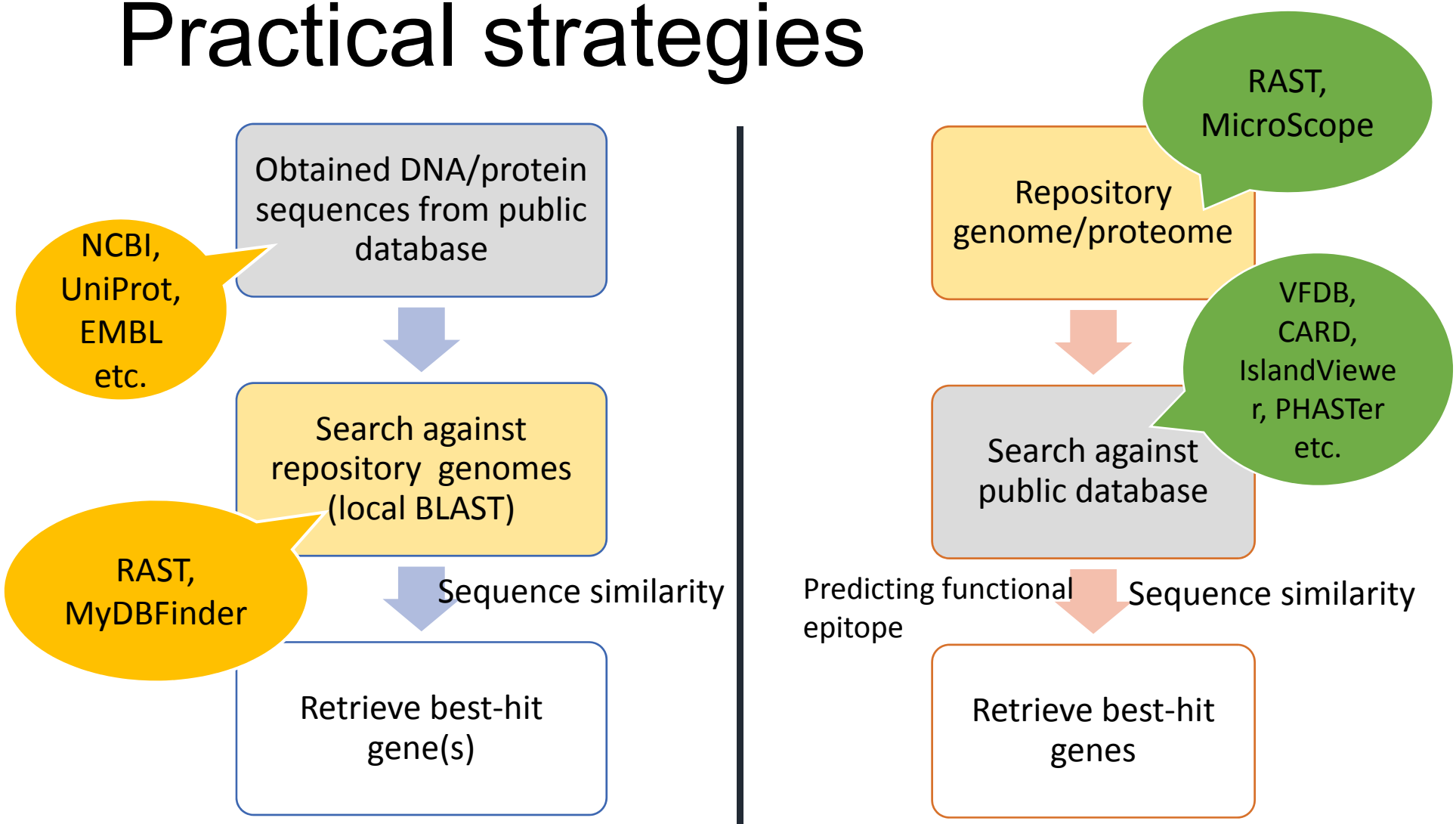
Public databases enlisting important bacterial elements

Prokaryotic regulatory proteins

- transcription factors (TFs) and two-component systems (TCSs)
- **P2RP** (Predicted Prokaryotic Regulatory Proteins)
  - http://www.p2rp.org/
  - Freely accessible web server, to provide a RP database and a platform for RPs prediction

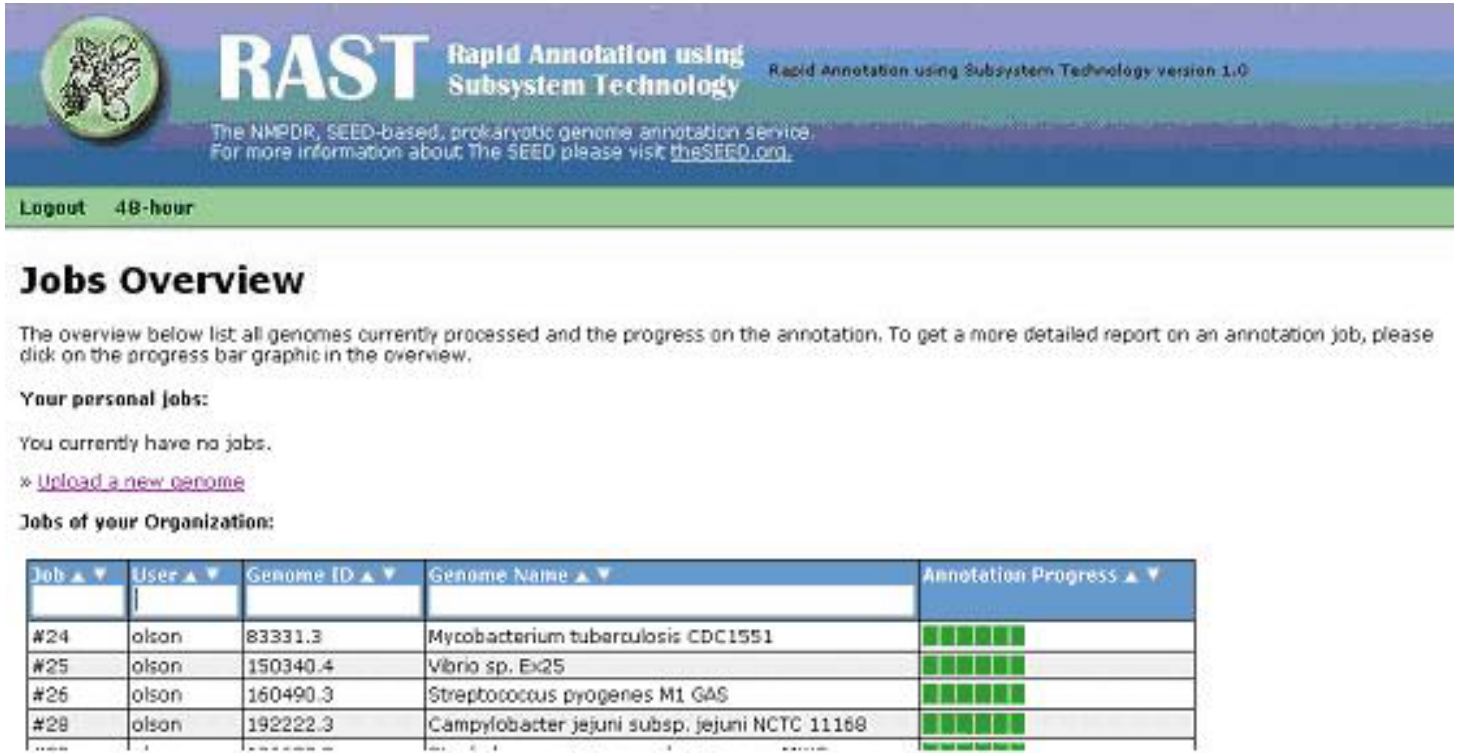


Searching gene of interest: Practical strategies



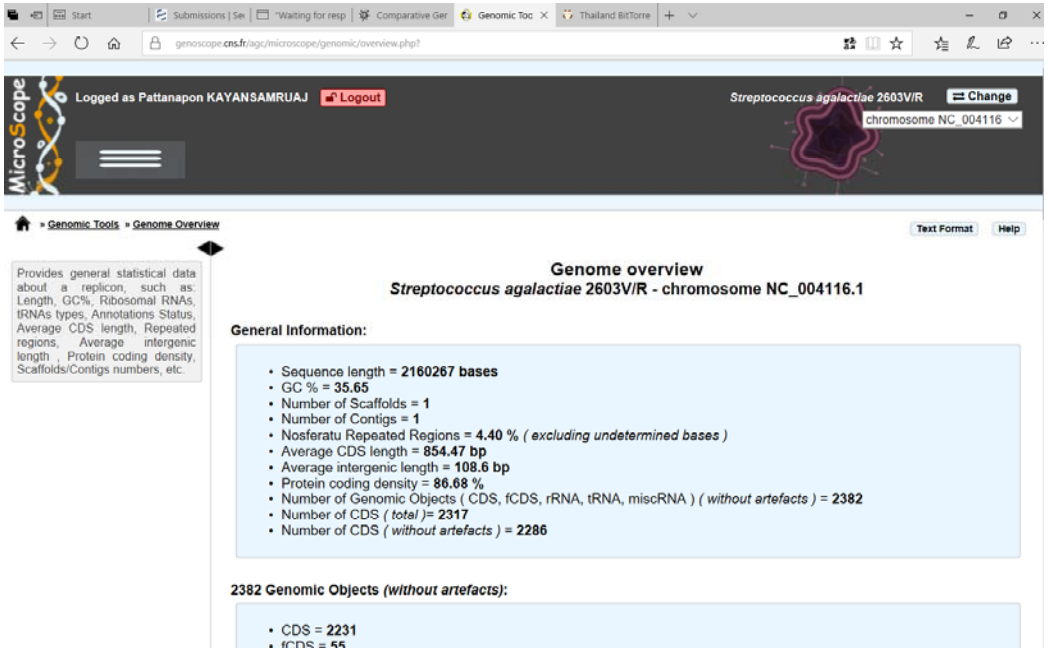
Identification gene of interest from bacterial genome

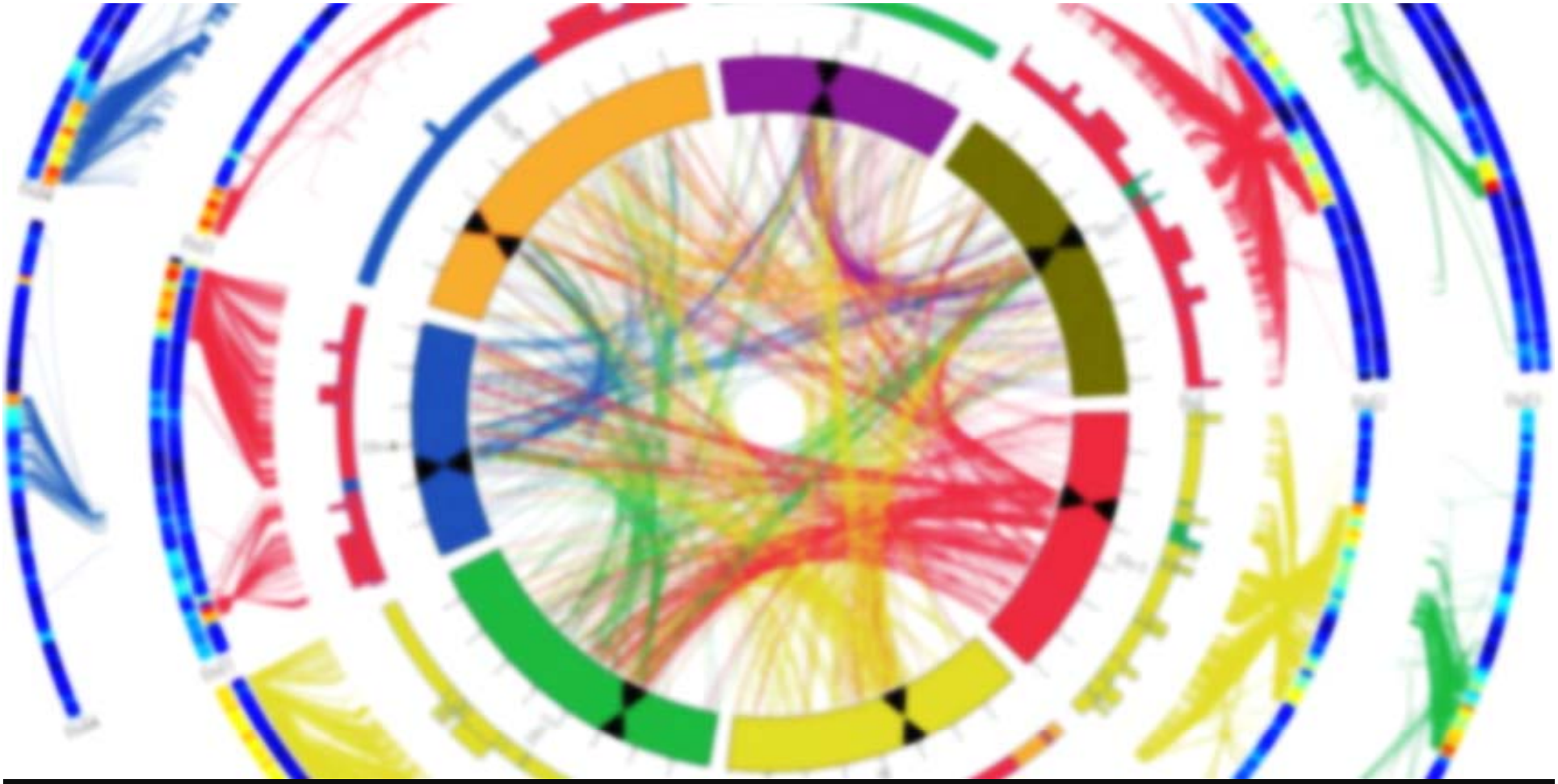
- Download and extract specific CDS from a genome curated in RAST server (http://rast.nmpdr.org/)



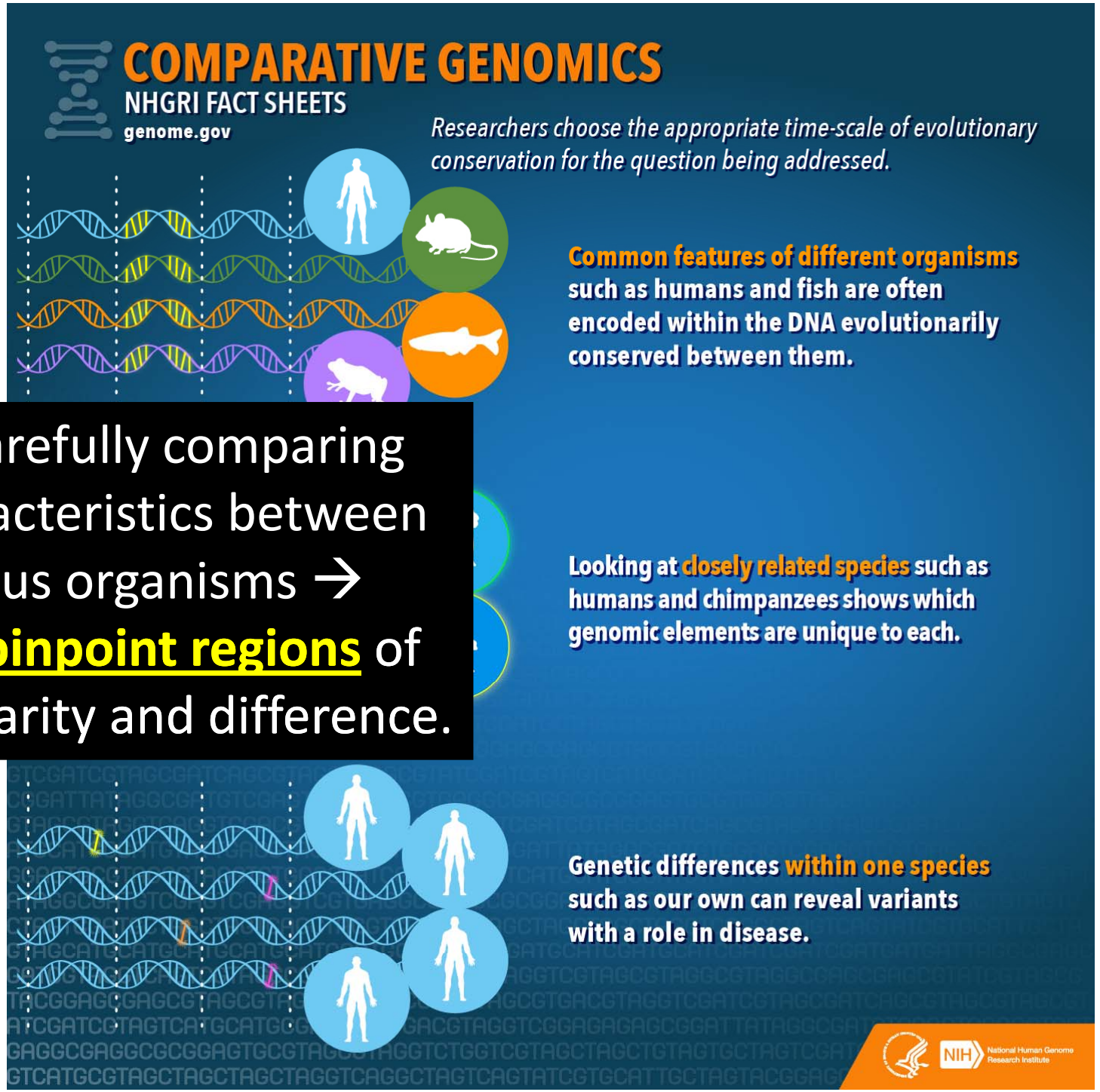
Identification gene of interest from bacterial genome

- **MicroScope** platform: a web-based platform for microbial comparative genome analysis and manual functional annotation (http://www.genoscope.cns.fr/agc/microscope/home/index.php)



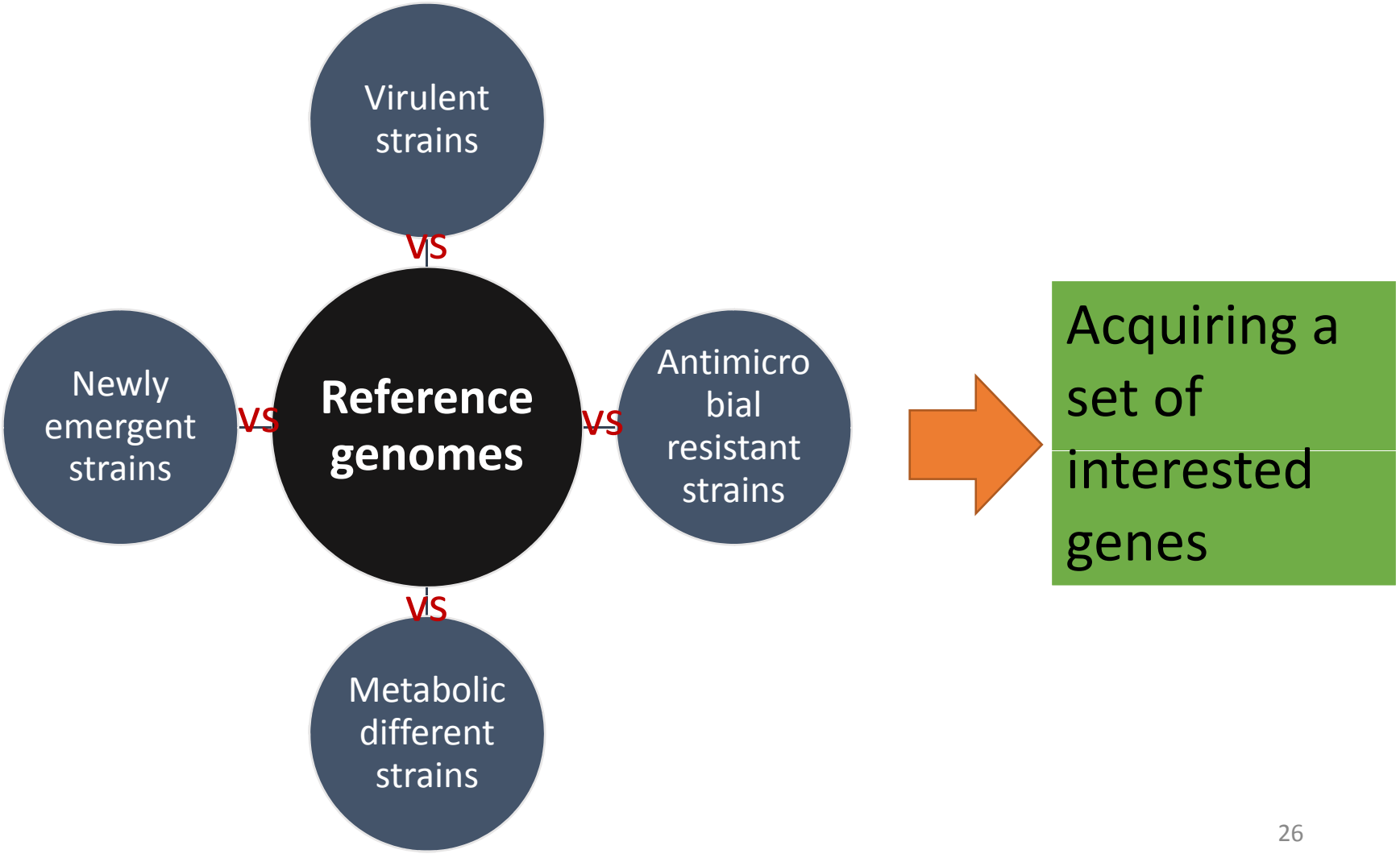


Comparative genomics and Phylogenomics



By carefully comparing characteristics between various organisms → can **pinpoint regions** of similarity and difference.

The benefits of comparative genomics

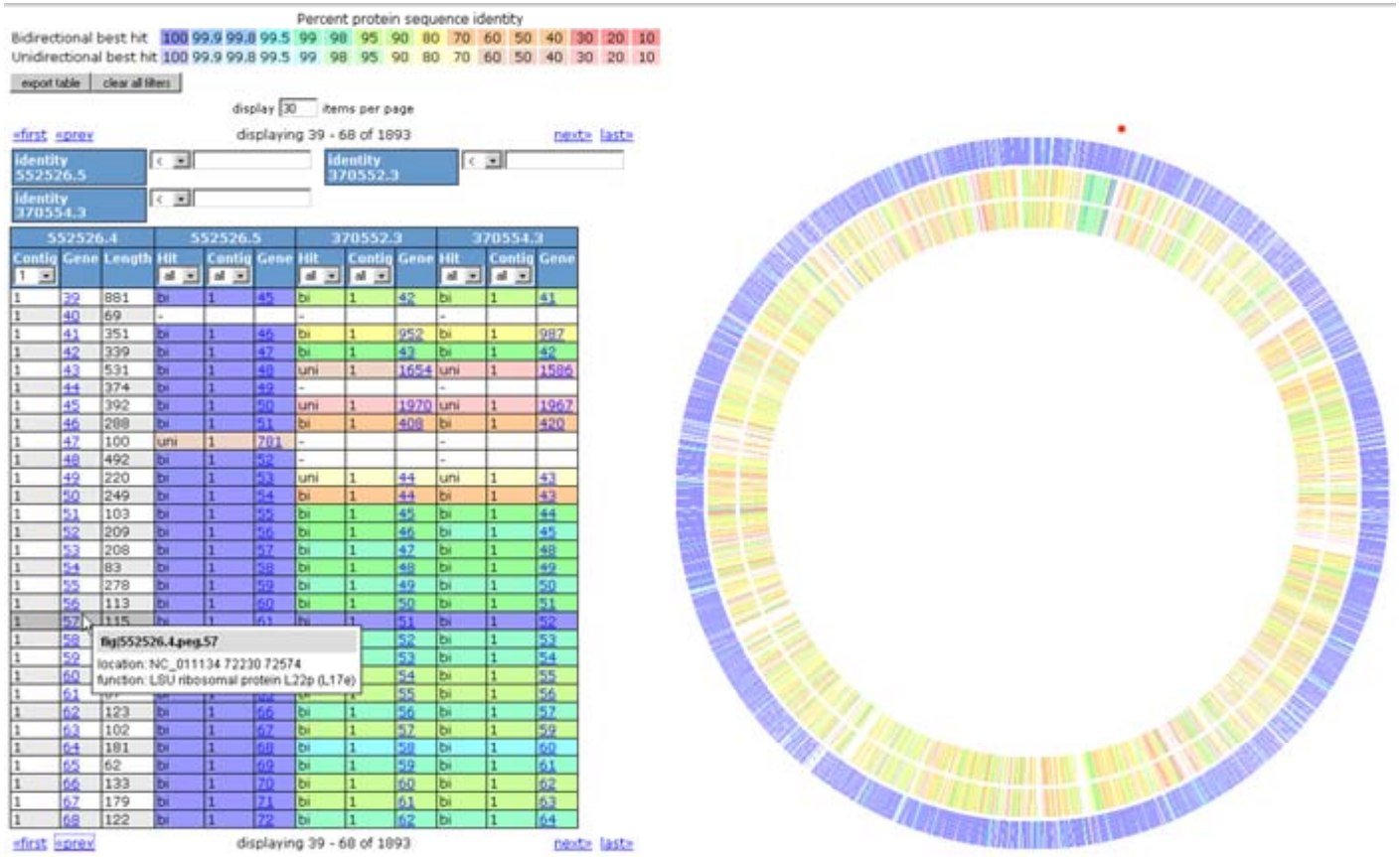


How to conduct comparative genome analysis: relevant bioinformatic tools

- Whole genome alignment (DNA-based)
- Ortholog clustering (protein-based)

Whole genome alignment

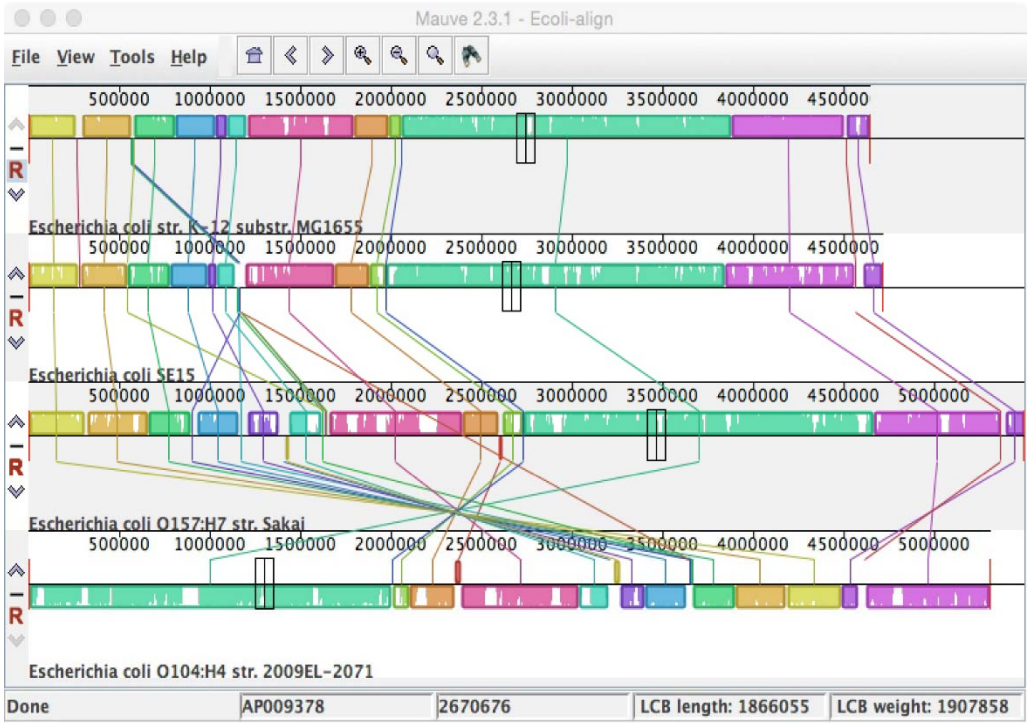
RAST & SEED viewer (http://rast.nmpdr.org)

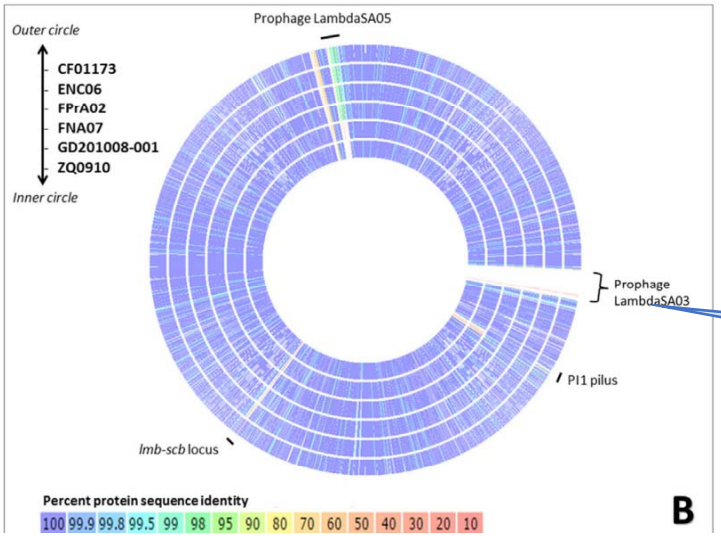


Whole genome alignment

MAUVE

- Input:
  - DNA fasta
  - GenBank file (.gbk)
- Visualizing genome rearrangement
- Java program
- http://darlinglab.org/mauve/mauve.html





Contents lists available at ScienceDirect

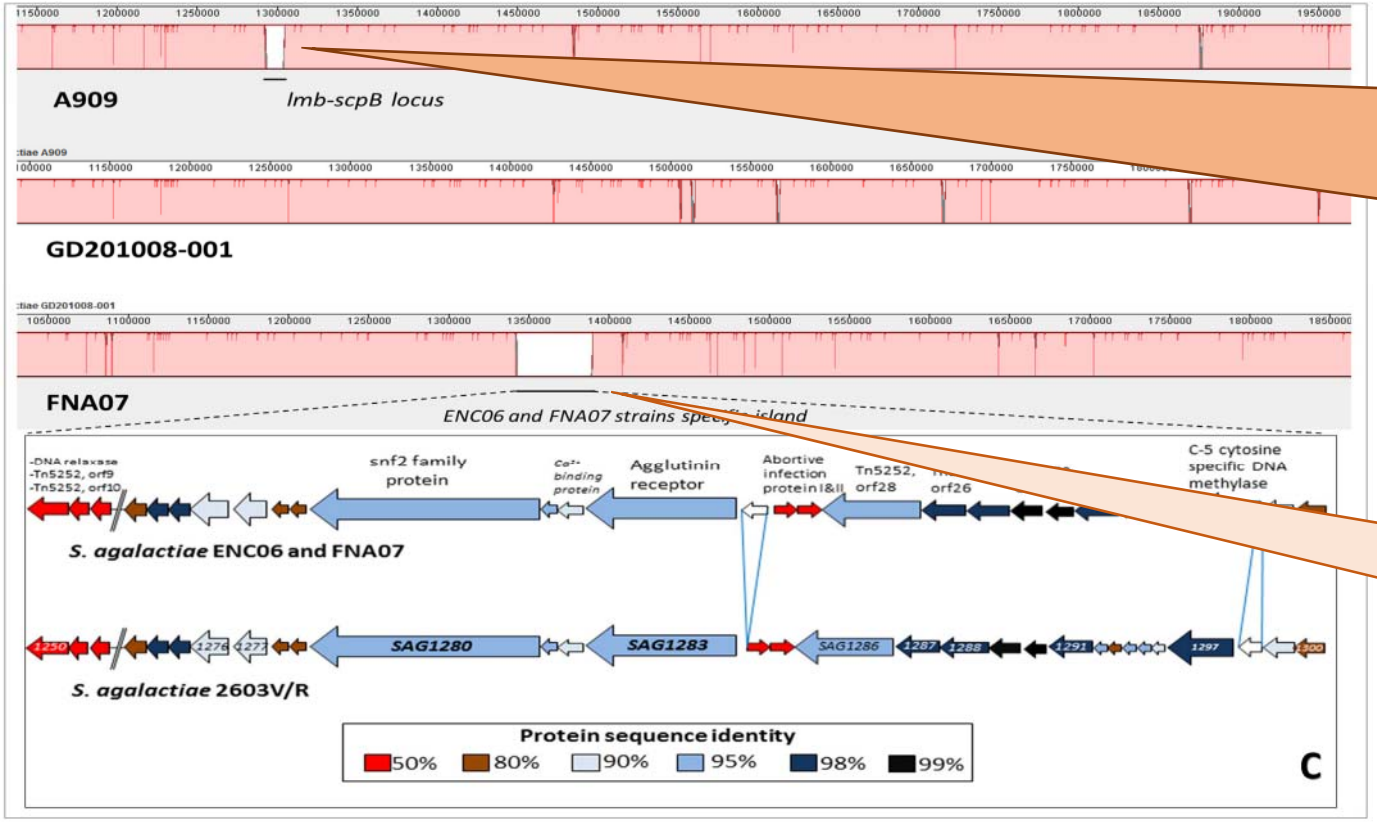
Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Genomic comparison between pathogenic *Streptococcus agalactiae* isolated from Nile tilapia in Thailand and fish-derived ST7 strains

Pattanapon Kayansamruaj [a], Nopadon Pirarat [b], Hidehiro Kondo [c], Ikuo Hirono [c], Channarong Rodkhum [a,*]

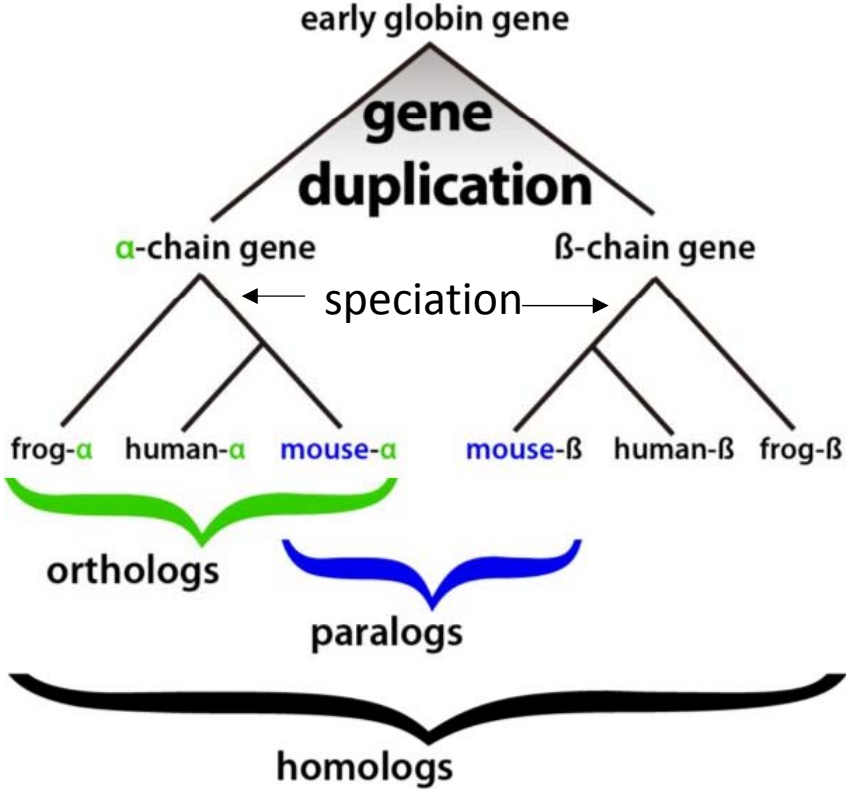
Prophage sequence is deleted in piscine *S. agalactiae* strains comparing to human strains



- *Laminin binding protein* and *C5a peptidase genes* were found in huma strains but not in fish strains
- Deletion involved in niche adaptation

Genomic island specifically present in *S. agalactiae* Thai isolate

Ortholog clustering



Group orthologous protein together

- All vs All protein comparison
- Information used: sequence similarity, synteny
- Difficulty increases with taxa distance
- ***DEPEND ON ANNOTATION QUALITY!!***

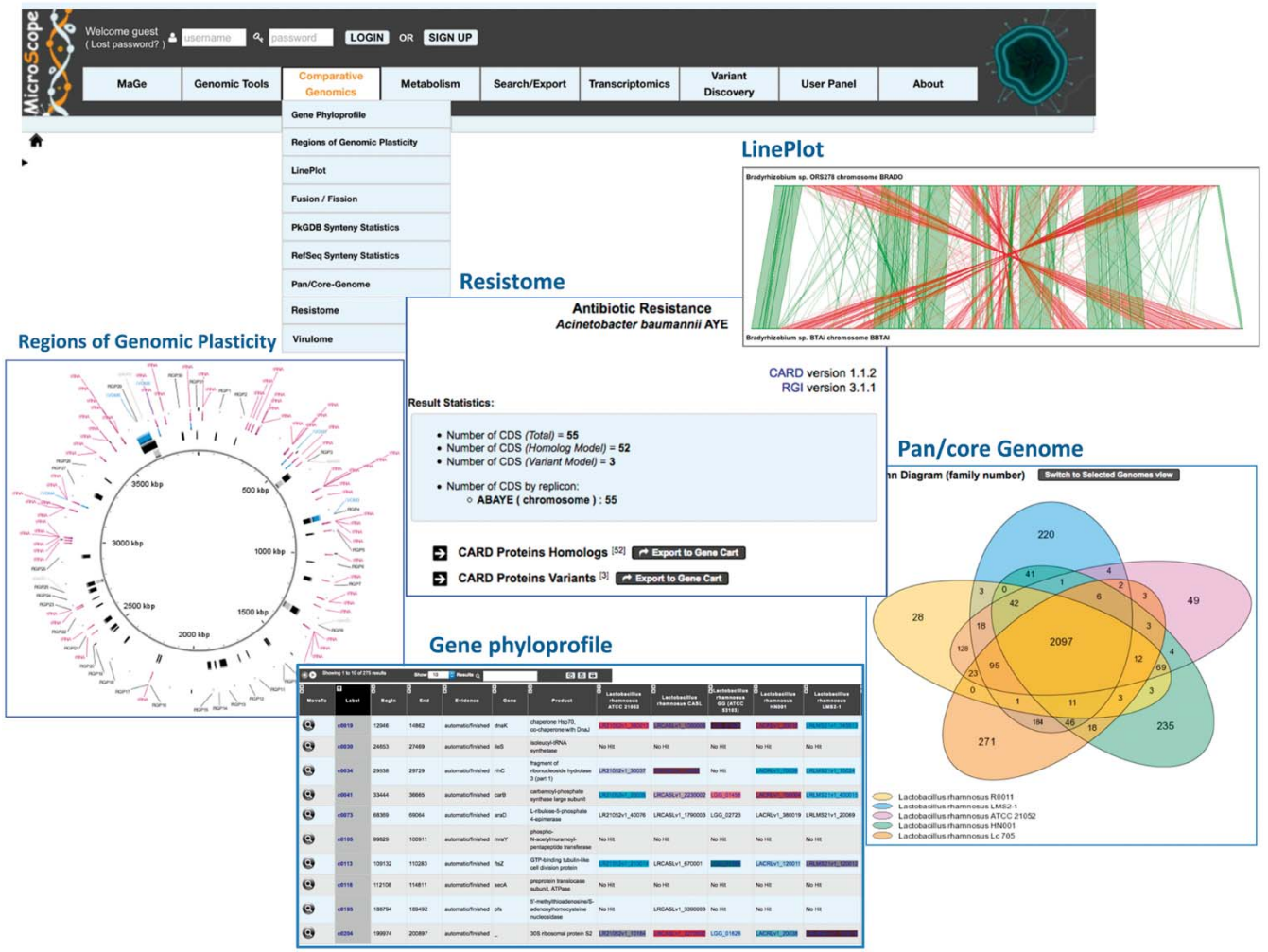
Software: OrthoMCL, cd-hit, uclust etc.

Ortholog clustering

(almost) all-in-one comparative genomics tool

MicroScope

http://www.genoscope.cns.fr/agc/microscope/home/index.php



Comparative Genomics and Transcriptional Analysis of *Flavobacterium columnare* Strain ATCC 49512

*Hasan C. Tekedar[1], Attila Karsi[1], Joseph S. Reddy[2], Seong W. Nho[1], Safak Kalindamar[1] and Mark L. Lawrence[1]**

frontiers in Microbiology

the strain ATCC 49512 genome to four other *Flavobacterium* genomes. In this analysis, we identified predicted proteins whose functions indicate *F. columnare* is capable of denitrification, which would enable anaerobic growth in aquatic pond sediments.

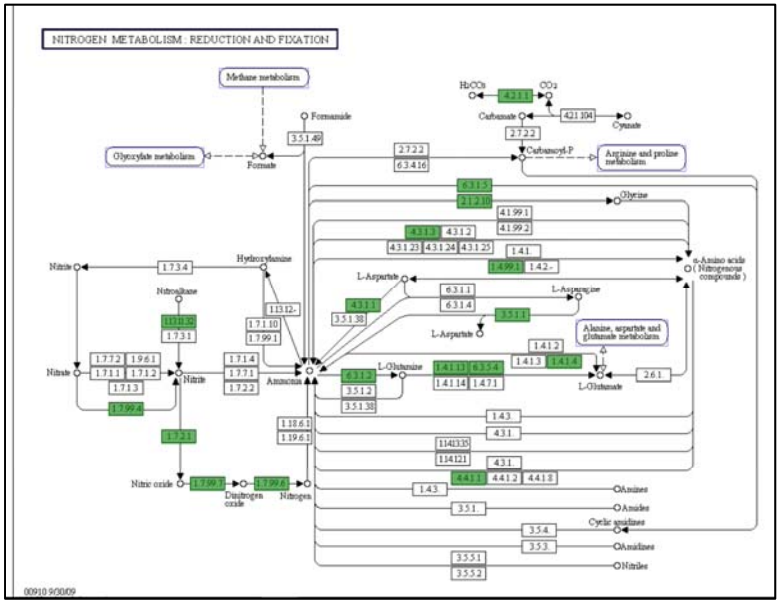
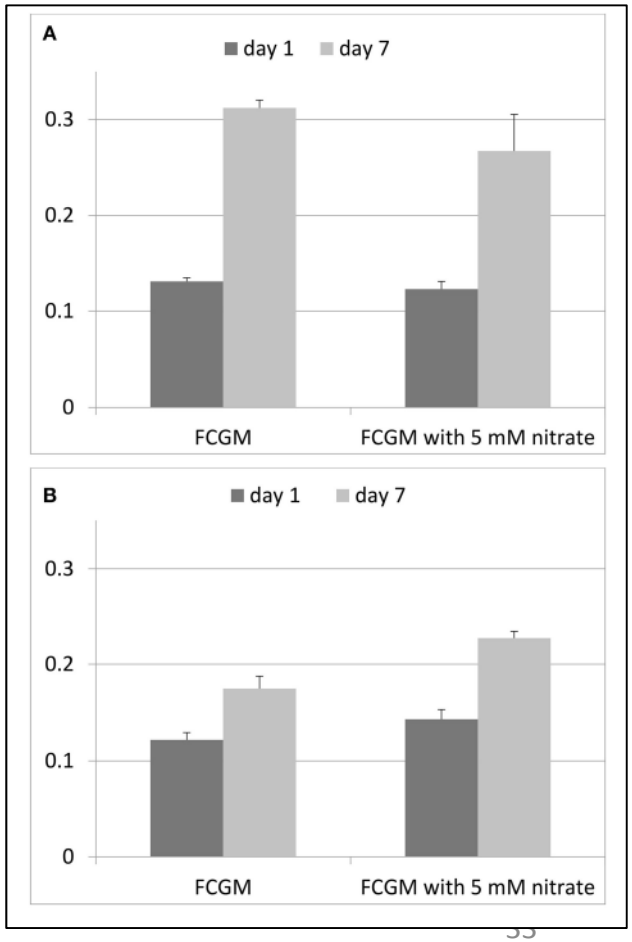


FIGURE 5 | Growth of *F. columnare* ATCC 49512 in FCGM and FCGM supplemented with 5 mM nitrate. (A) Growth under aerobic conditions and **(B)** growth under anaerobic conditions.



Taxonomic classification based on NGS data

Standard parameters for taxonomic revision in the genomic era

I. Average nucleotide identity (ANI)

- The ANI values between genomes of the same species are typically **above 95%**
- https://www.ezbiocloud.net/tools/ani
- http://ani.mypathogen.cn/

II. Digital DNA-DNA hybridization (dDDH)

- Result can be interpreted in the same scale as conventional DDH
- Cut-off: 70% (same species), 80% (same subspecies)
- http://ggdc.dsmz.de/ggdc.php

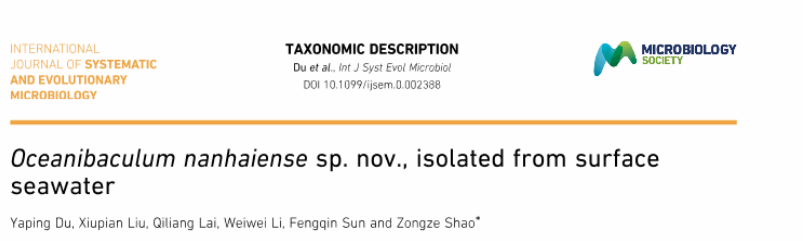
Inference of taxonomic position of unknown bacterium must be made by comparison with **TYPE STRAIN** only!!!

Taxonomic classification based on NGS data

Standard parameters for taxonomic revision in the genomic era

- Currently, **International Journal of Systematic and Evolutionary Microbiology** accepts the WGS as genotypic tool for species demarcation
- Can be used ***alternative*** to conventional ***DDH***

However, polyphasic approach (**phenotype**) is still **mandatory** to define nov. sp.



INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY

TAXONOMIC DESCRIPTION

Du *et al.*, *Int J Syst Evol Microbiol*

DOI 10.1099/ijsem.0.002388

MICROBIOLOGY SOCIETY

Oceanibaculum nanhaiense sp. nov., isolated from surface seawater

Yaping Du, Xiupian Liu, Qiliang Lai, Weiwei Li, Fengqin Sun and Zongze Shao*

International Journal of Systematic and Evolutionary Microbiology (2014), **64**, 316–324 DOI 10.1099/ijs.0.054171-0

Review

Integrating genomics into the taxonomy and systematics of the *Bacteria* and *Archaea*

Jongsik Chun[1] and Fred A. Rainey[2]

Research paper

Comparative genome analysis of fish pathogen *Flavobacterium columnare* reveals extensive sequence diversity within the species

Pattanapon Kayansamruaj [a,b,*], Ha Thanh Dong [c], Ikuo Hirono [d], Hidehiro Kondo [d], Saengchan Senapin [e], Channarong Rodkhum [a,**]

Genomovar	II				I		I/II	II			
Strains	1362	CF1	94-081	C#2	ATCC49512	Pf1	1215	1214	NK01	*F. johnsoniae* UW101	*F. psychrophilum* JIP02/86
1362											
CF1	**92.9** 0.15										
94-081	**97.6** 0.19	**91.3** 0.04									
C#2	**93.2** 0.36	**92.1** 0.21	**93.1** 0.17								
ATCC49512	**42.6** 0.85	**43.5** 0.71	**42.6** 0.66	**43** 0.5							
Pf1	**42.8** 0.96	**43.4** 0.82	**42.7** 0.77	**43.3** 0.61	**95.9** 0.11						
1215	**30.1** 0.02	**30.4** 0.12	**30** 0.17	**30.2** 0.33	**29.8** 0.83	**29.8** 0.94					
1214	**27.8** 0.62	**28.1** 0.76	**28.3** 0.81	**27.9** 0.98	**28** 1.47	**28.2** 1.58	**28** 0.64				
NK01	**27.7** 0.74	**28.1** 0.89	**28.3** 0.93	**27.9** 1.1	**28** 1.6	**28.1** 1.71	**28.1** 0.77	**94.5** 0.12			
F. johnsoniae UW101	**22.6** 3.5	**21.9** 3.36	**22.7** 3.31	**24** 3.14	**22.6** 2.65	**22.9** 2.54	**20.7** 3.48	**21.5** 4.12	**20.9** 4.24		
F. psychrophilum JIP02/86	**20.4** 1.92	**20.7** 1.78	**21.1** 1.73	**22.2** 1.56	**21.5** 1.07	**21.7** 0.96	**20.1** 1.9	**20** 2.54	**20** 2.66	**22** 1.58	

%DDH
% G+C content difference

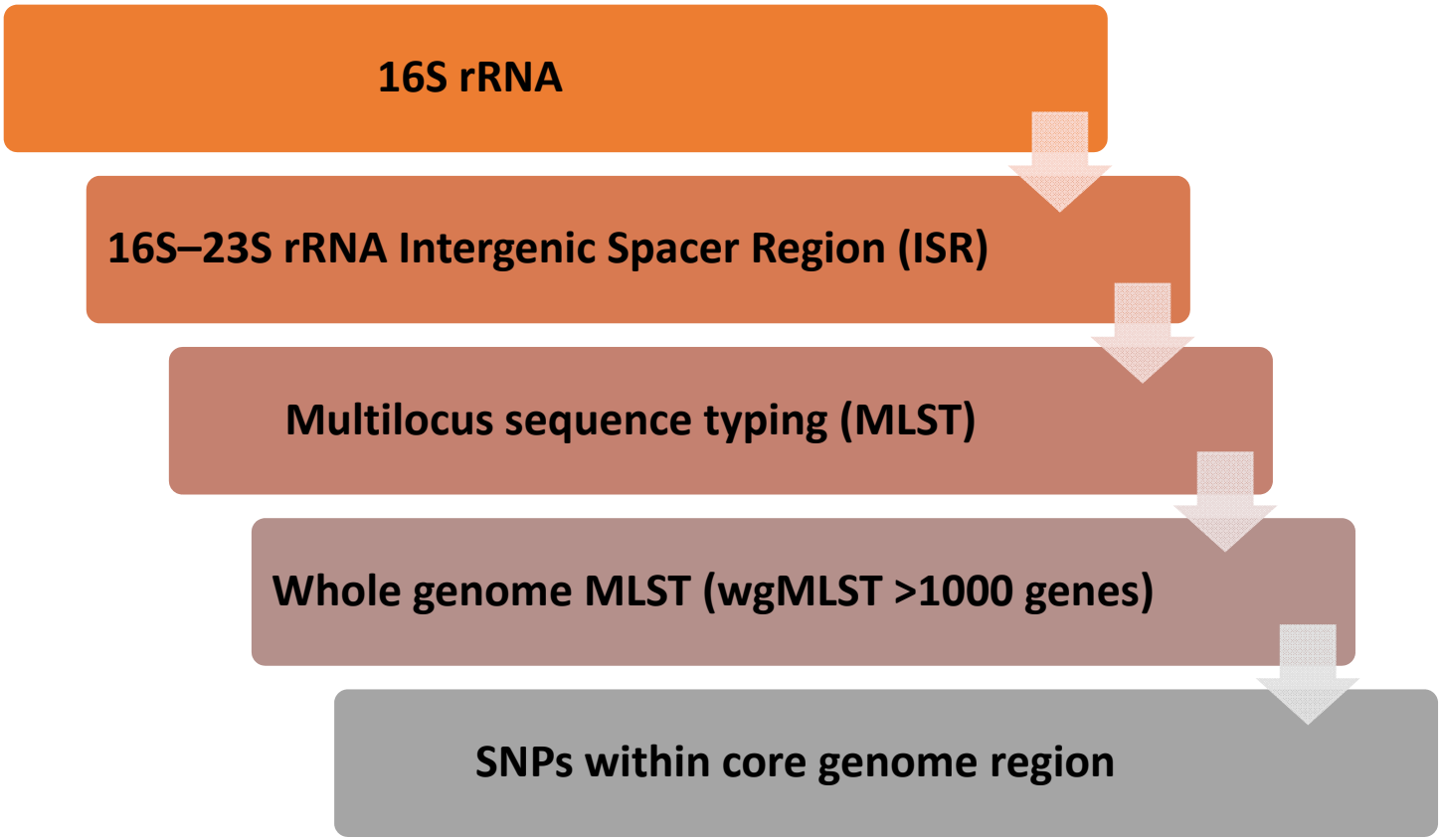
Phylogenomics

What is “phylogenomics”

- Evolution + Genomics
- Phylogenomics draws information by comparing entire genomes, or at least large portions of genomes, whereas phylogenetic compares only a small number of genes
- Can infer evolutionary relationships, gene family evolution (duplication, deletion, recombination), lateral gene transfer, taxonomy etc.

Genetic markers for reconstructing evolutionary history

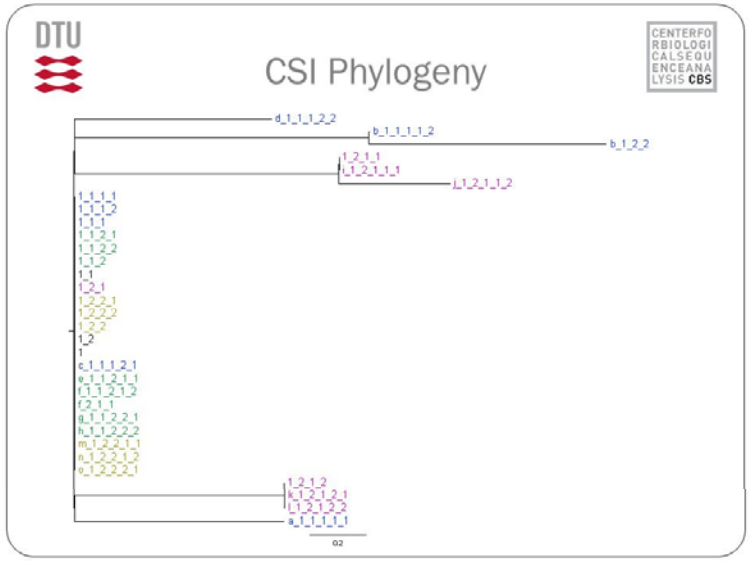
Phylogeny reconstruction based on sequence alignment



Genetic markers for reconstructing evolutionary history

Extracting concatenated SNP sequence from core genome region

- CSI phylogeny
- https://cge.cbs.dtu.dk/services/CSIPhylogeny/
- Automated web-based tool
- Obtained multiple alignment file to construct phylogenetic tree using secondary program

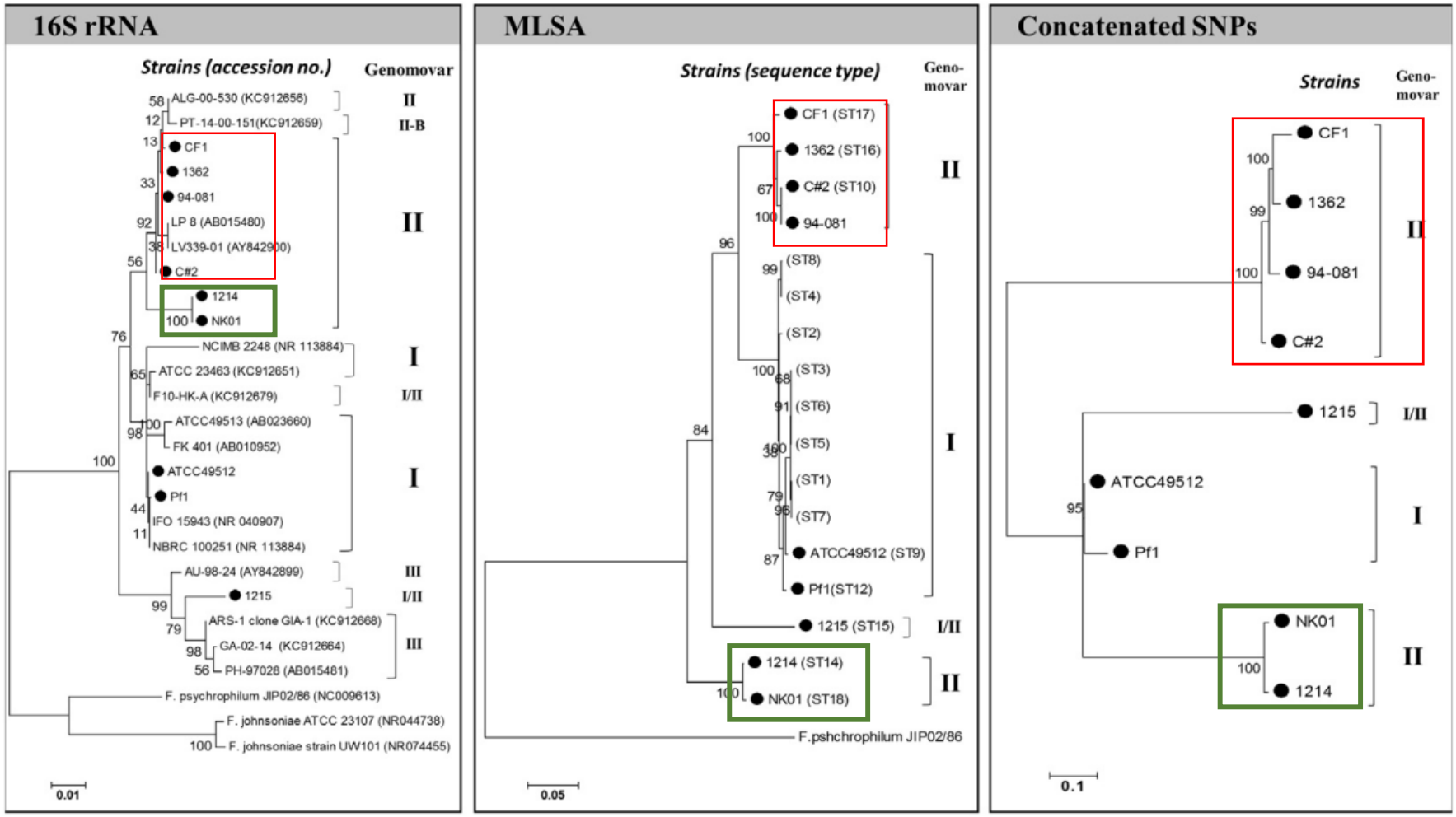


Research paper

Comparative genome analysis of fish pathogen *Flavobacterium columnare* reveals extensive sequence diversity within the species

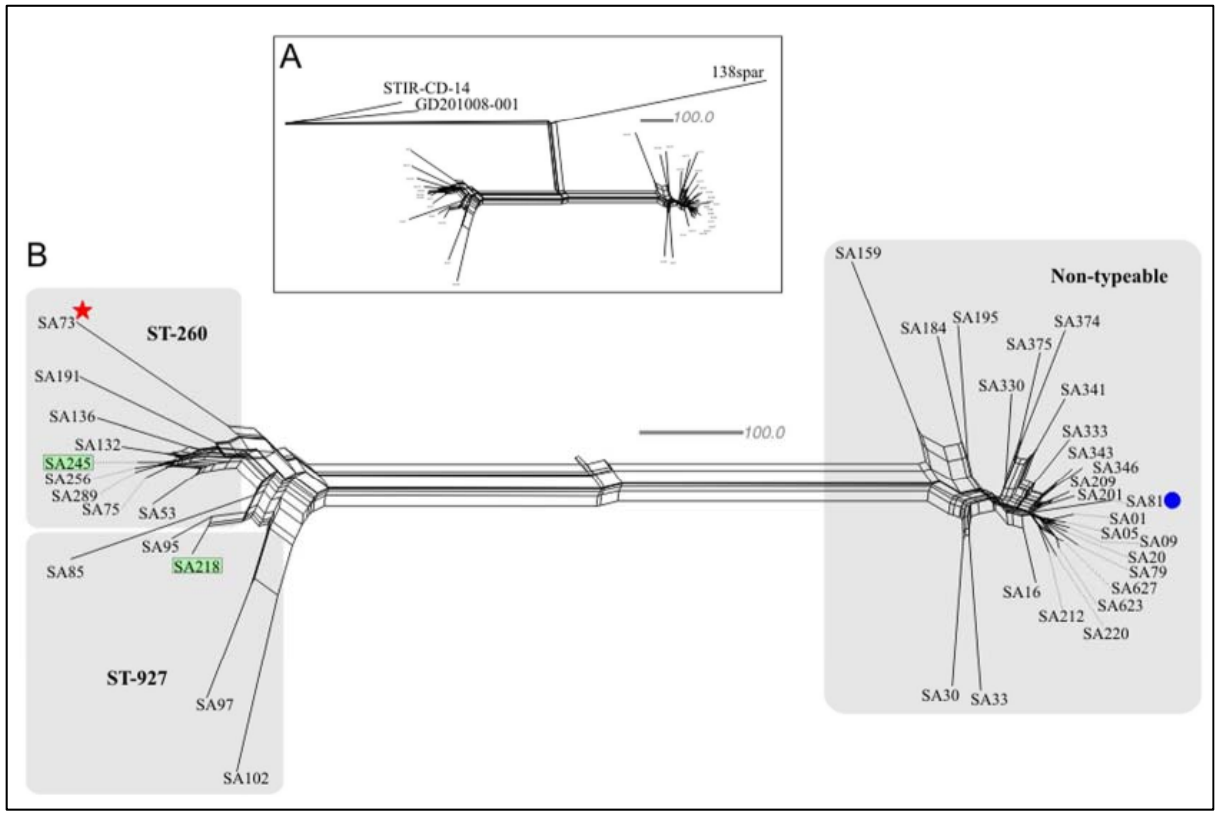
Pattanapon Kayansamruaj [a,b,*], Ha Thanh Dong [c], Ikuo Hirono [d], Hidehiro Kondo [d], Saengchan Senapin [e], Channarong Rodkhum [a,**]

Phylogenetic relationship of *F. columnare* strains generated from different genetic markers



Genetic markers for reconstructing evolutionary history

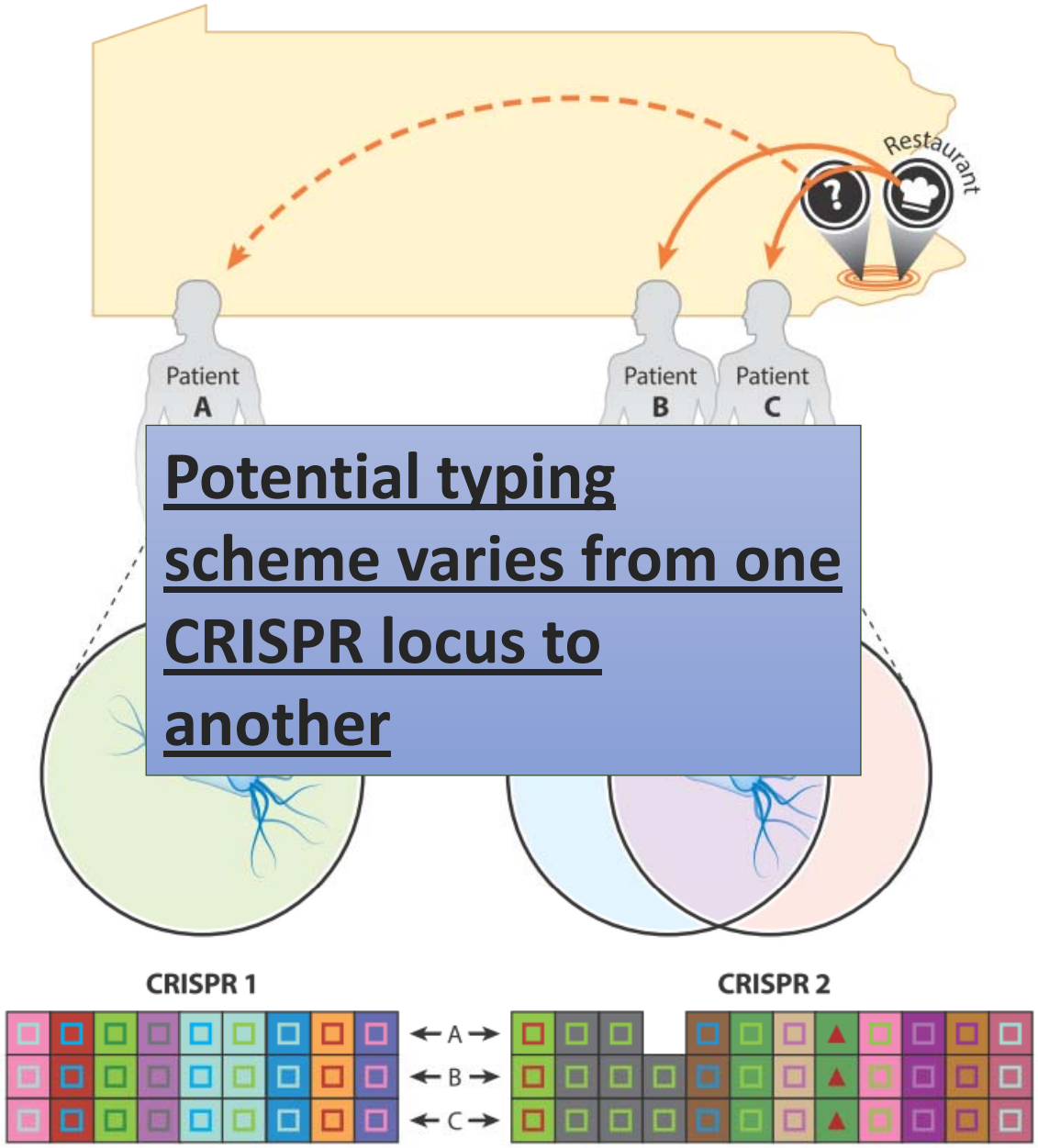
- **Revisited *PubMLST***
  - Phylogenetic network generated from wgMLST data



Phylogenomic NeighborNet network of whole genome MLST (wgMLST) data inferred that at least 2 population of fish pathogenic *S. agalactiae* were present in Brazil

Large-scale genomic analyses reveal the population structure and evolutionary trends of *Streptococcus agalactiae* strains in Brazilian fish farms

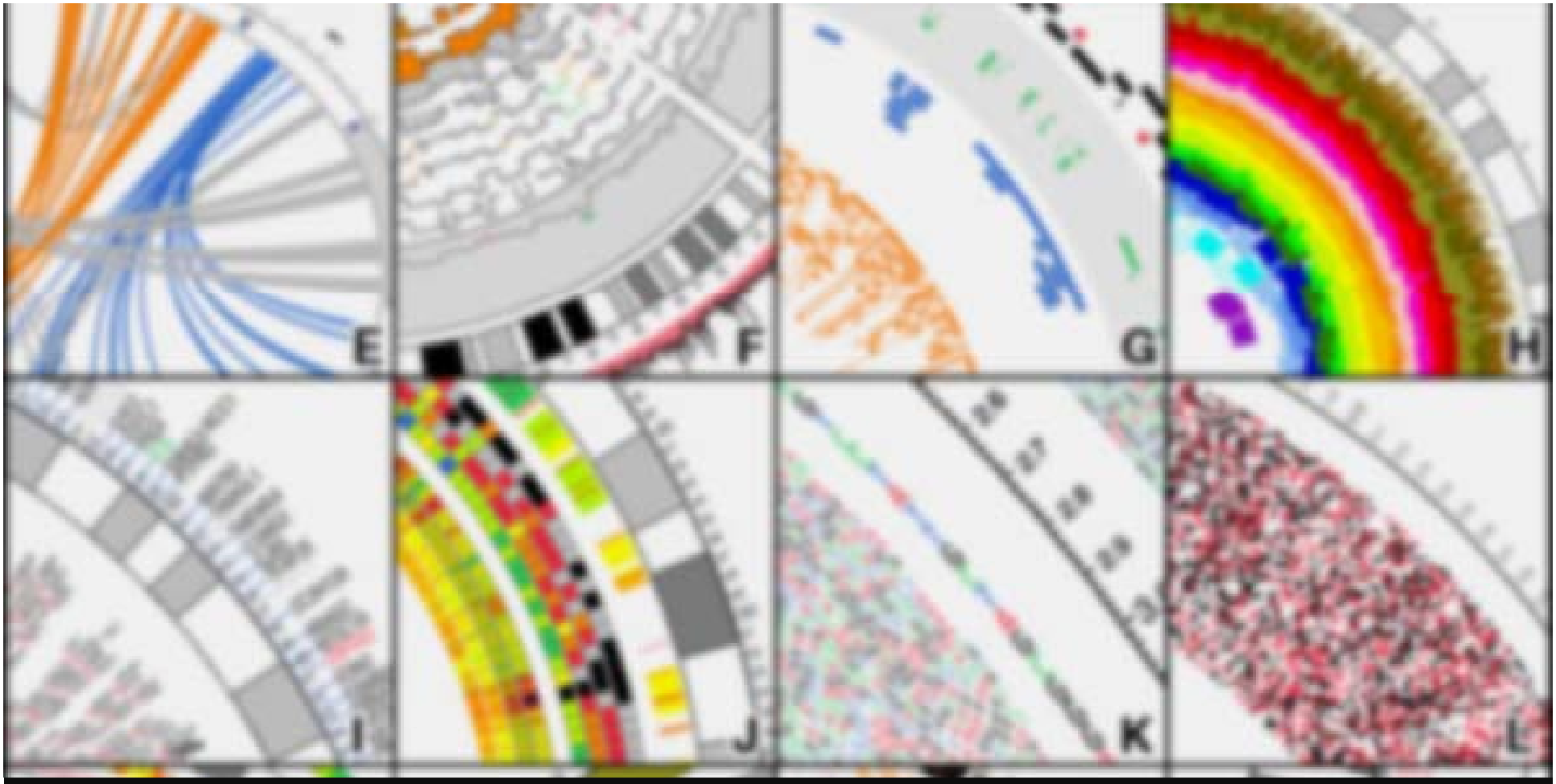
Gustavo M. Barony, Guilherme C. Tavares, Felipe L. Pereira, Alex F. Carvalho, Fernanda A. Dorella, Carlos A. G. Leal & Henrique C. P. Figueiredo



CRISPR-Based Typing and Next-Generation Tracking Technologies

Rodolphe Barrangou[1,2,*] and Edward G. Dudley[2]

- Three *Salmonella* Enteritidis isolates were
- received during the **same time frame**
- All three are **same pulsotype** (same PFGE pattern)
- Pattern B and C ate at the same restaurant
- No epidemiologic data linked to patient A
- CRISPR sequencing revealed a **one-spacer difference**
- **Two separate incidents** were responsible for these illnesses



Getting published

List of published genomic studies featured fish pathogenic flavobacteria (2016-2017)

- Tekedar HC, Karsi A, Reddy JS, Nho SW, Kalindamar S, Lawrence ML. **Comparative Genomics and Transcriptional Analysis of Flavobacterium columnare Strain ATCC 49512**. Frontiers in Microbiology. 2017;8:588. doi:10.3389/fmicb.2017.00588.

- Kumru S, Tekedar HC, Gulsoy N, Waldbieser GC, Lawrence ML and Karsi A (2017) **Comparative Analysis of the Flavobacterium** **75.**

- **ed** **gy** 8.

- Castillo D, Christiansen RH, Dalsgaard I, Madsen L, Espejo R & Middelboe M (2016) **Comparative genome analysis provides insights into the pathogenicity of Flavobacterium psychrophilum**. PLOS ONE 11: e0152515.

- Kayansamruaj P, Dong HT, Hirono I, Kondo H, Senapin S & Rodkhum C (2017) **Comparative genome analysis of fish pathogen Flavobacterium columnare reveals extensive sequence diversity within the species**. Infection, Genetics and Evolution 54: 7-17.

All these are **descriptive studies** and **not hypothesis testing** type of research

Is it possible to publish paper based on WGS data only?

If you have **only genome sequences** without further analysis

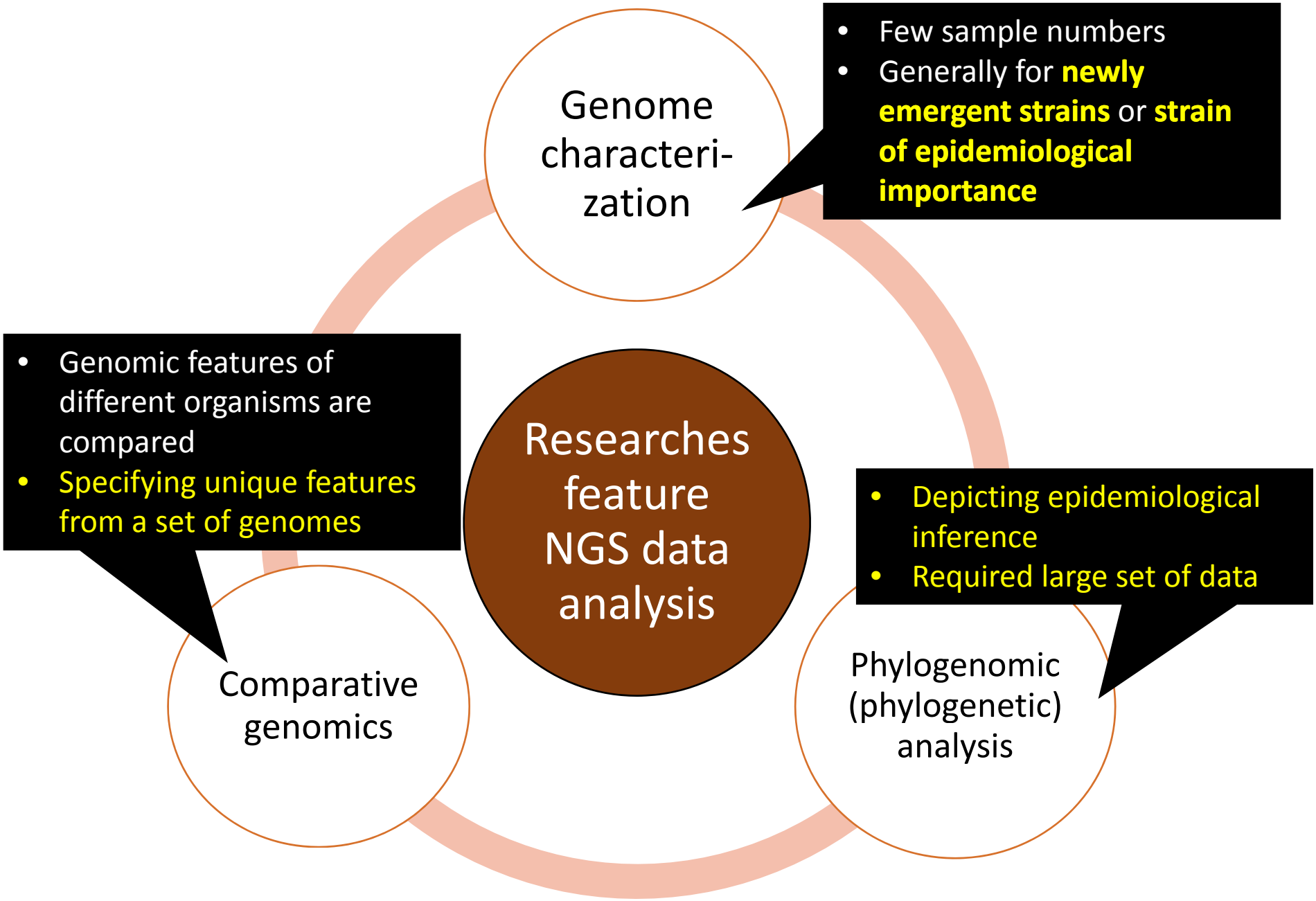
- Submit data to *NCBI*

If you provide **genome sequences** with **basic annotation**

- *Genome Announcements*
- *Standard in Genomic Sciences*

If you provide **genome sequences** with **specific analyses.** Study must be **constructive** and **informative**

- *Frontiers, PLOS ONE, BMC, Genome Research, MEEGID etc.*


Genome characteri-zation
Few sample numbers
Generally for newly emergent strains or strain of epidemiological importance
Researches feature NGS data analysis
Genomic features of different organisms are compared
Specifying unique features from a set of genomes
Comparative genomics
Depicting epidemiological inference
Required large set of data
Phylogenomic (phylogenetic) analysis

Segura *et al. Standards in Genomic Sciences* (2017) 12:61
DOI 10.1186/s40793-017-0272-0

SIGS
Standards in Genomic Sciences
An Open Access Journal of the Genomic Standards Consortium

EXTENDED GENOME REPORT Open Access

Draft genome sequence and characterization of commensal *Escherichia coli* strain BG1 isolated from bovine gastro-intestinal tract

Audrey Segura[1*†], Pauline Auffret[1†], Christophe Klopp[2], Yolande Bertin[1] and Evelyne Forano[1]

Table 6 Adherence systems encoded by the *E. coli* BG1 genome

Adherence system	Gene or genes cluster	Pathotype[a]	in vitro cell adherence[b]	Receptor
Curli fimbriae	*csgDEFG, csgBA*	EHEC, ETEC, aEPEC, APEC	T84	Matrix, plasma proteins
EhaA autotransporter	*ehaA*	EHEC, EAEC, ETEC, AIEC, EPEC	Primary bovine epithelial cells (terminal rectum)	Unknown
EhaB autotransporter	*ehaB*	EHEC, UPEC, ETEC, EIEC, EAEC	NA[c]	Collagen I, laminin
EhaC autotransporter	*ehaC (yfaL)*	EHEC, UPEC	Unknown	Unknown
ECP (*E. coli* Common Pilus)	*ecpRABCDE*	ETEC, EHEC, NMEC, EAEC, aEPEC, septicemia	HT29, Hep-2, HeLa, HTB-4	Arabinosyl residues
ELF (*E. coli* Laminin-binding Fimbriae)	*ycbQRST*	EHEC, aEPEC	HT29, Hep-2, MDBK	Laminin
F9 Fimbriae	*z2200-z2206*	EHEC, UPEC, APEC, AIEC, EAEC, EPEC	EBL	Bovine fibronectin, Galβ1-3GlcNAc

Standards in Genomic Sciences (2013) **8**:188-197 DOI:10.4056/sigs.3687314

Table 4 Number of genes associated with general COGs functional categories

Code	Value	% age[a]	Description
J	250	6.55	Translation, ribosomal structure and biogenesis
A	2	0.05	RNA processing and modification
K	293	7.68	Transcription
L	154	4.04	Replication, recombination and repair
B	0	0.00	Chromatin structure and dynamics
D	41	1.07	Cell cycle control, cell division, chromosome partitioning
V	93	2.44	Defense mechanisms
T	176	4.61	Signal transduction mechanisms
M	271	7.11	Cell wall/membrane/envelope biogenesis
N	156	4.09	Cell motility
U	60	1.57	Intracellular trafficking, secretion, and vesicular transport
O	153	4.01	Post-translational modifications, protein turnover, chaperones
C	282	7.39	Energy production and conversion
G	381	9.99	Carbohydrate transport and metabolism
E	335	8.78	Amino acid transport and metabolism
F	101	2.65	Nucleotide transport and metabolism
H	169	4.43	Coenzyme transport and metabolism
I	119	3.12	Lipid transport and metabolism
P	190	4.98	Inorganic ion transport and metabolism
	53	1.39	Secondary metabolites biosynthesis, transport and catabolism
	211	5.53	General function prediction only
	238	6.24	Function unknown
	750	7.43	Not in COGs

Complete genome sequence of *Streptococcus agalactiae* strain SA20-06, a fish pathogen associated to meningoencephalitis outbreaks

BioMed Central
The Open Access Publisher

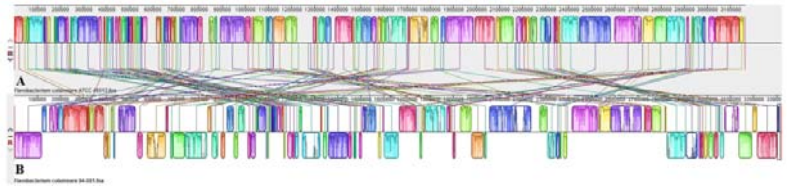
frontiers in Microbiology

PNAS

MEEGID Infection, Genetics and Evolution

ORIGINAL RESEARCH
published: 25 July 2017
doi: 10.3389/fmicb.2017.01375

SCIENTIFIC REPORTS

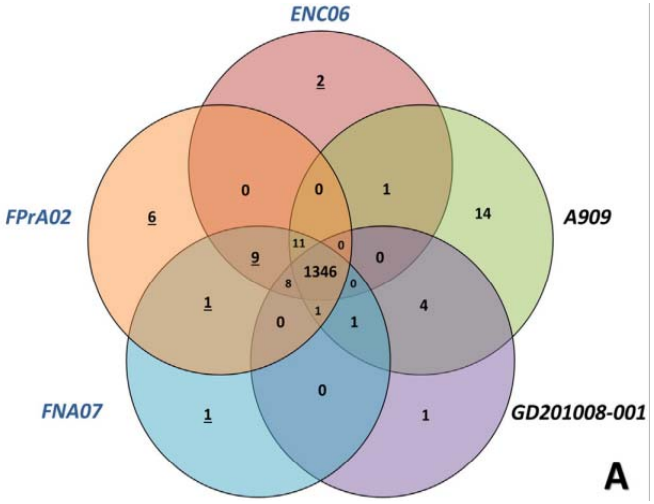


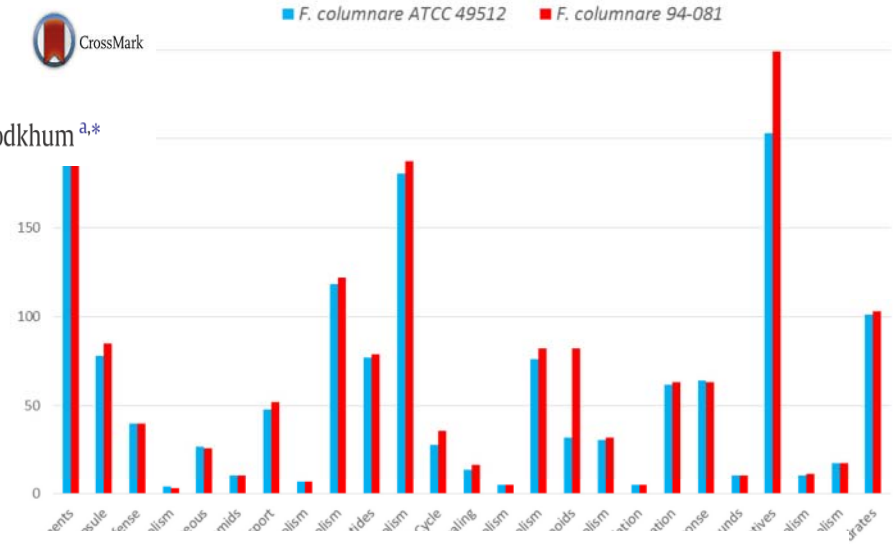
Comparative Analysis of the *Flavobacterium columnare* Genomovar I and II Genomes

Salih Kumru[1†], Hasan C. Tekedar[1†], Nagihan Gulsoy[2], Geoffrey C. Waldbieser[3], Mark L. Lawrence[1] and Attila Karsi[1*]*

Genomic comparison between pathogenic *Streptococcus agalactiae* isolated from Nile tilapia in Thailand and fish-derived ST7 strains

Pattanapon Kayansamruaj [a], Nopadon Pirarat [b], Hidehiro Kondo [c], Ikuo Hirono [c], Channarong Rodkhum [a,*]





Journal of Fish Diseases 2014
doi:10.1111/jfd.12319

Genomic comparison of virulent and non-virulent *Streptococcus agalactiae* in fish

C M J Delannoy[1,2], R N Zadoks[2,3], M Crumlish[1], D Rodgers[2], F A Lainson[2], H W Ferguson[4], J Turnbull[1] and M C Fontaine[2]

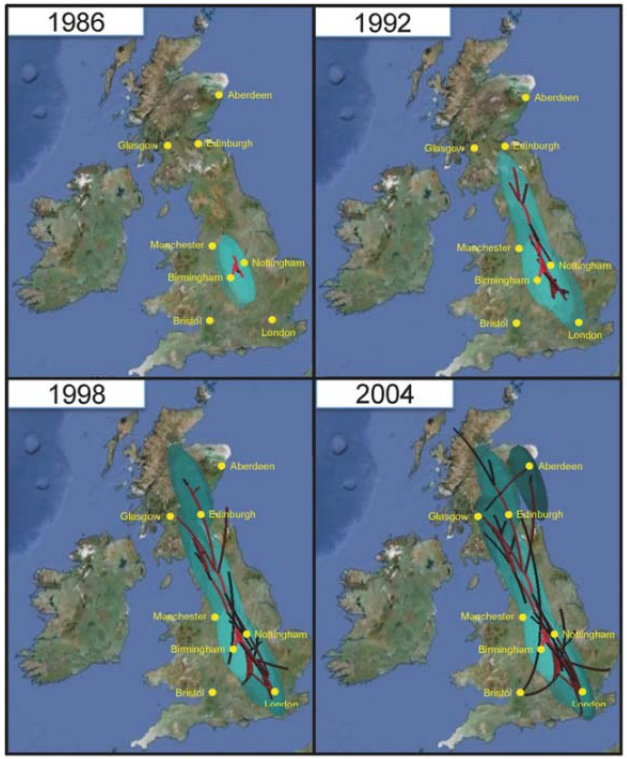
Research

Genome Research
www.genome.org

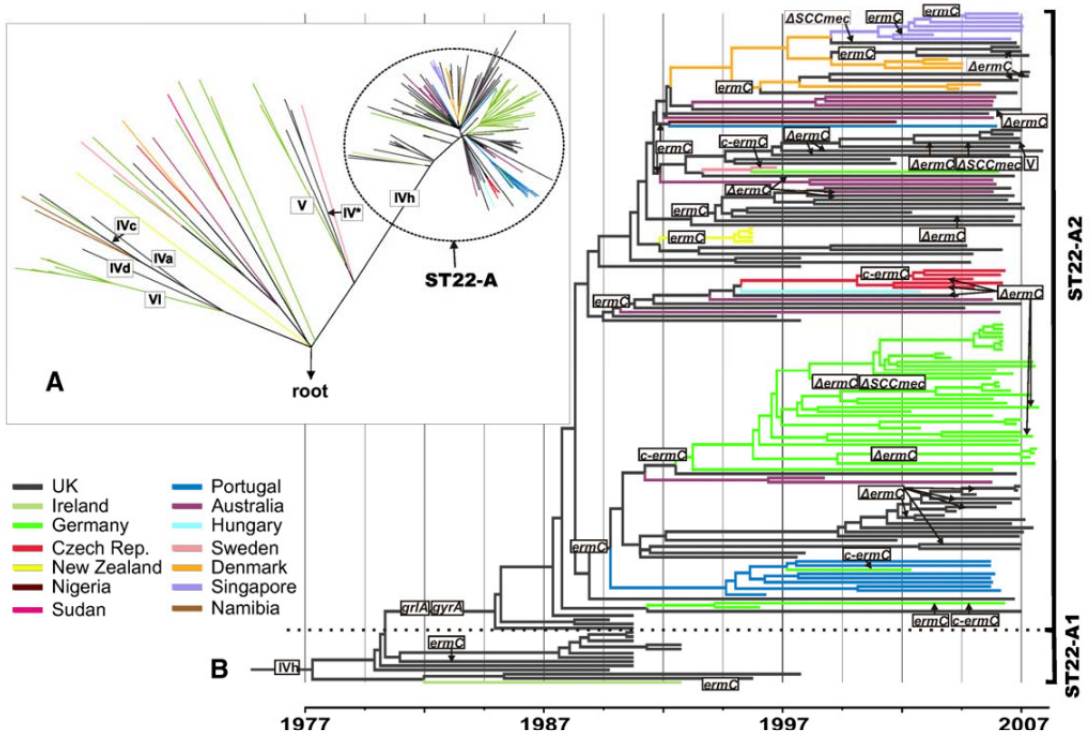
A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic

Matthew T.G. Holden,[1,25] Li-Yang Hsu,[1,2] Kevin Kurt,[3] Lucy A. Weinert,[4,22]

122 isolates of S. aureus ST22 were sequenced



SPREAD: Spatial Phylogenetic Reconstruction of Evolutionary Dynamics

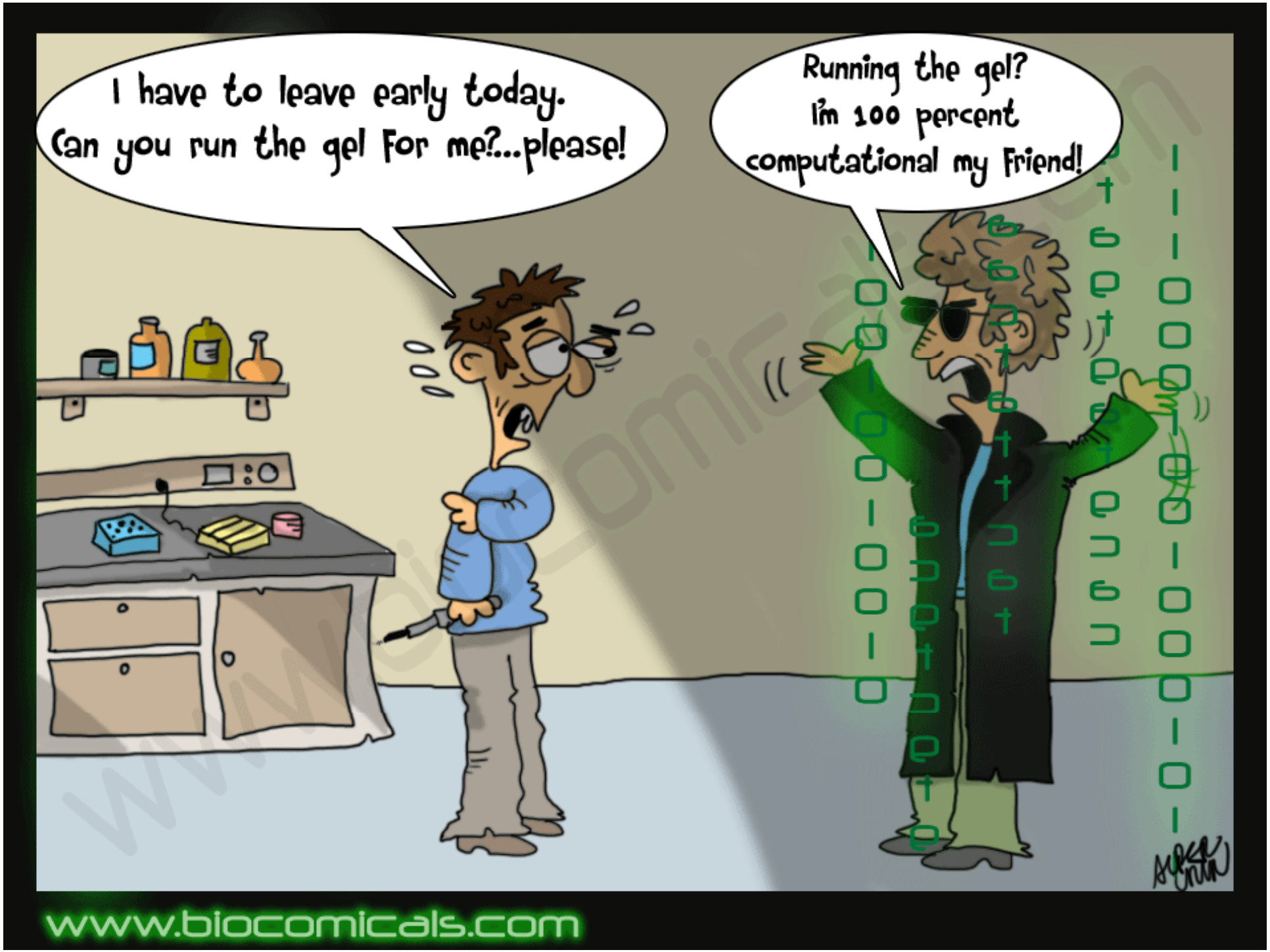


BEAST analysis based on core genome sequence

Today's topics

- Impact of Next-Generation sequencing to science
- Generic workflow in bacterial genome analysis
- High-throughput screening of bacterial genomes
- Comparative genomics and Phylogenomics
- Getting published

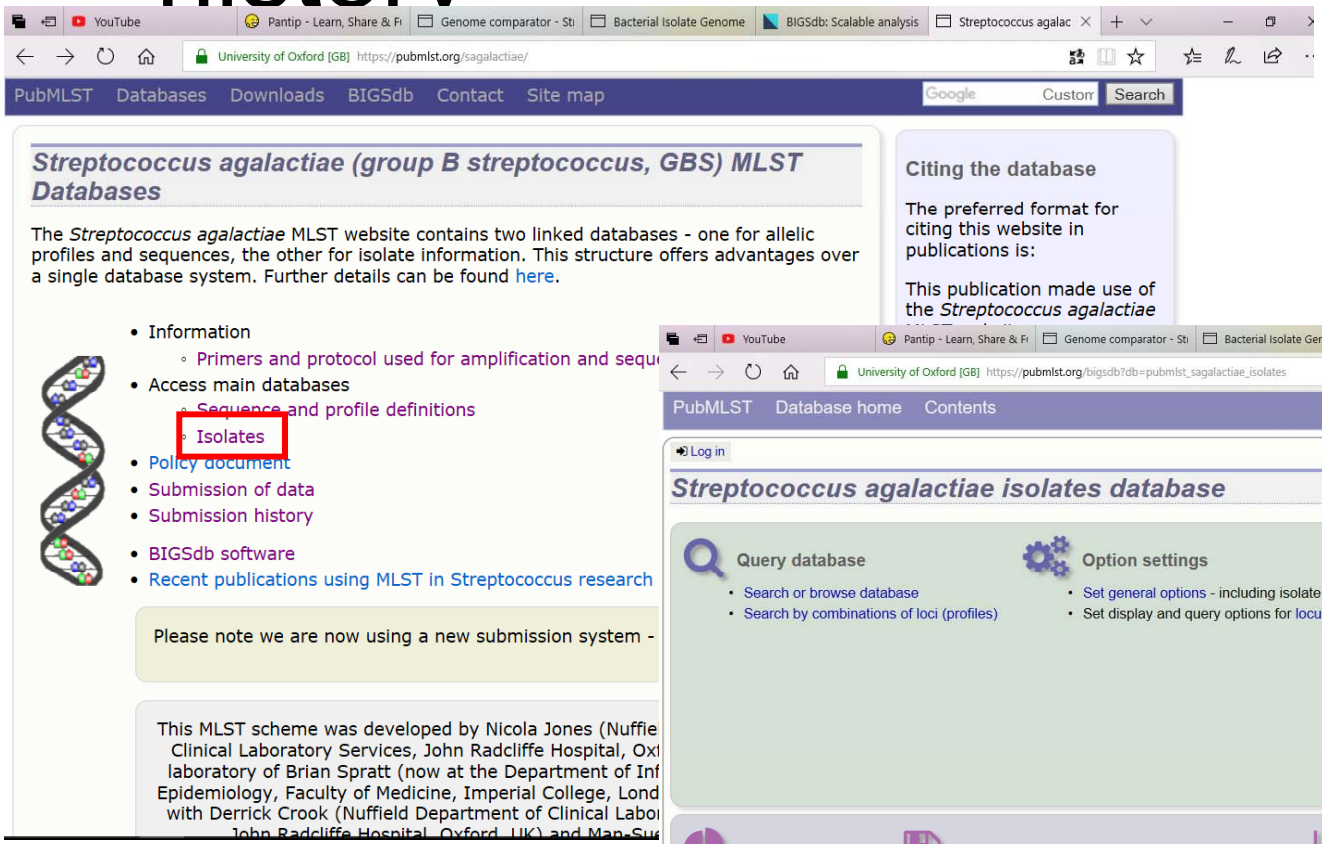
THANK YOU FOR YOUR ATTENTION! ANY QUESTIONS?

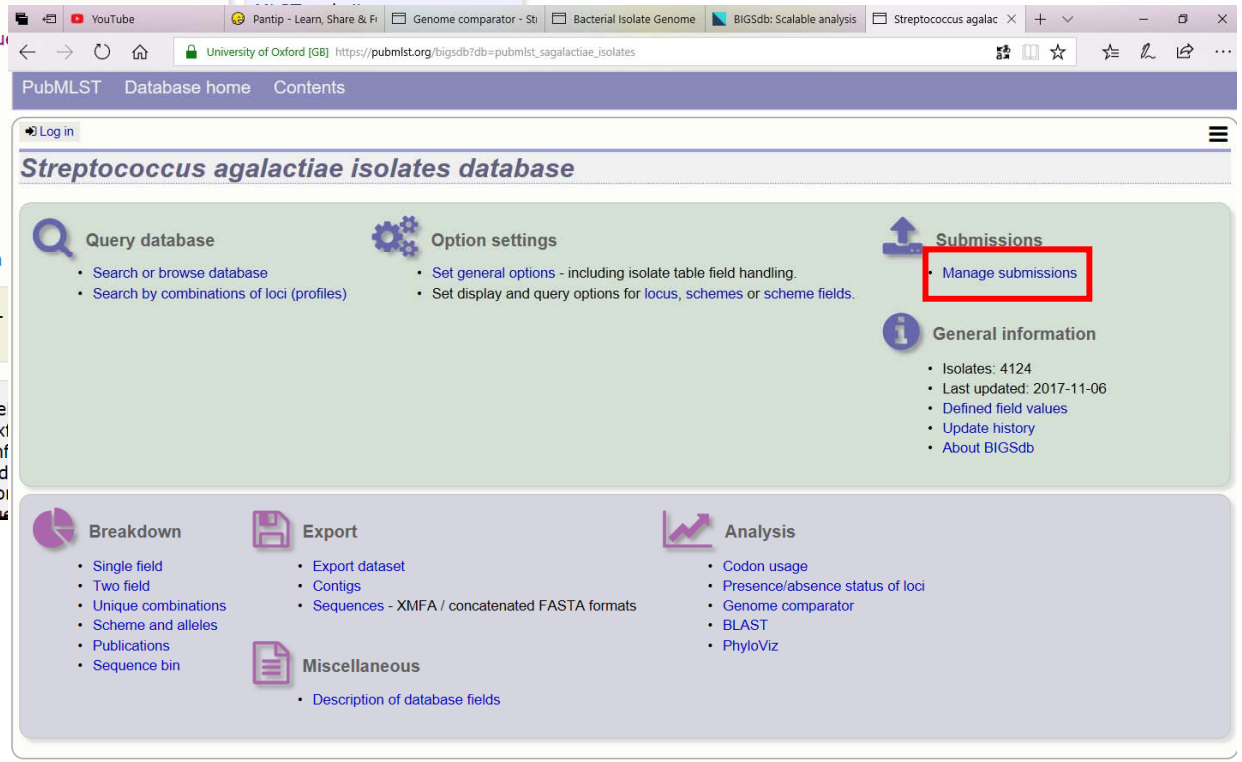

I have to leave early today. Can you run the gel for me?...please!
Running the gel? I'm 100 percent computational my Friend!
www.biocomicals.com

Genetic markers for reconstructing evolutionary history

- **Revisited *PubMLST***
  - Comparison of bacterial genotypes based on several typing schemes:
    - *MLST, capsule genes, antibiotic resistance genes, antigen typing, metabolic pathways typing, whole genome MLST (wgMLST) and much more…*
  - Locus allocation of bacterial isolates can be accomplished automatically
    - Submit your genome assembly to **BIGSdb**
  - https://pubmlst.org/

Genetic markers for reconstructing evolutionary history





Genome submission to BIGSdb

https://pubmlst.org/

Submission template.xls

Genetic markers for reconstructing evolutionary history

Output

Analysis against defined loci

All loci

Allele numbers are used where these have been defined, otherwise sequences will be marked as 'New#1, 'New#2' etc. Missing alleles are marked as 'X'. Incomplete alleles (located at end of contig) are marked as 'I'.

Locus	4469 (ERR1672933)	4470 (ERR1672935)	4645 (SBVN)	4646 (3896VN)
BACT000001 (rpsA)	3211	102	new#1	102
BACT000002 (rpsB)	816	111	112	112
BACT000003 (rpsC)	754	66	69	69
BACT000004 (rpsD)	76	76	77	77
BACT000005 (rpsE)	56	55	56	56
BACT000006 (rpsF)	22	63	22	22
BACT000007 (rpsG)	1951	59	59	59
BACT000008 (rpsH)	49	49	49	49
BACT000009 (rpsI)	4093	676	67	67
BACT000010 (rpsJ)	61	62	63	63
BACT000011 (rpsK)	50	50	50	50
BACT000012 (rpsL)	61	61	61	61
BACT000013 (rpsM)	47	47	47	47
BACT000014 (rpsN)	45;217	45;217	45;217	45;217
BACT000015 (rpsO)	46	46	47	47
BACT000016 (rpsP)	43	43	43	43
BACT000017 (rpsQ)	47	47	47	47
BACT000018 (rpsR)	42	42	42	42
BACT000019 (rpsS)	37	37	37	37
BACT000020 (rpsT)	59	59	59	59
BACT000021 (rpsU)	34	34	34	34
BACT000030 (rplA)	85	84	85	85
BACT000031 (rplB)	846	74	75	75
BACT000032 (rplC)	75	4463	75	75
BACT000033 (rplD)	57	56	57	57
BACT000034 (rplE)	61	61	61	61
BACT000035 (rplF)	2066	68	69	69
BACT000036 (rplL)	new#1	58	58	58
BACT000038 (rplI)	141	141	142	142
BACT000039 (rplJ)	810	59	59	59
BACT000040 (rplK)	73	74	4327	4327
BACT000042 (rplM)	792	792	53	53
BACT000043 (rplN)	50	50	50	50
BACT000044 (rplO)	59	786	59	59
BACT000045 (rplP)	1747	49	49	49
BACT000046 (rplQ)	42	42	42	42
BACT000047 (rplR)	54	54	54	54
BACT000048 (rplS)	814	814	90	90
BACT000049 (rplT)	68	69	68	68
BACT000050 (rplU)	new#1	54	54	54

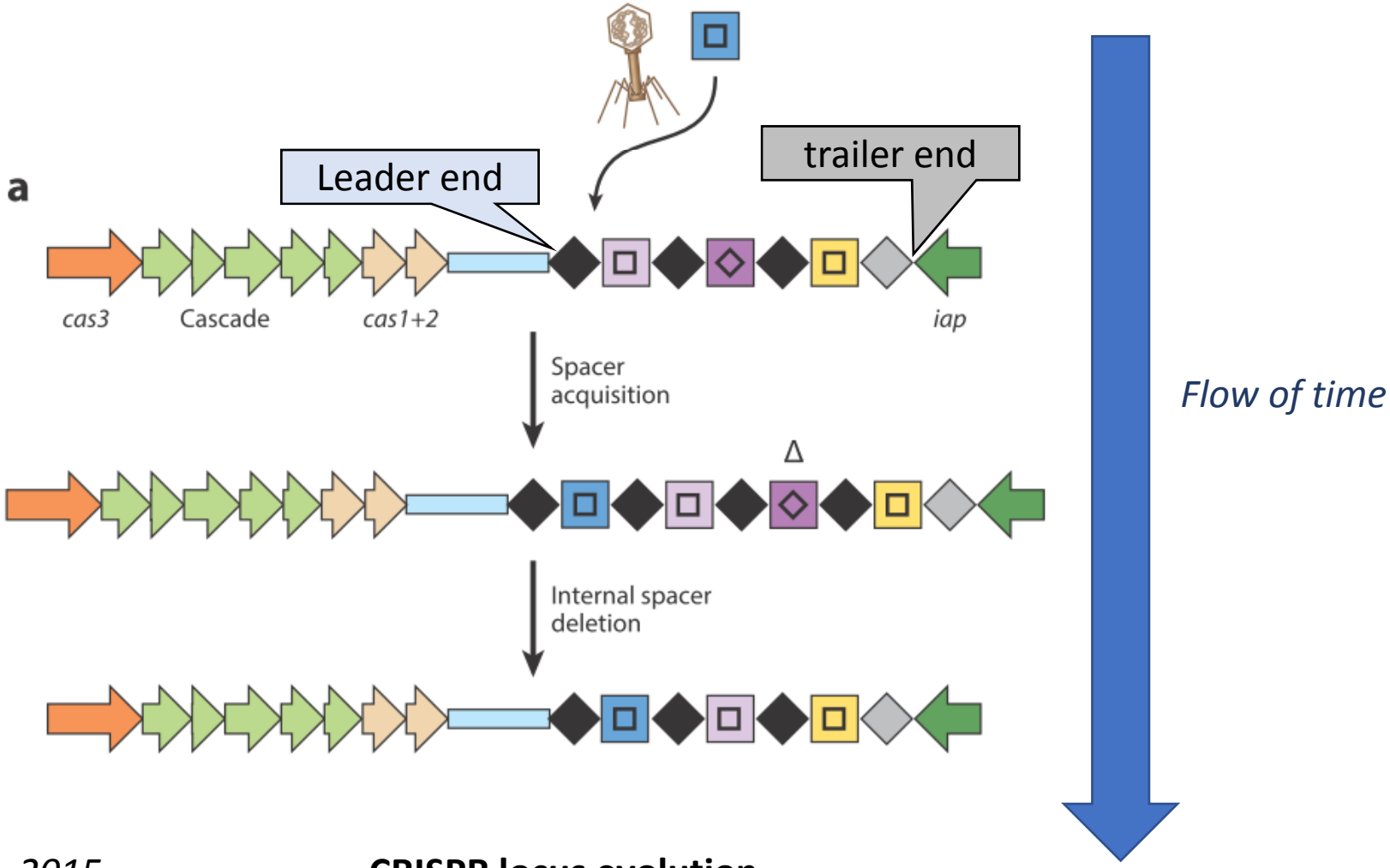
BIGSdb: Genome Comparator output

Genetic markers for reconstructing evolutionary history

High-resolution identification and typing using **CRISPR array**

- CRISPR array acquires novel ***spacer*** in an ordinal manner → infer the chronological record of bacterium
- Spacer can be identified using CRISPRFinder tool

Genetic markers for reconstructing evolutionary history



Lier et al., 2015

CRISPR locus evolution