UNIVERSITÀ CATTOLICA del Sacro Cuore

IN SILICO APPLICATIONS IN MICROBIOLOGICAL FOOD SAFETY AND RISK ASSESSMENT: THE ANALYSIS OF WHOLE GENOME SEQUENCE

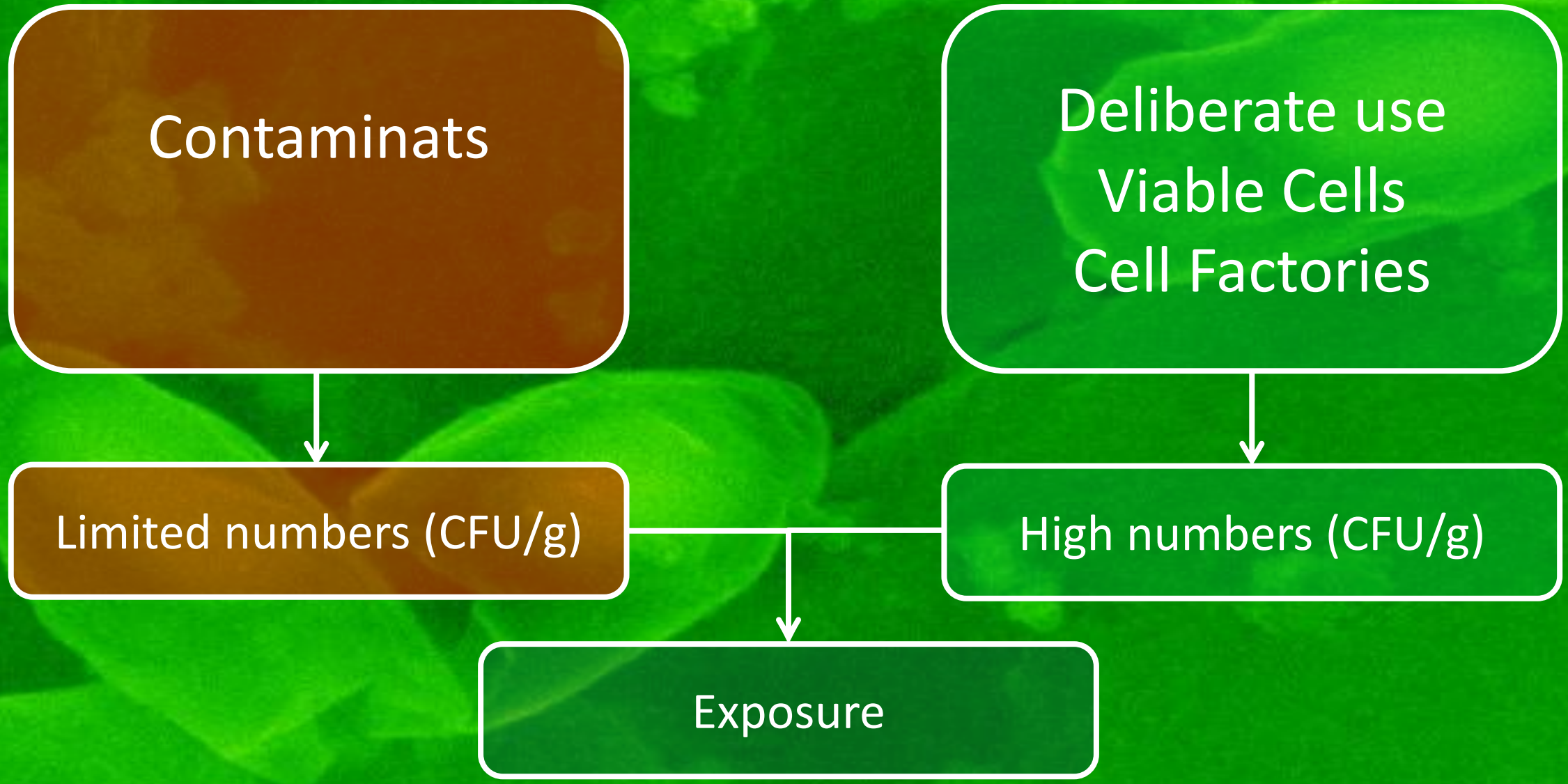
Pier Sandro Cocconcelli

Università Cattolica del Sacro Cuore

Italy

UNIVERSITÀ CATTOLICA del Sacro Cuore

Microorganisms in the Food Chain



UNIVERSITÀ CATTOLICA del Sacro Cuore

THE INTENTIONAL USE OF MICROORGANISMS IN THE FOOD CHAIN

USE	Authorisation
Food starter cultures	Not required
Human probiotics	Health Claims
Animal probiotics	Feed Additives
Feed fermentation-silage	Feed Additives
Enzyme production	Feed Additives - Food Additives
Amino acid production	Feed Additives - Food Additives
Vitamin production	Feed Additives - Food Additives
Biopesticide	Plant Protection Products
Genetically Modified Microorg.	Feed Additives - Food Additives- GMO

UNIVERSITÀ CATTOLICA del Sacro Cuore

The theoretic risk assessment

KNOWLEDGE AND DATA

HAZARD IDENTIFICATION

HAZARD CHARACTERIZATION

EXPOSURE ASSESSMENT

RISK CHARACTERIZATION

OPINION FOR RISK MANAGERS

UNIVERSITÀ CATTOLICA del Sacro Cuore

Microbiological Risk Assessment in Practice:

CONTAMINANTS, BIOLOGICAL HAZARDS…….

SCIENTIFIC KNOWLEDGE AND DATA COLLECTION	available? reliable? harmonised?
HAZARD IDENTIFICATION	emerging risks?
HAZARD CHARACTERIZATION	uncertainities
EXPOSURE ASSESSMENT	quantitative data? consumption info?
RISK CHARACTERIZATION	✓
OPINION FOR RISK MANAGERS	Regulation

UNIVERSITÀ CATTOLICA del Sacro Cuore

Microbiological Risk Assessment in practice:

Regulated Products

KNOWLEDGE AND DATA	proprietary data: completeness, trustable?
HAZARD IDENTIFICATION	✓
HAZARD CHARACTERIZATION	uncertainities
EXPOSURE ASSESSMENT	✓
RISK CHARACTERIZATION	✓
OPINION FOR RISK MANAGERS	Regulation

UNIVERSITÀ CATTOLICA del Sacro Cuore

A wide variety of microorganisms are intentionally added at different stages into the food chain, either directly or as a source of additives or enzymes.

- More than 130 bacterial species

  Arthrobacter globiformis ..... Streptococcus thermophilus

- Filamentous fungi

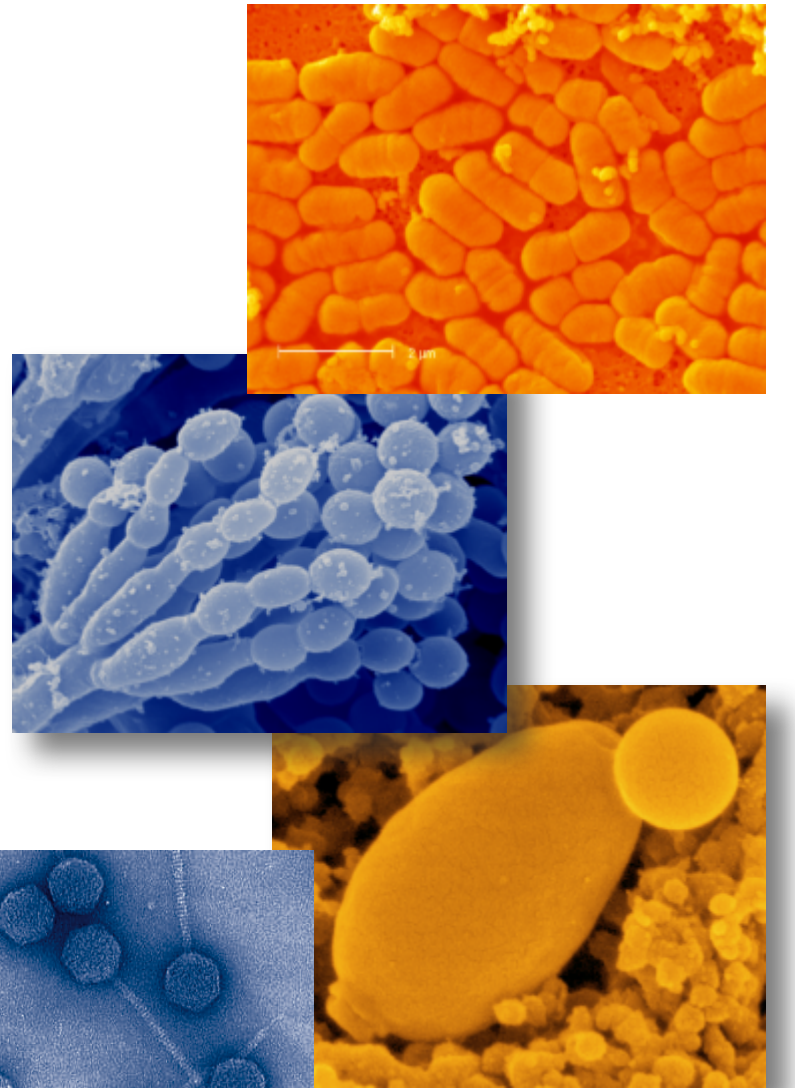
  Fusarium solani ..... Trichothecium domesticum

- Yeasts

  Candida famata .....Williopsis mrakii

- Virus

  Bacteriophages – Baculovirus -Potovirus



UNIVERSITÀ CATTOLICA del Sacro Cuore

Safety Assessment Guidelines and Guidance

Safety of microbial agents intended for deliberate introduction in the food chain and notified to EC-EFSA:

- for the consumer
- for the animals (when applicable)
- for the user/worker,
- for the environment

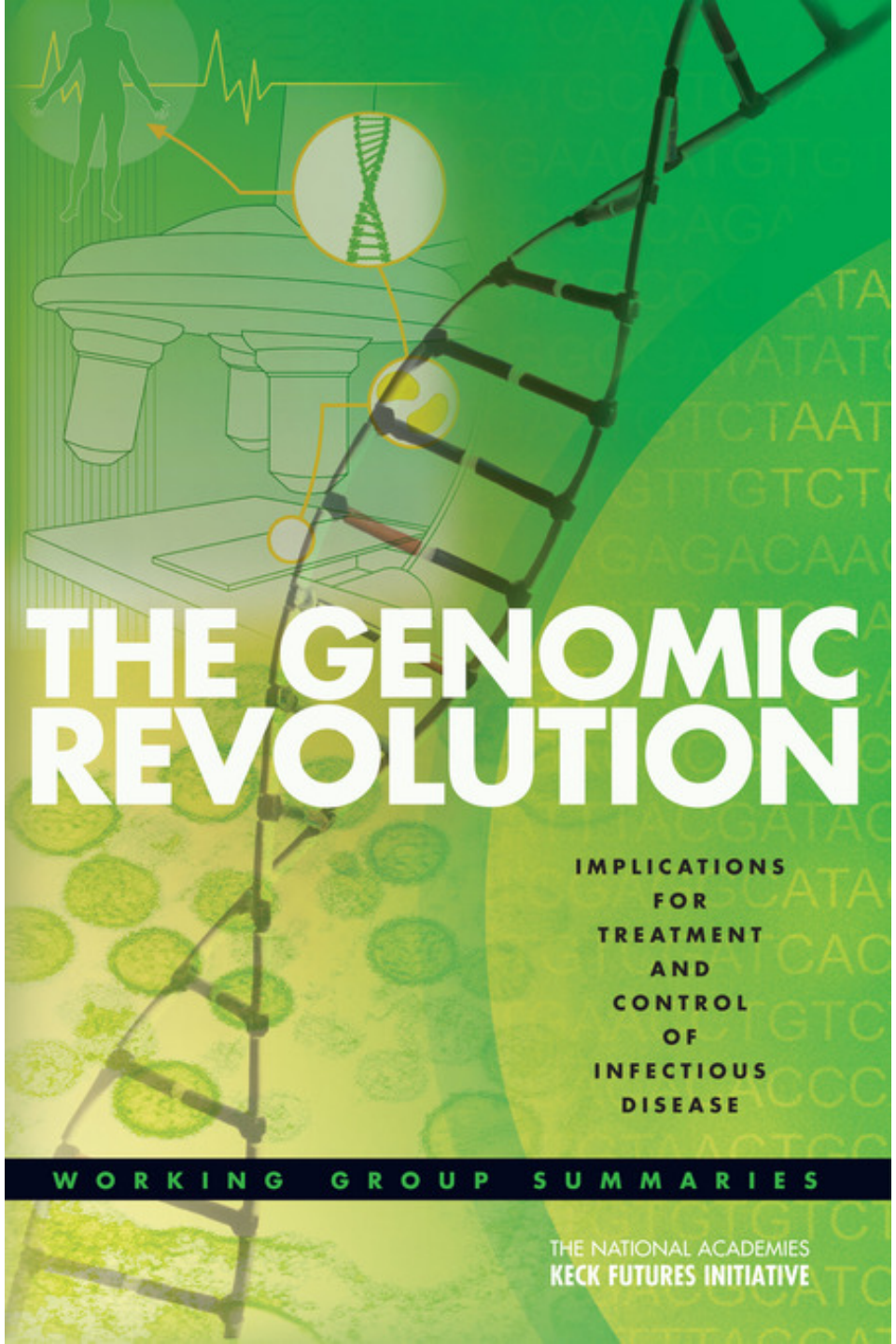
UNIVERSITÀ CATTOLICA del Sacro Cuore

Risk Assessment of Microbial products

- Microbial Agent
- Manufacturing process
- Product
  - Metabolism and residue studies
  - Toxicological studies
    - Acute toxicity
    - Genotoxicity studies (mutagenicity, clastogenicity)
    - Subchronic oral toxicity
    - Chronic oral toxicity/carcinogenicity
    - Reproduction toxicity including teratogenicity
  - Known toxins and virulence factors
  - Antibiotic susceptibility
  - Antibiotic production

How to assess the safety of a microbial product?

UNIVERSITÀ CATTOLICA del Sacro Cuore



Platform	Read length (bp)	Throughput	Reads	Runtime	Error profile	Instrument cost (US$)	Cost per Gb (US$, approx.)
Sequencing by ligation							
SOLiD 5500 Wildfire	50 (SE)	80 Gb	~700 M*	6 d*	≤0.1%, AT bias‡	NA§	$130‡
	75 (SE)	120 Gb					
	50 (SE)*	160 Gb*					
SOLiD 5500 xl	50 (SE)	160 Gb	~1.4 B*	10 d*	≤0.1%, AT bias‡	$251,000‡	$70‡
	75 (SE)	240 Gb					
	50 (SE)*	320 Gb*					
BGISEQ-500 FCS[155]	50–100 (SE/PE)*	8–40 Gb*	NA‖	24 h*	≤0.1%, AT bias‡	$250 (REF. 155)	NA‖
BGISEQ-500 FCL[155]	50–100 (SE/PE)*	40–200 Gb*	NA‖	24 h*	≤0.1%, AT bias‡	$250,000 (REF. 155)	NA‖
Sequencing by synthesis: CRT							
Illumina MiniSeq Mid output	150 (SE)*	2.1–2.4 Gb*	14–16 M*	17 h*	<1%, substitution‡	$50,000 (REF. 118)	$200–300 (REF. 118)
Illumina MiniSeq High output	75 (SE)	1.6–1.8 Gb	22–25 M (SE)*	7 h	<1%, substitution‡	$50,000 (REF. 118)	$200–300 (REF. 118)
	75 (PE)	3.3–3.7 Gb	44–50 M (PE)*	13 h			
	150 (PE)*	6.6–7.5 Gb*		24 h*			
Illumina MiSeq v2	36 (SE)	540–610 Mb	12–15 M (SE)	4 h	0.1%, substitution‡	$99,000‡	~$1,000
	25 (PE)	750–850 Mb	24–30 M (PE)*	5.5 h			$996
	150 (PE)	4.5–5.1 Gb		24 h			$212
	250 (PE)*	7.5–8.5 Gb*		39 h*			$142‡
Illumina MiSeq v3	75 (PE)	3.3–3.8 Gb	44–50 M (PE)*	21–56 h*	0.1%, substitution‡	$99,000‡	$250
	300 (PE)*	13.2–15 Gb*					$110‡
Illumina NextSeq 500/550 Mid output	75 (PE)	16–20 Gb	Up to 260 M (PE)*	15 h	<1%, substitution‡	$250‡	$42
	150 (PE)*	32–40 Gb*		26 h*			$40‡
Illumina NextSeq 500/550 High output	75 (SE)	25–30 Gb	400 M (SE)*	11 h	<1%, substitution‡	$250‡	$43
	75 (PE)	50–60 Gb	800 M (PE)*	18 h			$41
	150 (PE)*	100–120 Gb*		29 h*			$33‡
Illumina HiSeq2500 v2 Rapid run	36 (SE)	9–11Gb	300 M (SE)*	7 h	0.1%, substitution‡	$690‡	$230
	50 (PE)	25–30 Gb	600 M (PE)*	16 h			$90
	100 (PE)	50–60 Gb		27 h			$52
	150 (PE)	75–90 Gb		40 h			$45
	250 (PE)*	125–150 Gb*		60 h*			$40‡
Illumina HiSeq2500 v3	36 (SE)	47–52 Gb	1.5 B (SE)	2 d	0.1%, substitution‡	$690‡	$180
	50 (PE)	135–150 Gb	3 B (PE)*	5.5 d			$78
	100 (PE)*	270–300 Gb		11 d*			$45‡
Illumina HiSeq2500 v4	36 (SE)	64–72 Gb	2 B (SE)	29 h	0.1%, substitution‡	$690‡	$150
	50 (PE)	180–200 Gb	4 B (PE)*	2.5 d			$58
	100 (PE)	360–400 Gb		5 d			$45
	125 (PE)*	450–500 Gb*		6 d*			$30‡
Illumina HiSeq3000/4000	50 (SE)	105–125 Gb	2.5 B (SE)*	1–3.5 d*	0.1%, substitution‡	$740/$900 (REF. 156)	$50
	75 (PE)	325–375 Gb					$31
	150 (PE)*	650–750 Gb*					$22 (REF. 157)

Platform	Read length (bp)	Throughput	Reads	Runtime	Error profile	Instrument cost (US$)	Cost per Gb (US$, approx.)
Sequencing by synthesis: SNA (cont.)							
Illumina HiSeq X	150 (PE)*	800–900 Gb per flow cell*	2.6–3 B (PE)*	<3 d*	0.1%, substitution‡	$1,000‡,¶	$7.0‡
Qiagen GeneReader	NA‖	12 genes; 1,250 mutations[22]	NA‖	Several days[22]	Similar to other SBS systems[22]	NA‖	$400–$600 per panel[22]
Sequencing by synthesis: SNA							
454 GS Junior	Up to 600; 400 average (SE, PE)*	35 Mb*	~0.1 M*	10 h*	1%, indel‡	NA§	$40,000‡
454 GS Junior+	Up to 1,000; 700 average (SE, PE)*	70 Mb*	~0.1 M*	18 h*	1%, indel‡	$108,000‡	$19,500‡
454 GS FLX Titanium XLR70	Up to 600; 450 mode (SE, PE)*	450 Mb*	~1 M*	10 h*	1%, indel‡	NA§	$15,500‡
454 GS FLX Titanium XL+	Up to 1,000; 700 mode (SE, PE)*	700 Mb*	~1 M*	23 h*	1%, indel‡	$450,000‡	$9,500‡
Ion PGM 314	200 (SE)	30–50	400,000–550,000*	23 h	1%, indel‡	$49‡	$25–3,500‡
	400 (SE)	60–100 Mb*		3.7 h*			
Ion PGM 316	200 (SE)	300–500 Mb	2–3 M*	3 h	1%, indel‡	$49‡	$700–1,000‡
	400 (SE)*	600 Mb–1 Gb*		4.9 h*			
Ion PGM 318	200 (SE)	600 Mb–1 Gb	4–5.5 M*	4 h	1%, indel‡	$49‡	$450–800‡
	400 (SE)*	1–2 Gb*		7.3 h*			
Ion Proton	Up to 200 (SE)	Up to 10 Gb*	60–80 M*	2–4 h*	1%, indel‡	$224‡	$80‡
Ion S5 520	200 (SE)	600 Mb–1 Gb	3–5 M*	2.5 h	1%, indel‡	$65 (REF. 158)	$2,400*
	400 (SE)*	1.2–2 Gb*		4 h*			$1,200*
Ion S5 530	200 (SE)	3–4 Gb	15–20 M*	2.5 h	1%, indel‡	$65 (REF. 158)	$950*
	400 (SE)*	6–8 Gb*		4 h*			$475*
Ion S5 540	200 (SE)*	10–15 Gb*	60–80 M*	2.5 h*	1%, indel‡	$65 (REF. 158)	$300*
Single-molecule real-time long reads							
Pacific BioSciences RS II	~20 Kb	500 Mb–1 Gb*	~55,000*	4 h*	13% single pass, ≤1% circular consensus read, indel‡	$695‡	$1,000‡
Pacific Biosciences Sequel	8–12 Kb[69]	3.5–7 Gb*	~350,000*	0.5–6 h*	NA‖	$350 (REF. 69)	NA‖
Oxford Nanopore MK 1 MinION	Up to 200 Kb[159]	Up to 1.5 Gb[159]	>100,000 (REF. 159)	Up to 48 h[160]	~12%, indel[159]	$1,000*	$750*
Oxford Nanopore PromethION	NA‖	Up to 4 Tb*	NA‖	NA‖	NA‖	$75*	NA‖
Synthetic long reads							
Illumina Synthetic Long-Read	~100 Kb synthetic length*	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500 (possible barcoding and partitioning errors)	No additional instrument required	~$1,000*
10X Genomics	Up to 100 Kb synthetic length*	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500 (possible barcoding and partitioning errors)	$75 (REFS 72,161)	See HiSeq 2500 +$500 per sample[161]

Goodwin, McPherson and McCombie
NATURE REVIEWS | GENETICS VOLUME 17 | JUNE 2016

Platform	Read length (bp)	Throughput	Reads	Runtime	Error profile	Instrument cost (US$)	Cost per Gb (US$, approx.)
Sequencing by ligation							
SOLiD 5500 Wildfire	50 (SE)	80 Gb	~700 M*	6 d*	≤0.1%, AT bias‡	NA§	$130‡
	75 (SE)	120 Gb					
	50 (SE)*	160 Gb*					
SOLiD 5500 xl	50 (SE)	160 Gb	~1.4 B*	10 d*	≤0.1%, AT bias‡	$251,000‡	$70‡
	75 (SE)	240 Gb					
	50 (SE)*	320 Gb*					
BGISEQ-500 FCS[155]	50–100 (SE/PE)*	8–40 Gb*	NA‖	24 h*	≤0.1%, AT bias‡	$250 (REF. 155)	NA‖
BGISEQ-500 FCL[155]	50–100 (SE/PE)*	40–200 Gb*					
Sequencing by synthesis: CRT							
Illumina MiniSeq Mid output	150 (SE)*	2.1–2.4 Gb*					
Illumina MiniSeq High output	75 (SE)	1.6–1.8 Gb					
	75 (PE)	3.3–3.7 Gb					
	150 (PE)*	6.6–7.5 Gb*					
Illumina MiSeq v2	36 (SE)	540–610 Mb					
	25 (PE)	750–850 Mb					
	150 (PE)	4.5–5.1 Gb					
	250 (PE)*	7.5–8.5 Gb*					
Illumina MiSeq v3	75 (PE)	3.3–3.8 Gb					
	300 (PE)*	13.2–15 Gb*					
Illumina NextSeq 500/550 Mid output	75 (PE)	16–20 Gb					
	150 (PE)*	32–40 Gb*					
Illumina NextSeq 500/550 High output	75 (SE)	25–30 Gb					
	75 (PE)	50–60 Gb					
	150 (PE)*	100–120 Gb*					
Illumina HiSeq2500 v2 Rapid run	36 (SE)	9–11 Gb					
	50 (PE)	25–30 Gb					
	100 (PE)	50–60 Gb					
	150 (PE)	75–90 Gb					
	250 (PE)*	125–150 Gb*		60 h*			$40‡
Illumina HiSeq2500 v3	36 (SE)	47–52 Gb	1.5 B (SE)	2 d	0.1%, substitution‡	$690‡	$180
	50 (PE)	135–150 Gb	3 B (PE)*	5.5 d			$78
	100 (PE)*	270–300 Gb		11 d*			$45‡
Illumina HiSeq2500 v4	36 (SE)	64–72 Gb	2 B (SE)	29 h	0.1%, substitution‡	$690‡	$150
	50 (PE)	180–200 Gb	4 B (PE)*	2.5 d			$58
	100 (PE)	360–400 Gb		5 d			$45
	125 (PE)*	450–500 Gb*		6 d*			$30‡
Illumina HiSeq3000/4000	50 (SE)	105–125 Gb	2.5 B (SE)*	1–3.5 d*	0.1%, substitution‡	$740/$900 (REF. 156)	$50
	75 (PE)	325–375 Gb					$31
	150 (PE)*	650–750 Gb*					$22 (REF. 157)

Platform	Read length (bp)	Throughput	Reads	Runtime	Error profile	Instrument cost (US$)	Cost per Gb (US$, approx.)
Sequencing by synthesis: SNA (cont.)							
Illumina HiSeq X	150 (PE)*	800–900 Gb per flow cell*	2.6–3 B (PE)*	<3 d*	0.1%, substitution‡	$1,000‡,¶	$7.0‡
Qiagen GeneReader	NA‖	12 genes; 1,250 mutations[22]	NA‖	Several days[22]	Similar to other SBS systems[22]	NA‖	$400–$600 per panel[22]
Sequencing by synthesis: SNA							
454 GS Junior	Up to 600; 400 average (SE, PE)*	35 Mb*	~0.1 M*	10 h*	1%, indel‡	NA§	$40,000‡
454 GS Junior+	Up to 1,000; 700 average (SE, PE)*	70 Mb*	~0.1 M*	18 h*	1%, indel‡	$108,000‡	$19,500‡
				10 h*	1%, indel‡	NA§	$15,500‡
				23 h*	1%, indel‡	$450,000‡	$9,500‡
				23 h	1%, indel‡	$49‡	$25–3,500‡
				8.7 h*			
				8 h	1%, indel‡	$49‡	$700–1,000‡
				4.9 h*			
				4 h	1%, indel‡	$49‡	$450–800‡
				7.3 h*			
				2–4 h*	1%, indel‡	$224‡	$80‡
				2.5 h	1%, indel‡	$65 (REF. 158)	$2,400*
				4 h*			$1,200*
				2.5 h	1%, indel‡	$65 (REF. 158)	$950*
				4 h*			$475*
				2.5 h*	1%, indel‡	$65 (REF. 158)	$300*
				4 h*	13% single pass, ≤1% circular consensus read, indel‡	$695‡	$1,000‡
				0.5–6 h*	NA‖	$350 (REF. 69)	NA‖
Oxford Nanopore MK 1 MinION	Up to 200 Kb[159]	Up to 1.5 Gb[159]	>100,000 (REF. 159)	Up to 48 h[160]	~12%, indel[159]	$1,000*	$750*
Oxford Nanopore PromethION	NA‖	Up to 4 Tb*	NA‖	NA‖	NA‖	$75*	NA‖
Synthetic long reads							
Illumina Synthetic Long-Read	~100 Kb synthetic length*	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500 (possible barcoding and partitioning errors)	No additional instrument required	~$1,000*
10X Genomics	Up to 100 Kb synthetic length*	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500	See HiSeq 2500 (possible barcoding and partitioning errors)	$75 (REFS 72,161)	See HiSeq 2500 +$500 per sample[161]

UP TO 900Gb per run
COST ≥7 €/Gb
ERROR PROFILE <0.1%

UNIVERSITÀ CATTOLICA del Sacro Cuore

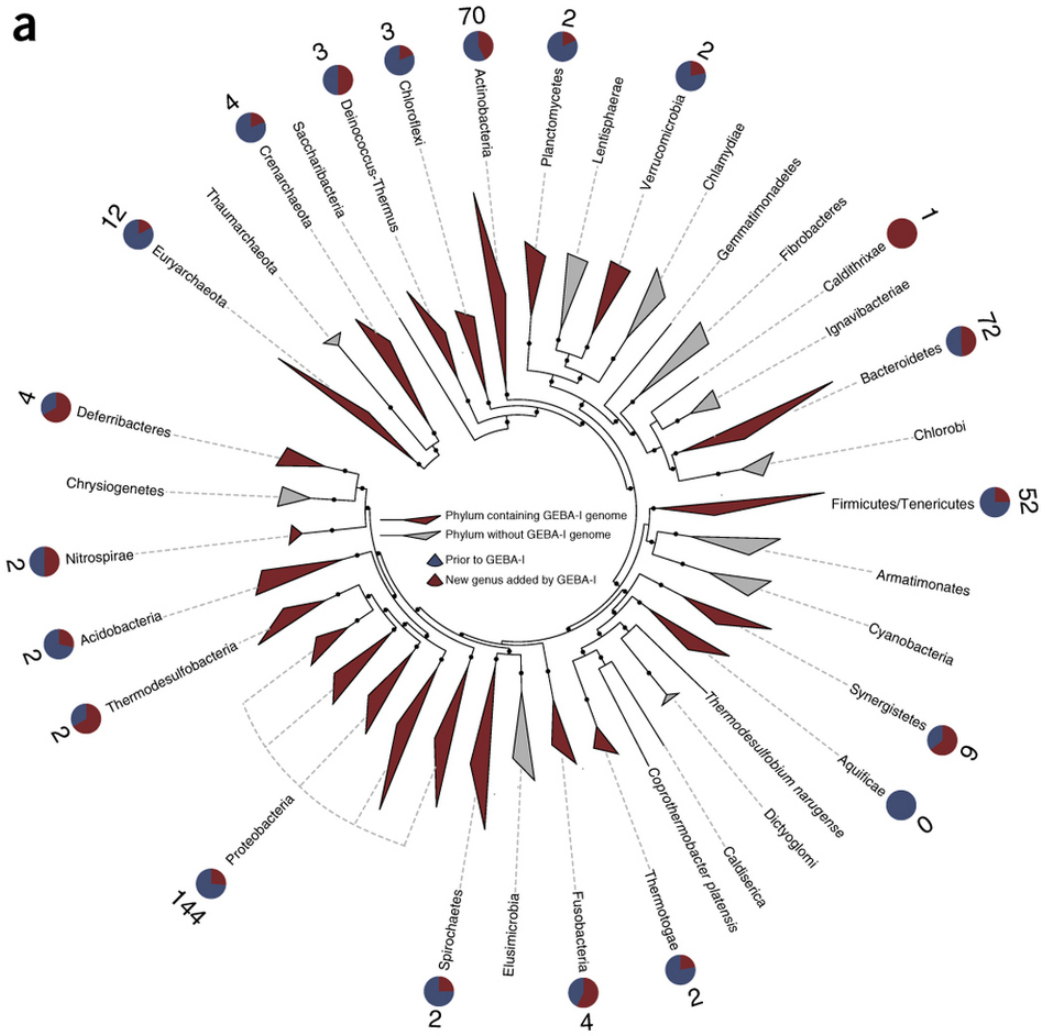
nature biotechnology

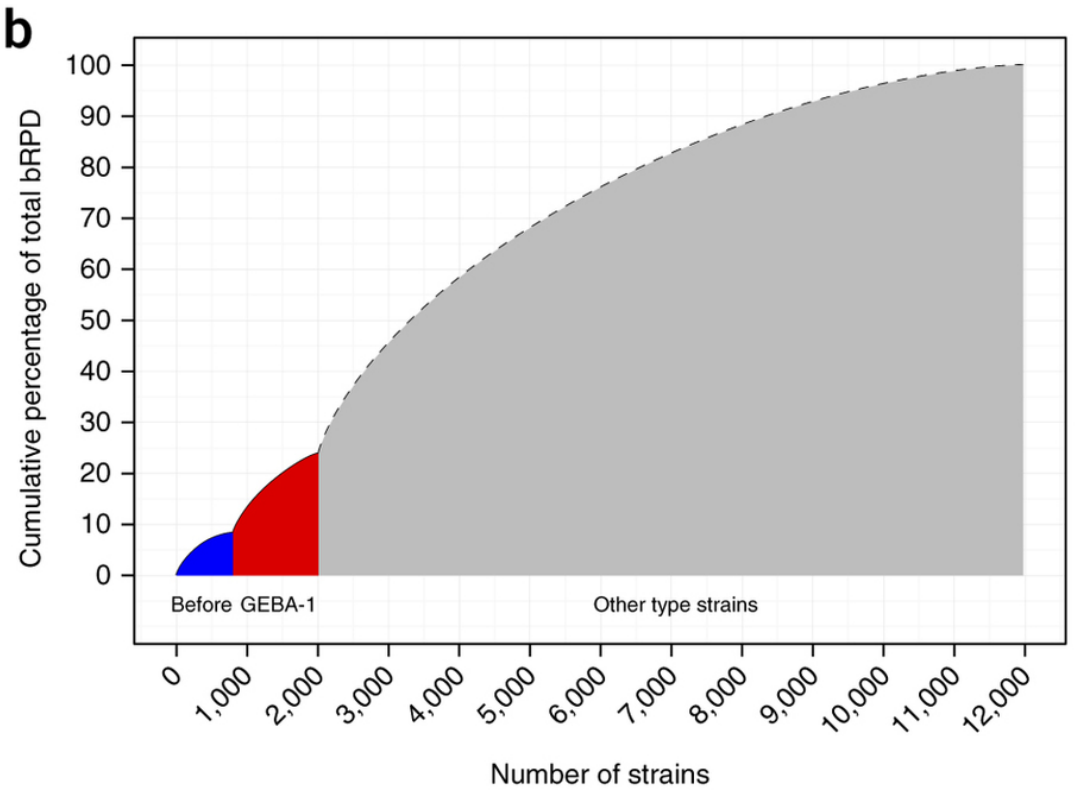
OPEN

Genomic Encyclopedia of Bacteria and Archaea (GEBA)

1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life

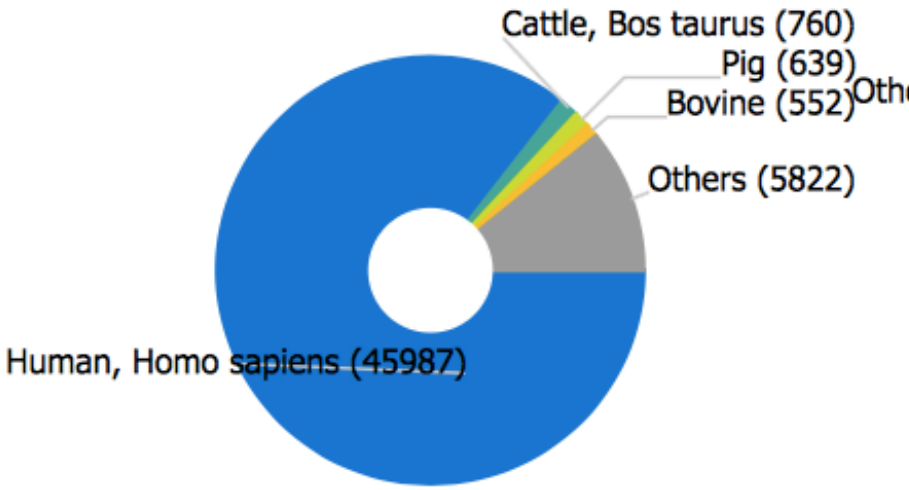
Supratim Mukherjee[1,10], Rekha Seshadri[1,10], Neha J Varghese[1], Emiley A Eloe-Fadrosh[1], Jan P Meier-Kolthoff[2], Markus Göker[2], R Cameron Coates[1,9], Michalis Hadjithomas[1], Georgios A Pavlopoulos[1], David Paez-Espino[1], Yasuo Yoshikuni[1], Axel Visel[1], William B Whitman[3], George M Garrity[4,5], Jonathan A Eisen[6], Philip Hugenholtz[7], Amrita Pati[1,9], Natalia N Ivanova[1], Tanja Woyke[1], Hans-Peter Klenk[8] & Nikos C Kyrpides[1]



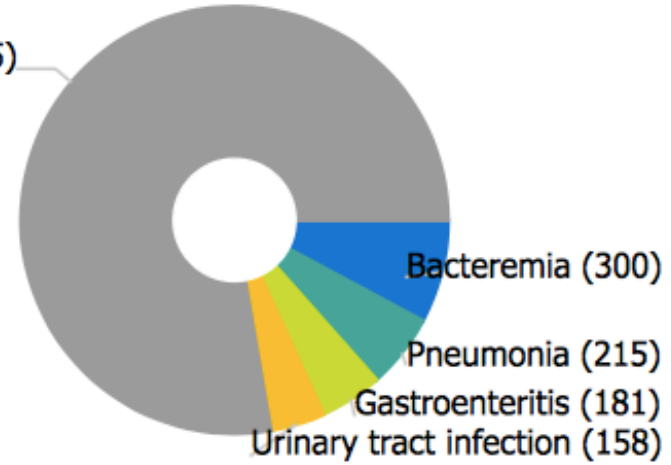


Genomes By Metadata

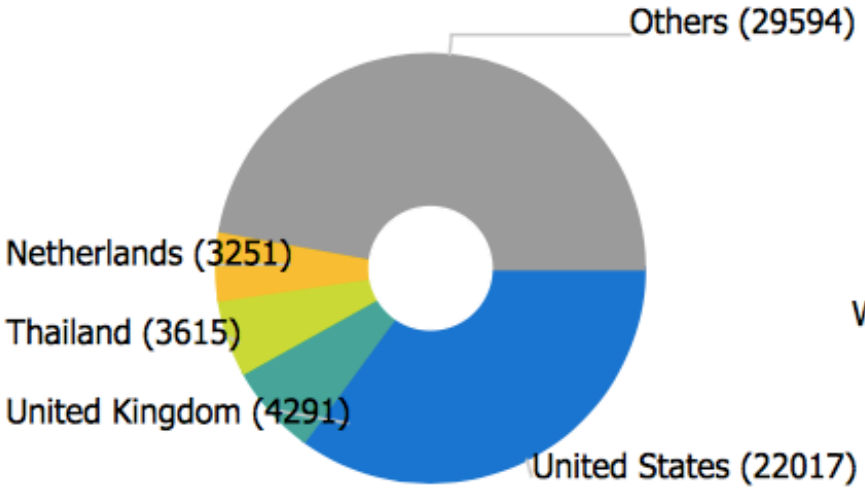
Host Name



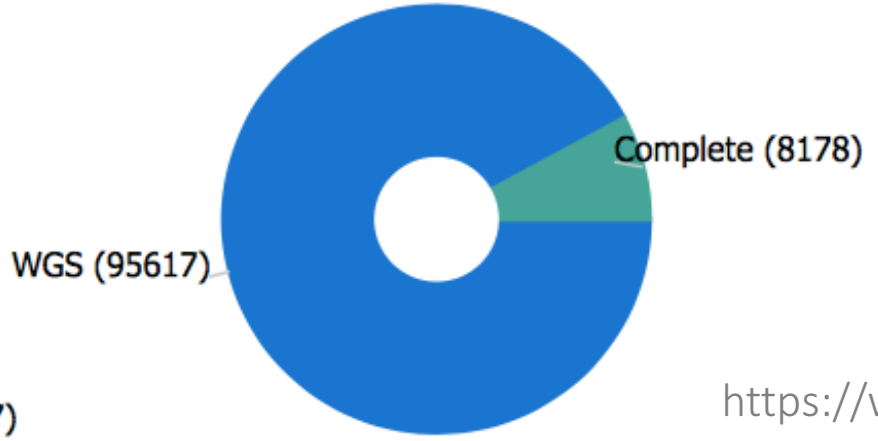
Disease



Isolation Country

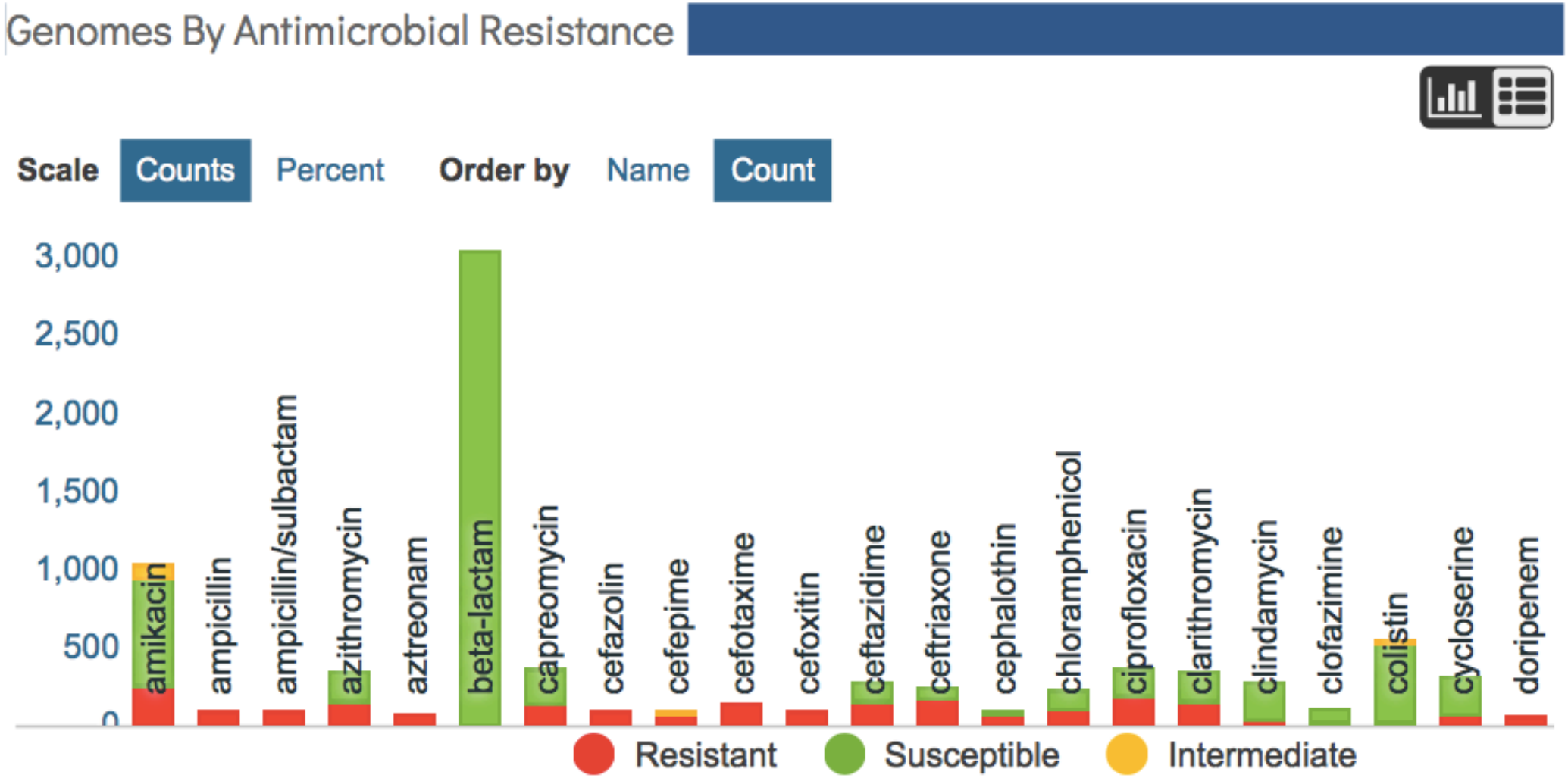


Genome Status



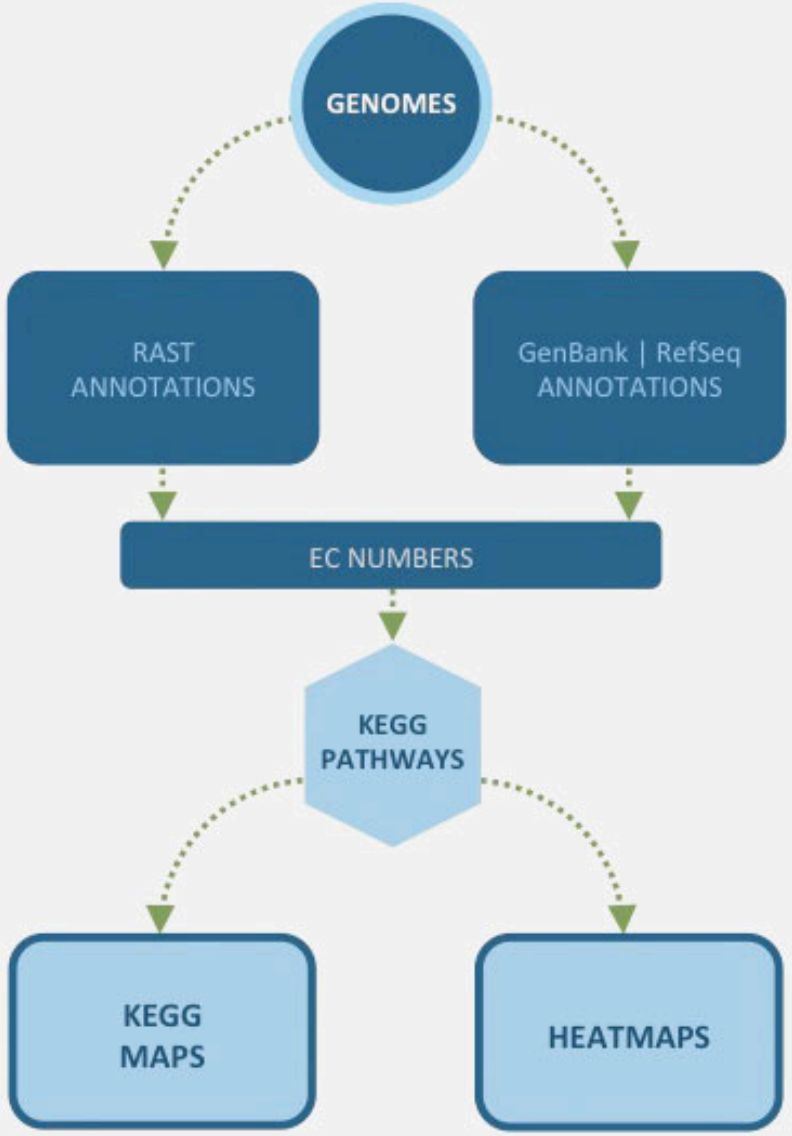
https://www.patricbrc.org

UNIVERSITÀ CATTOLICA del Sacro Cuore



https://www.patricbrc.org

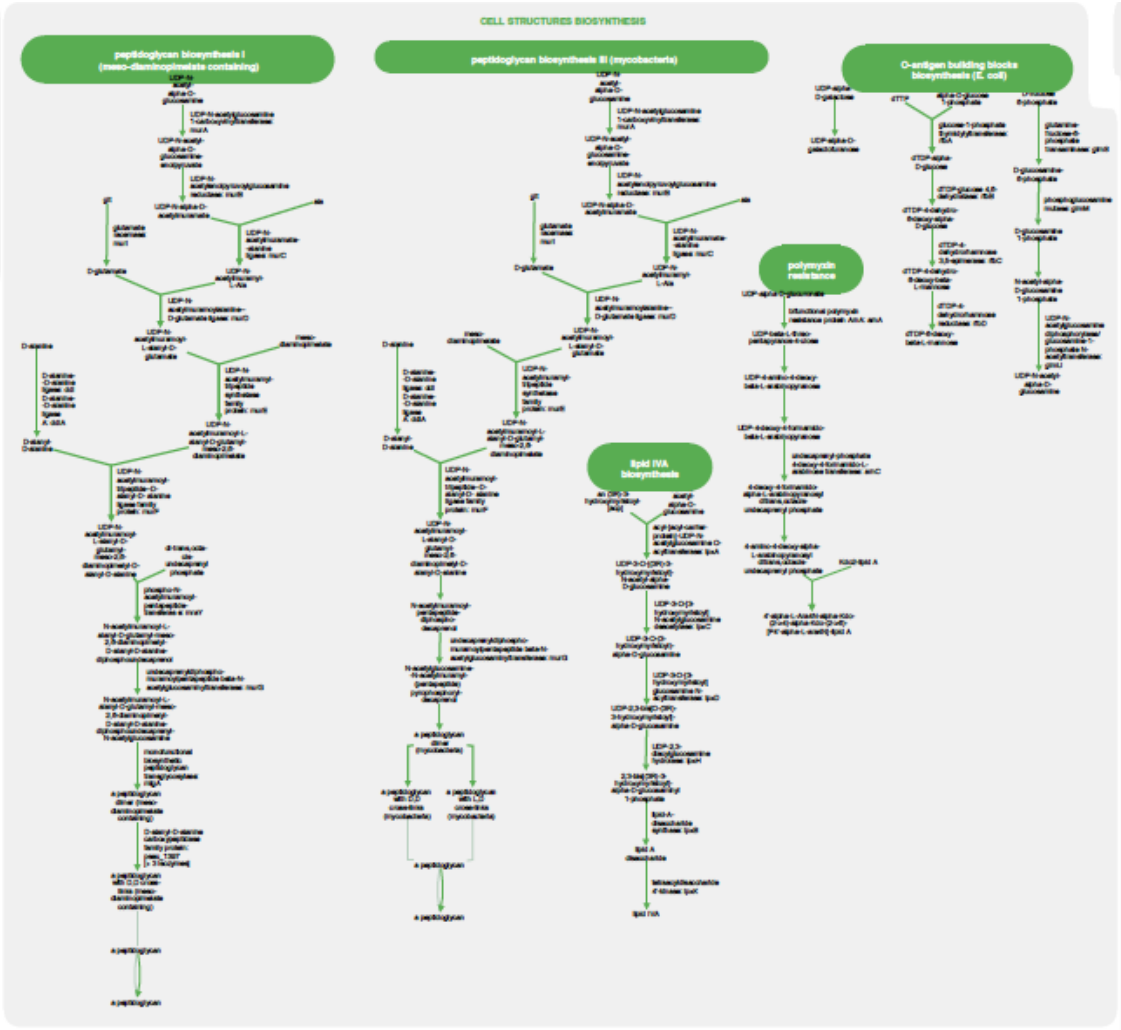
How Do We Process and Clean Up Our Data?

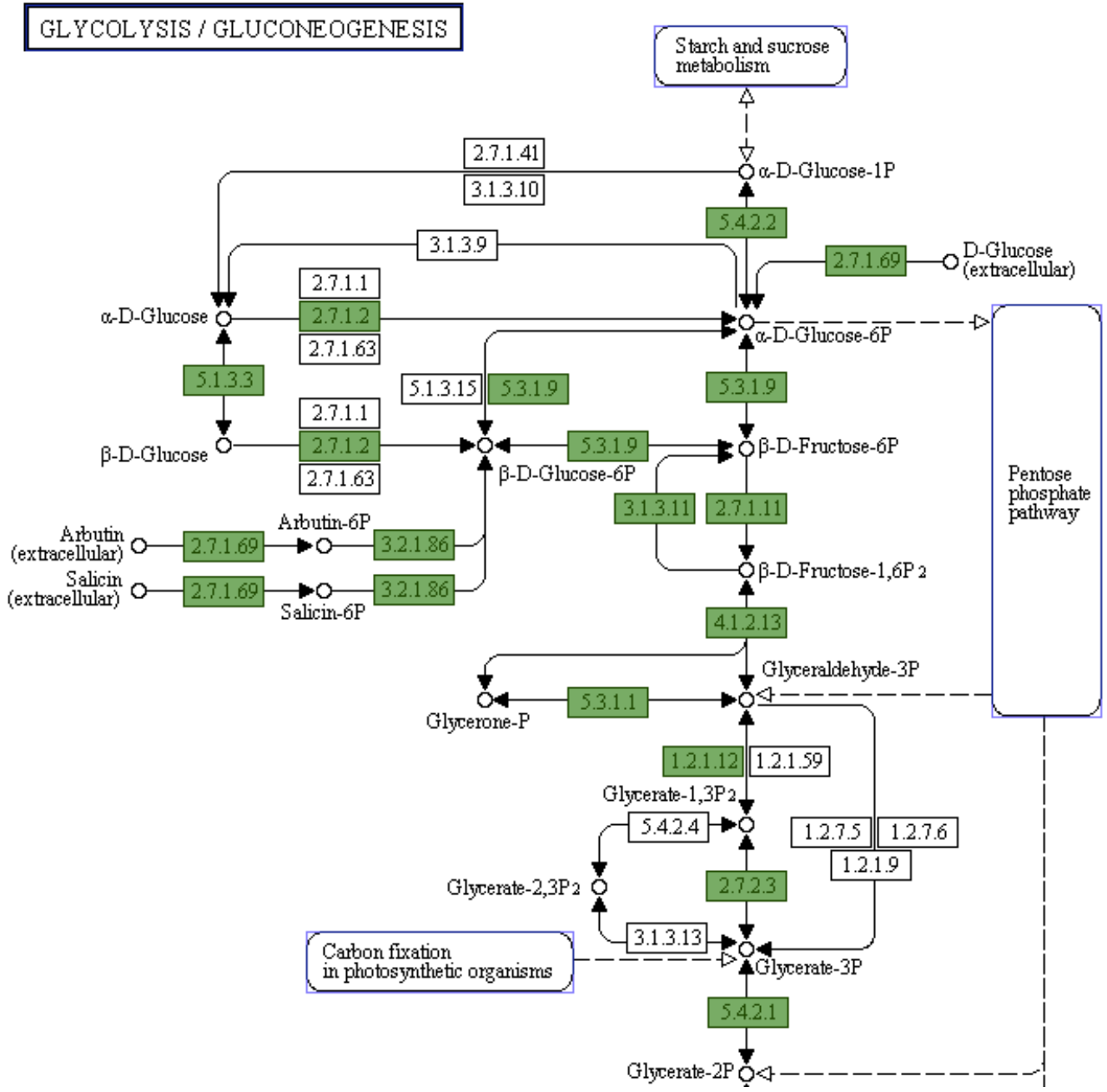


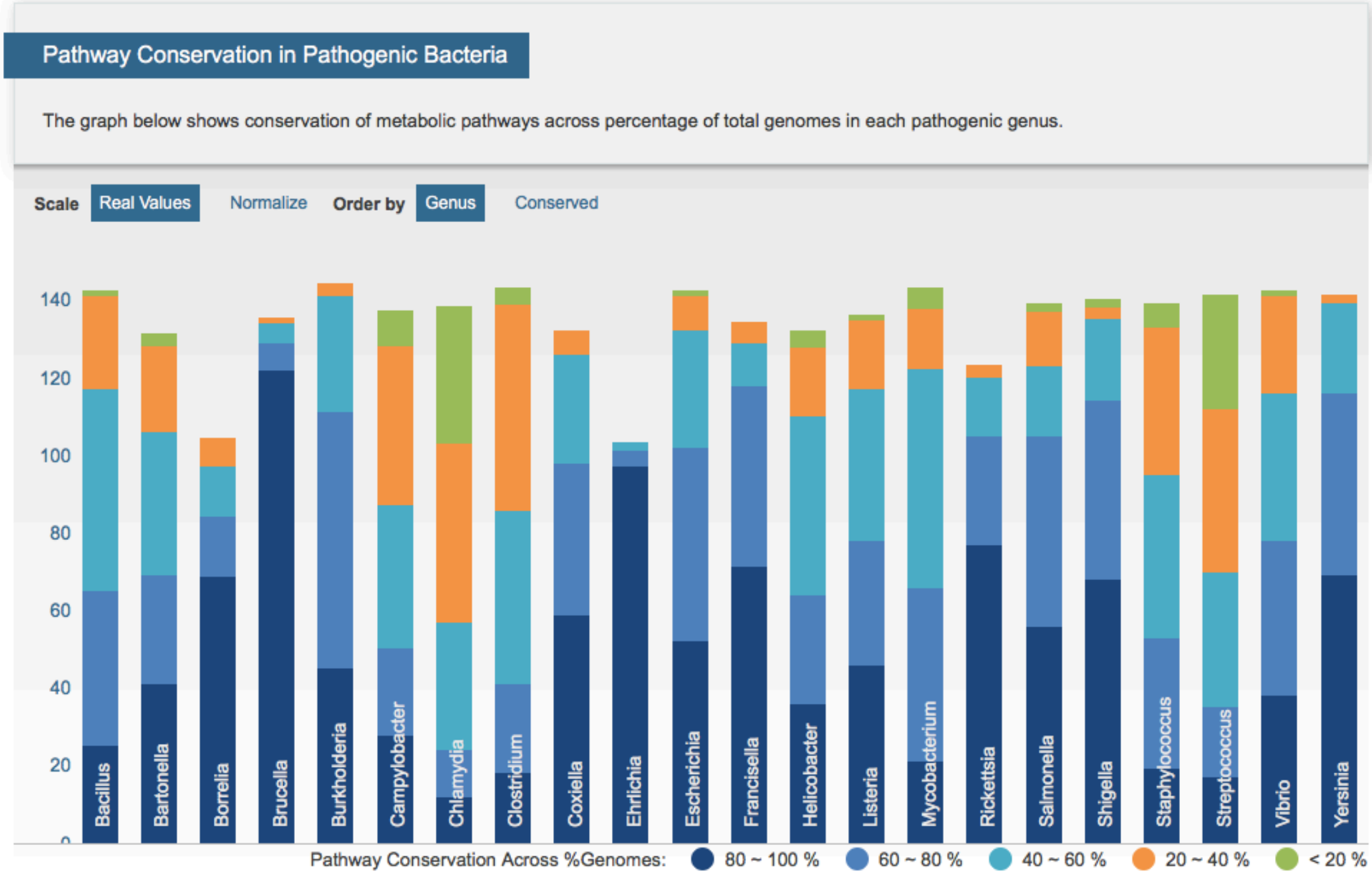
https://www.patricbrc.org

UNIVERSITÀ
CATTOLICA
del Sacro Cuore

UNIVERSITÀ CATTOLICA del Sacro Cuore







https://www.patricbrc.org

UNIVERSITÀ CATTOLICA del Sacro Cuore

What's needed nowadays for a whole genome sequence?

- An interesting bacterial strain
- A smart PhD student for DNA extraction
- Outsourcing the sequencing
- A PhD Student with some experience in bioinformatic

- But……

How to genomic data can be used to assess the safety of a microbial product?

UNIVERSITÀ CATTOLICA del Sacro Cuore

How a WGS looks like?

>g30571_seq5
ttcgtgttctcttagacataaaaattcctcctggtatgttttcttctagtctacacaccttaatttttctgtctagttaattgtagcctatccaatagattacaaatagtatcgtatttcgcaatcttaa
aattatcttcttactgtgaataagataataaaatttgtatttgtccttttaggtaacttattttaatctatataccttctggatcaacgaaagtatactttatatagtcataacgccgatggtaacta
tgcactgtaacgataatgctgctaaaatatctttttacgaaaaacatcacaaggaaaaacttaactatggtattatctaaacataatttttcggtatttaatatttaattaggaggtaatattttgctt
ttttatagtaatttatcttaatagtagcgatattattacttctgaatatatggatttttgataggagtagaaatgcaagtataggttttagaactaaacgaagtctgtcttctaaaaaaaattgggtat
actcacaaactatttttatggaggaattgttttaattagtttactttcttcaacactttattctttgaatattatagatgtttccacctcaaattcaatatctattattggaatcattattgctgcaat
tattacacaactgtttttagtgtttggagagaagaaaagaagtaaaaaatagaaaaactagtgatagttttggtgatataagaaaaacatatatgtataaaaagaagaaagtagcttctctttgtacata
tatggtttgtttttatggatgccgagctttttataagataaggctaccgcccataacgacccaaggatttgtggaactaagtaaatacctgcaagaagcaagatcagctgggtcggttttaaaggatga
tagaagtaacgccattcagtaaaattggcgcaaaataaataggaagtgacagtagatataatattggcaatgaataaagatataatacgcctgttcctttgaaaataatatccagcaaaaagagataaaa
atatcaatgtgatgaacccgcttacggaatgataacgaaaatcaagcagtaaggcgaaatgcacataaggaacaaaagataaaagcaacggaaggtttttccagcaagttttatcatctttcatcacca
cctaacaagtgttattattttgttagttttattgtaaattgaatcctgcaataatcaattacttttttcattatttctagattaaatgaataaactctcatgccattatctctaaaaaaaaagtttttccc
cctaatttttttagatagaaaattgtatagtatgatgagagaaaaagtagtttgggggaacgttaatgagaagaaaagcagtaatcgatcggctttatttatttttttggtttttttcctatcgttttgct
accttttacatgggtaaatgctgttgtacaaggtgcggagatattaggcctgatccaaggaagcggattagtggtttctaggaattttccagctttgactgttttctatttagtgataagtgtaggtctc
ttattaccaaaatctctgatcaatcactacaattggcagctgctttacttttcttgcttattagtatttacatctatgttttctttgctgcctcgcttgattatcggccctaatggtttgagcgagagtg
gattgaaggattcagcttattcctccatgtatttcttgacgatcaaaccatggttctatttagtaattgcctcatttattcttgcaacatttttttacttctggacagaaatcaggataaaaagtgaacg
atctgtatctataaaaaataactattgatggtggatcatcagtaaaaaagcatatgagctgcgtaaaatgcatgtttctcatatgctttttttattattatgtaaagaatgcgaaaaacgtttacatgat
ttttacataacttaactatatttttacagcaaccatacgagggatttacagtcgactgataaactttagttgtaaataagaaaggagcaataaaatgaaaaaagtattgtttactctaggtatgatgggga
attcttttagccggatgcggctcaggatcaggaaatacagatggctctgccacacaaaataattcaagtgactcaaatgaacaagtaaaaatcgttgcggtaggttcgacagcattacagccgttagtcg
atgcagcacaagaaagctttgtacaagaaaatccaaattatcaaatctctgtacaaggaggaggaagcggaactggtttgagccaagtcgaagctggagcagttactatcggaaactctgatgtatttgc
agaagaaagagacggagttgatgcttcaaaattagtggatcataaagtagcagtggtcgggatggcaccaatcgtcaataaagatacagatgtcaaagatatcaccaaacaggaattgattgatatcttt
acaggaaaaatcactaactggaaggaagttggcgggaaagatcaaaaaatcaacgtcgtgaaccgtgcaaacggcagcggtacccgcgcgacattcgaaaaatggggactggatggtgctacccctgtgc
agtcccaagaacaagattcatcaggaacagttcggcagctggtaagtcaaacaccaggagcaatcagctatttagctttctcatatctcgatgattccacacaagcactaagtatcgacggtgtagaacc
aaaagaagagaatgtagcagataacagctggggaatctggtcatatgaacatatgtacacgaacggaaaaccttcccctgaagtccaaaaattcttagactacatgatgacagaagaaatccaagaagga
cctgtcaaagaattgggctatttaccattcacaatgatggaagtagaacgtgatcacgaaggaaatatcaaataactagcagataaacatgccacggtagtcgttactattaaggaaaaatcaactatgt
ttccccacatatacatttttcctttaattgtgaaaaatagttcctcatattttgccgtttttgagctaataccgagggtcagctgattgaacaccgtttattcgaaaaaaagagctacgacatagaact
gtcgtagctttttttttgctatatgacaggaatgaaaggtgtttgcaagcttaatgtgaggaacactttccatatctgttctatactggataggagaagtttggaaatatccagaaaacgaaatcaatcg
aggtagccacatgaagaataaatggataaatcgaagtaaaaaaaagaaagtaacacaaagatcatcagataaagtcatccgtctttttccatcgtataaagaaggattaacggcagatcaagtcaaggaa
cgaatagaaaaaggtgcagcaaataattcagtcgatcccacattcaaaacaaatcagcaaatcgtattggaaaacatatttacgtatttcaatttgatctttcttatcttagccattttgctttgtttag
tagaatcctataagaacttgacttttttaccagttatcat……..

UNIVERSITÀ CATTOLICA del Sacro Cuore

Assessment of Regulated MO

	Feed additives containing viable microorganisms		Fermentation products	
	Bacteria	Fungi - yeasts	Bacteria	Fungi - yeasts
Identification	✓	✓	✓	✓
Antimicrobial susceptibility	✓		✓	
Antimicrobial production	✓	✓	✓	✓
Toxigenicity and pathogenicity	✓	✓	✓	✓
Genetic modification			For GMMs only	For GMMs only
Absence of the production strain			✓	✓
Presence of DNA from the production strain			where relevant	where relevant

UNIVERSITÀ CATTOLICA del Sacro Cuore

WGS sequence in the RA of microorganims

	Feed additives containing viable microorganisms			Fermentation products		
	Bacteria	Yeasts	Fungi	Bacteria	Yeasts	Fungi
Identification	✓	✓	✓	✓	✓	✓
Antimicrobial susceptibility	✓			✓		
Antimicrobial production	✓	✓	✓	✓	✓	✓
Toxigenicity and pathogenicity	✓	✓	✓	✓	✓	✓
Genetic modification				For GMMs only	For GMMs only	For GMMs only
Absence of the production strain				✓	✓	✓
Presence of DNA from the production strain				where relevant	where relevant	where relevant

UNIVERSITÀ CATTOLICA del Sacro Cuore

Use of whole genome sequence for characterisation of microorganisms

- Whole genome sequence analysis (chromosome(s) and/or extra-chromosomal genetic elements e.g. plasmids) for bacterial and yeast strains intended for use either as products or production strains.
- WGS analysis is also recommended for filamentous fungi.
- WGS data provide information
  - unequivocal taxonomic identification of the strain,
  - virulence factors,
  - production of or resistance to antimicrobials of clinical relevance,
  - production of known toxic metabolites

UNIVERSITÀ CATTOLICA del Sacro Cuore

Quality Criteria

- the sequencing strategy and instrumentation used
- the assembly method applied (e.g., the bioinformatic approach, *de novo* or re-seq strategy)
- the statistical measure of sequence quality (e.g., number of reads, coverage, N50 and K-mer)
- the number of contigs and scaffolds required to represent the genome, their absolute length and their length relative to the genome size
- the annotation protocol used
- for fungi: information on the quality of the annotations obtained from relevant databases (e.g., BUSCO)

UNIVERSITÀ CATTOLICA del Sacro Cuore

Identification

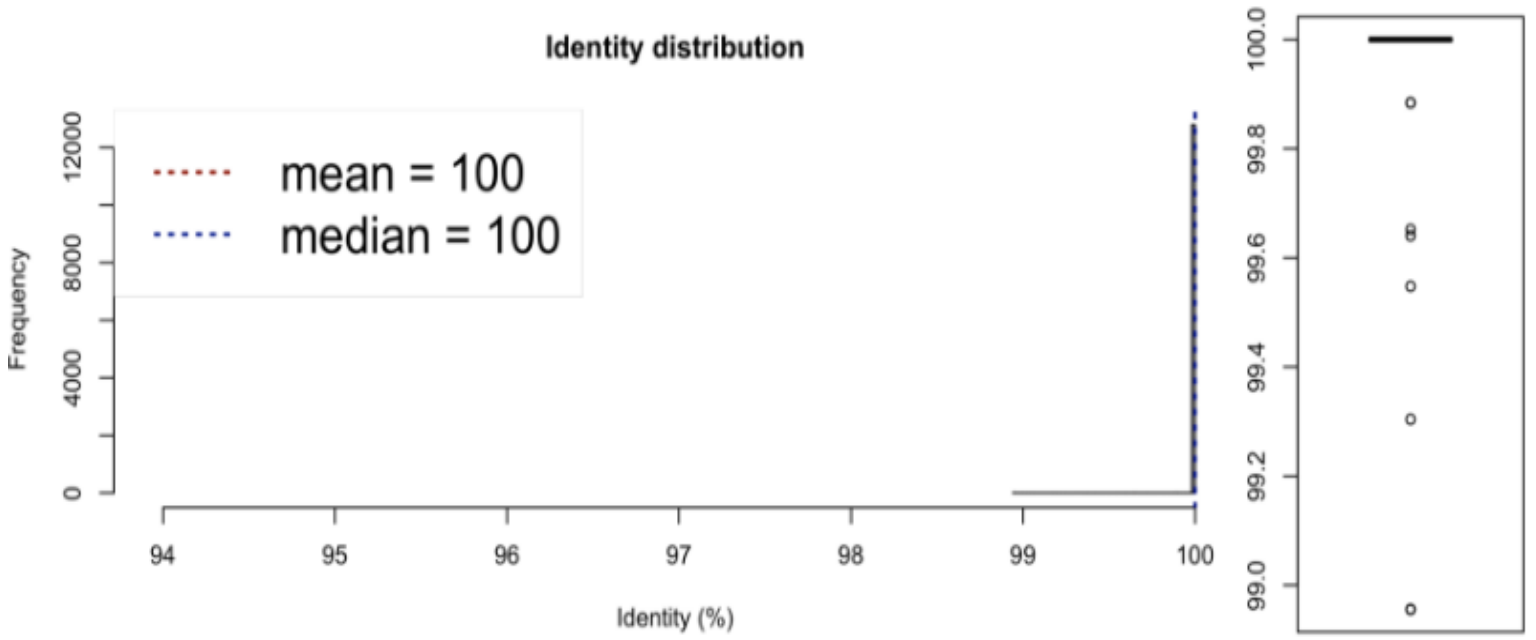
- Bacteria: Data from whole genome sequence (WGS)
  - 16S rRNA gene, housekeeping genes
  - computational approach for taxonomic assignments (e.g., phylogenomics or average nucleotide identity [ANI]).
- Yeasts: Data from whole genome sequence (WGS)
  - phylogenomic analysis (e.g. using a concatenation of several conserved genes to produce a phylogeny against available related genomes).
- Filamentous fungi:
  - When WGS is available, a phylogenomic analysis comparing the genome against available related genomes.

UNIVERSITÀ CATTOLICA del Sacro Cuore

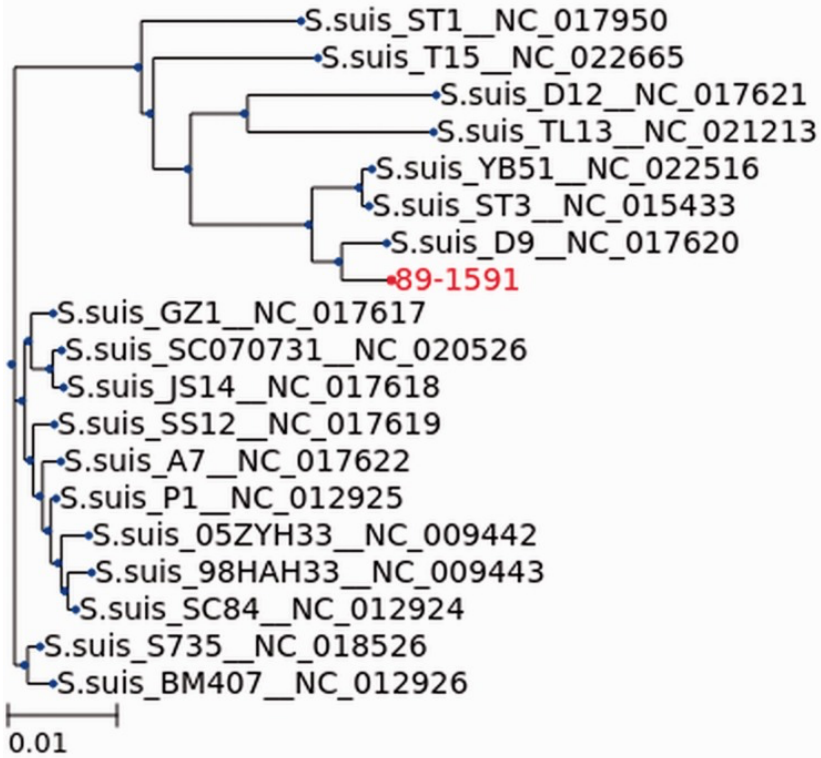
ANI: Average Nuceoltide Identity

Average nucleotide identity (ANI), calculated from pair-wise comparisons of all sequences shared between any two strains, is ciìonsidered as the new metrics for bacterial species definition and classification.

菌株名/Strain Name	菌株ID/Strain ID in NCBI	ANI value comparing with query strain
S.suis_D9	NC_017620	99.05%
S.suis_YB51	NC_022516	98.70%
S.suis_ST3	NC_015433	98.67%
S.suis_BM407	NC_012926	96.24%
S.suis_S735	NC_018526	96.21%
S.suis_SS12	NC_017619	96.20%
S.suis_P1	NC_012925	96.18%
S.suis_GZ1	NC_017617	96.13%
S.suis_A7	NC_017622	96.11%
S.suis_SC070731	NC_020526	96.06%
S.suis_SC84	NC_012924	96.05%
S.suis_JS14	NC_017618	96.04%
S.suis_ST1	NC_017950	96.00%
S.suis_98HAH33	NC_009443	95.94%
S.suis_05ZYH33	NC_009442	95.88%
S.suis_T15	NC_022665	95.87%
S.suis_D12	NC_017621	95.53%
S.suis_TL13	NC_021213	95.49%



3.进化关系树/Phylogenetic tree based on ANI matrix



UNIVERSITÀ CATTOLICA del Sacro Cuore

Phylogenomics

UNIVERSITÀ CATTOLICA del Sacro Cuore

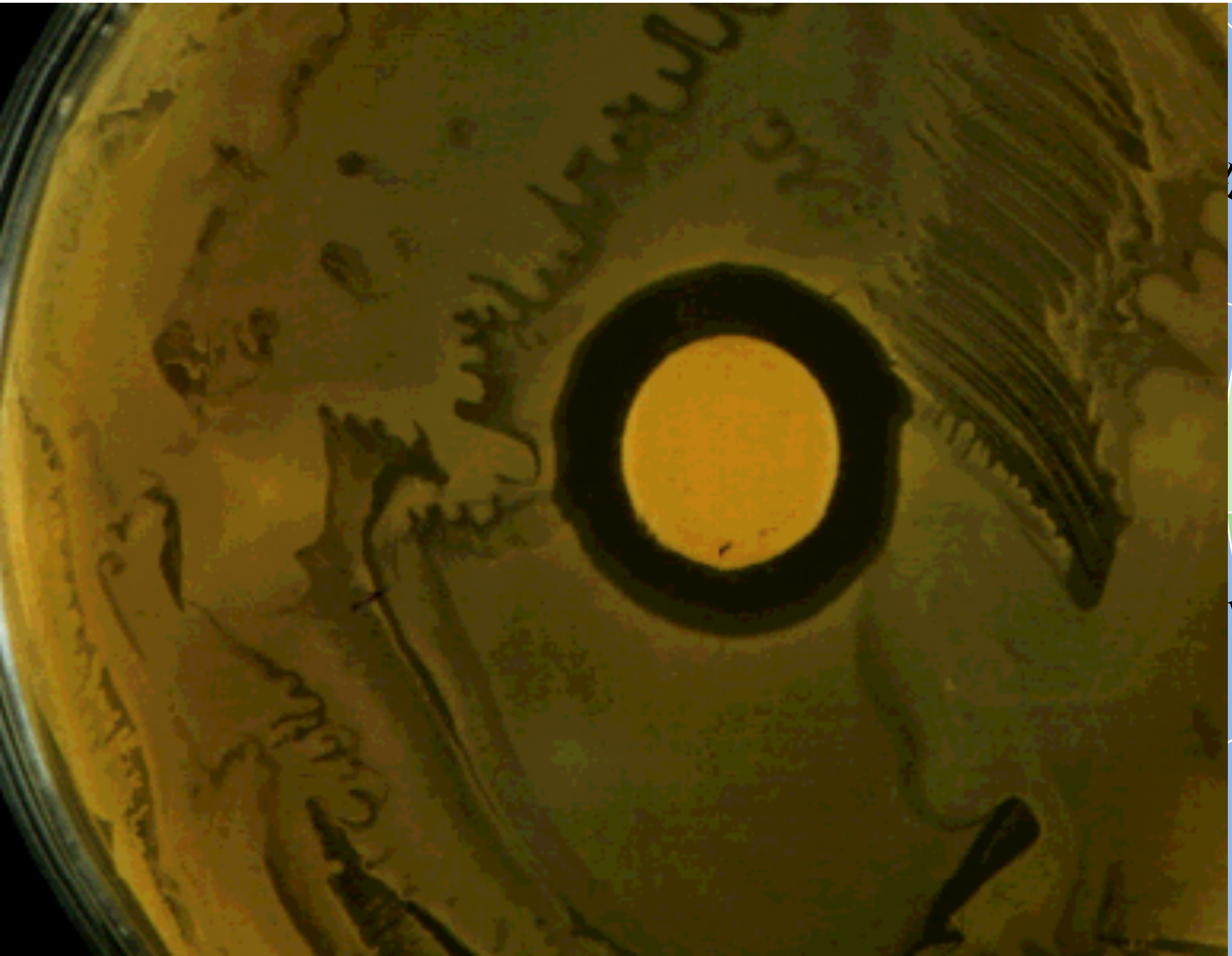
Antimicrobial Susceptibility

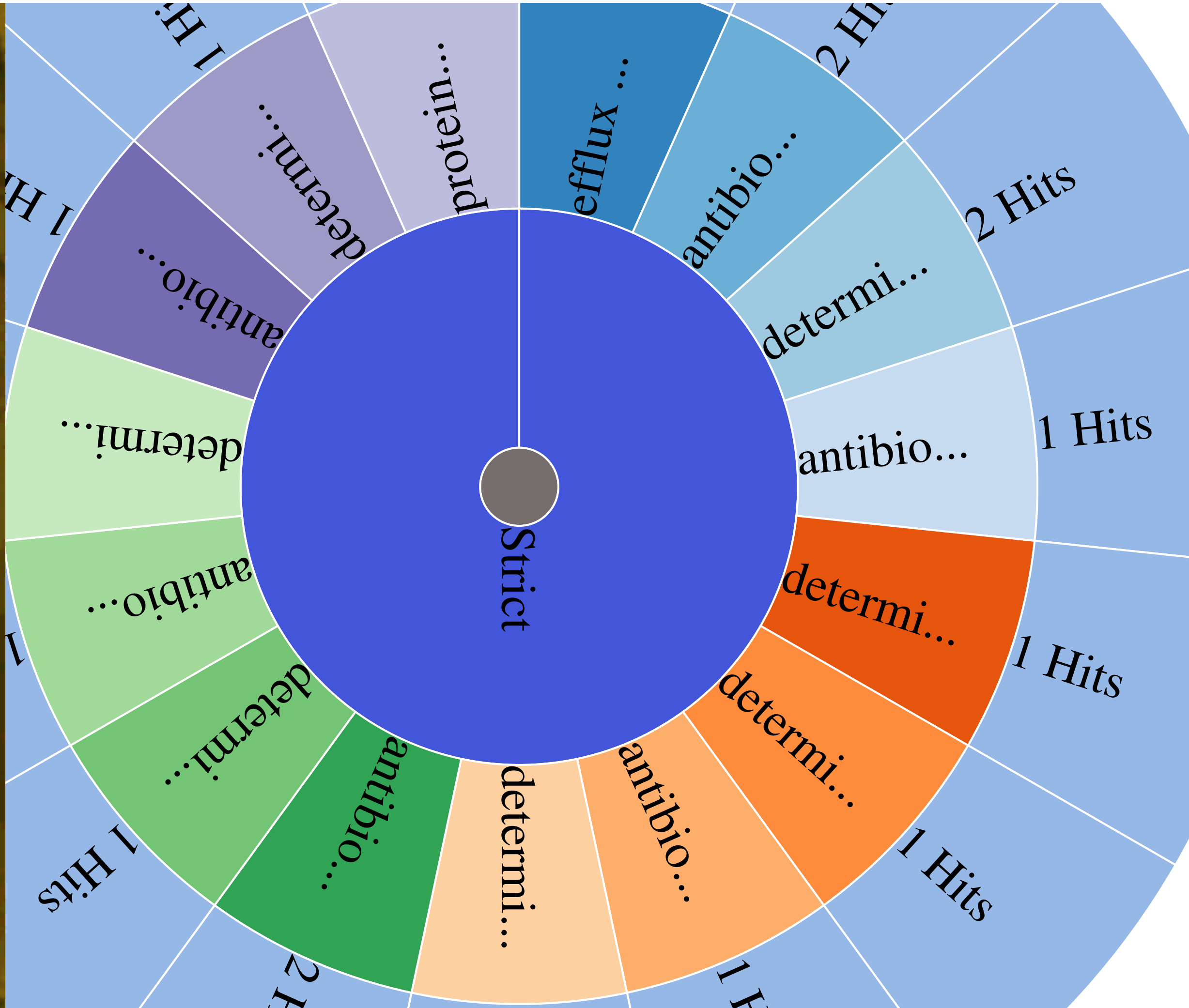
- Strains of bacteria carrying an acquired resistance to antimicrobial(s) should not be used, unless it can be demonstrated that it is a result of chromosomal mutation(s).

- Antimicrobial Resistance as a food risk

UNIVERSITÀ CATTOLICA del Sacro Cuore

Absence of acquired antimicrobial resistances in bacteria: combined used of phenotypic and genomic data





UNIVERSITÀ CATTOLICA del Sacro Cuore

AM susceptibility by WGS interrogation

- Presence of genes coding for or contributing to resistance to antimicrobials relevant to their use in humans and animals (CIAs or HIAs).
- Data focusing on complete genes coding for resistance to antimicrobials.
- Data should include at least the gene identification, function of the encoded protein, percentage of identity and e-value.

https://card.mcmaster.ca/

http://en.mediterranee-infection.com/article.php?laref=283%26titre=arg-annot

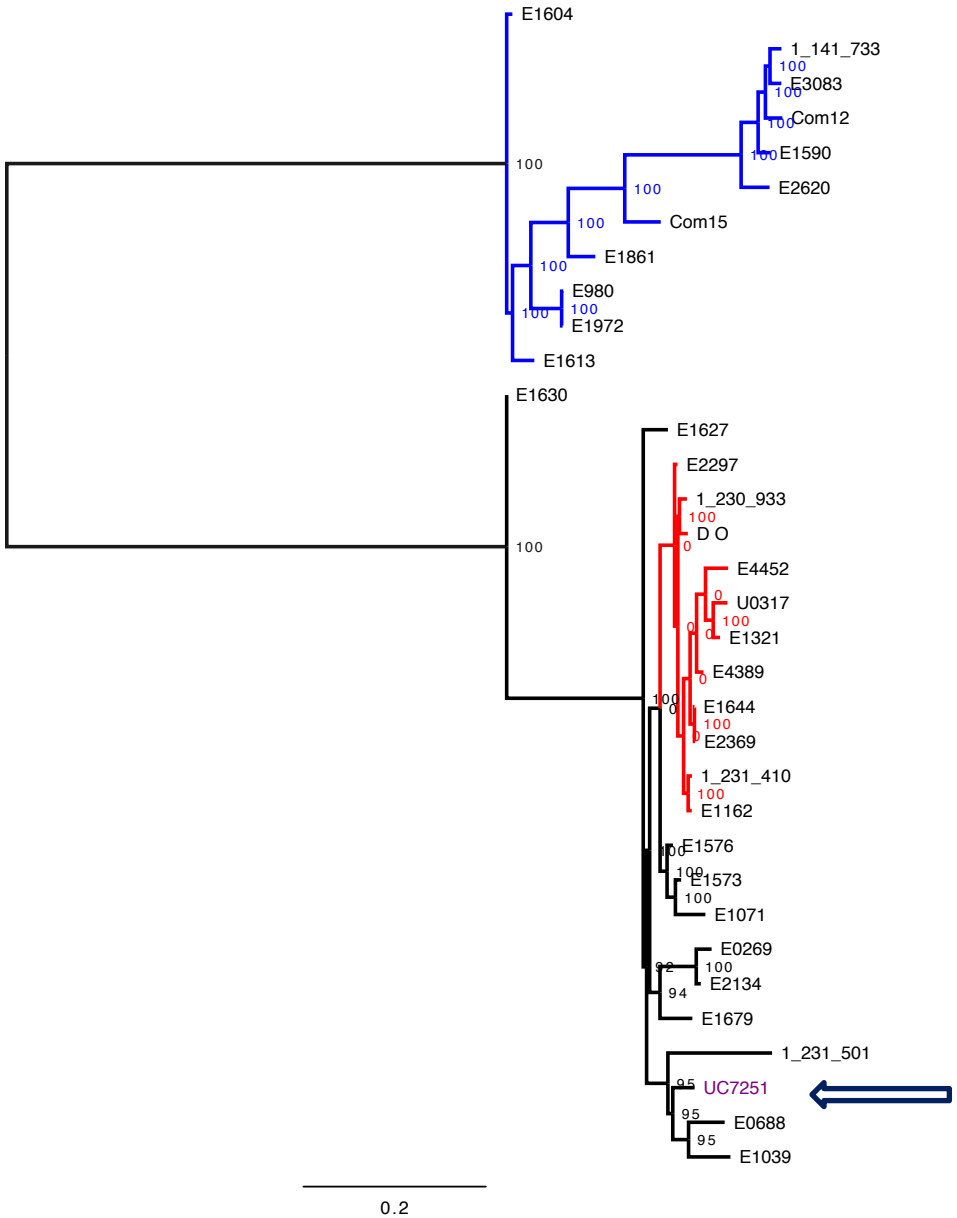
https://cge.cbs.dtu.dk/services/ResFinder/

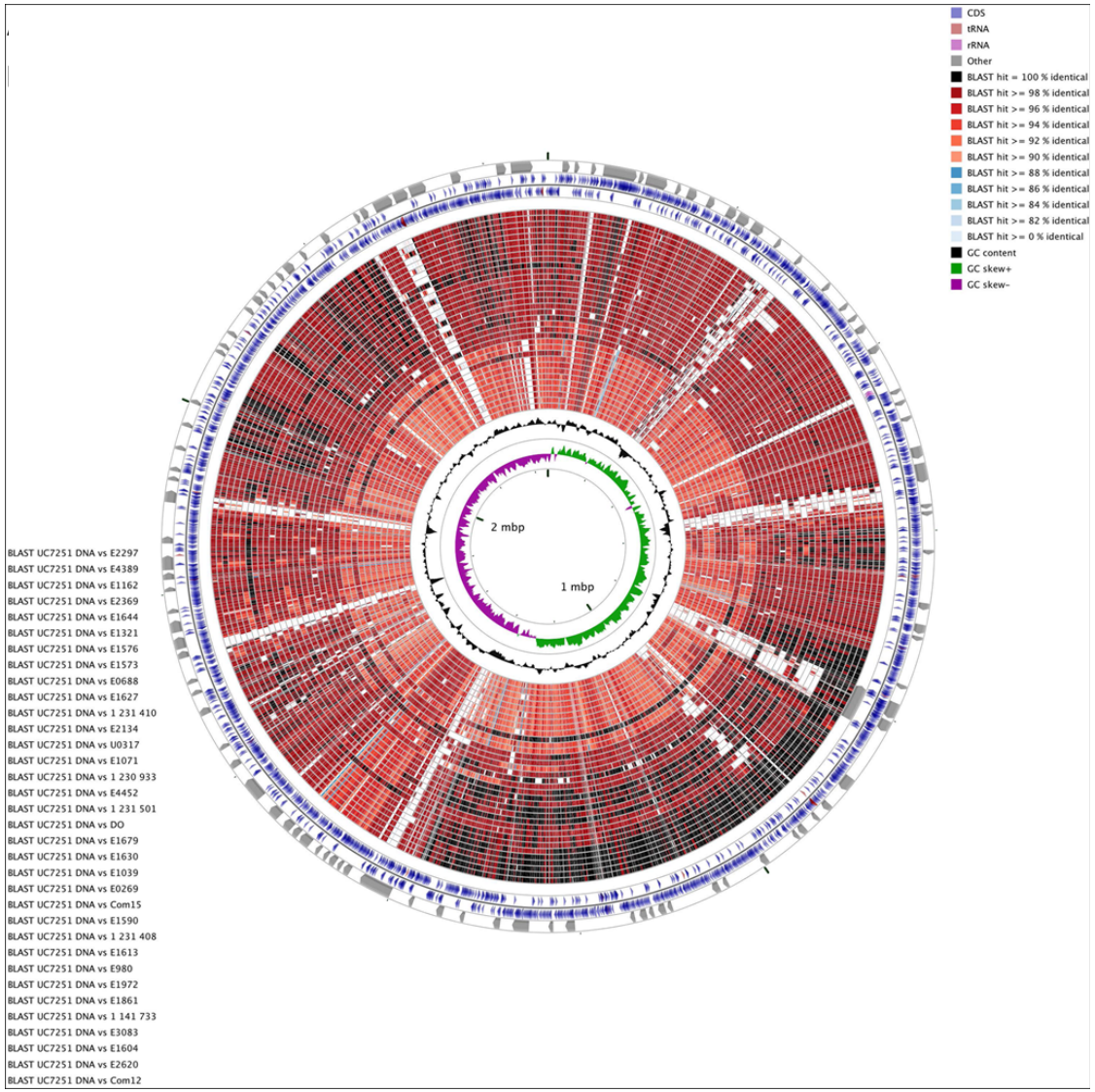
UNIVERSITÀ CATTOLICA del Sacro Cuore

Antimicrobial Resistance Gene Search

UNIVERSITÀ CATTOLICA del Sacro Cuore

Genome and phylogeny





UNIVERSITÀ CATTOLICA del Sacro Cuore

Virulence

- WGS analysis should be used to identify genes coding for known virulence factors. For this purpose, comparison against specific up-to-date databases (e.g. VFDB, PAI DB, MvirDB) should be performed.
  - Virulence factors
  - Toxins and secondary metabolites

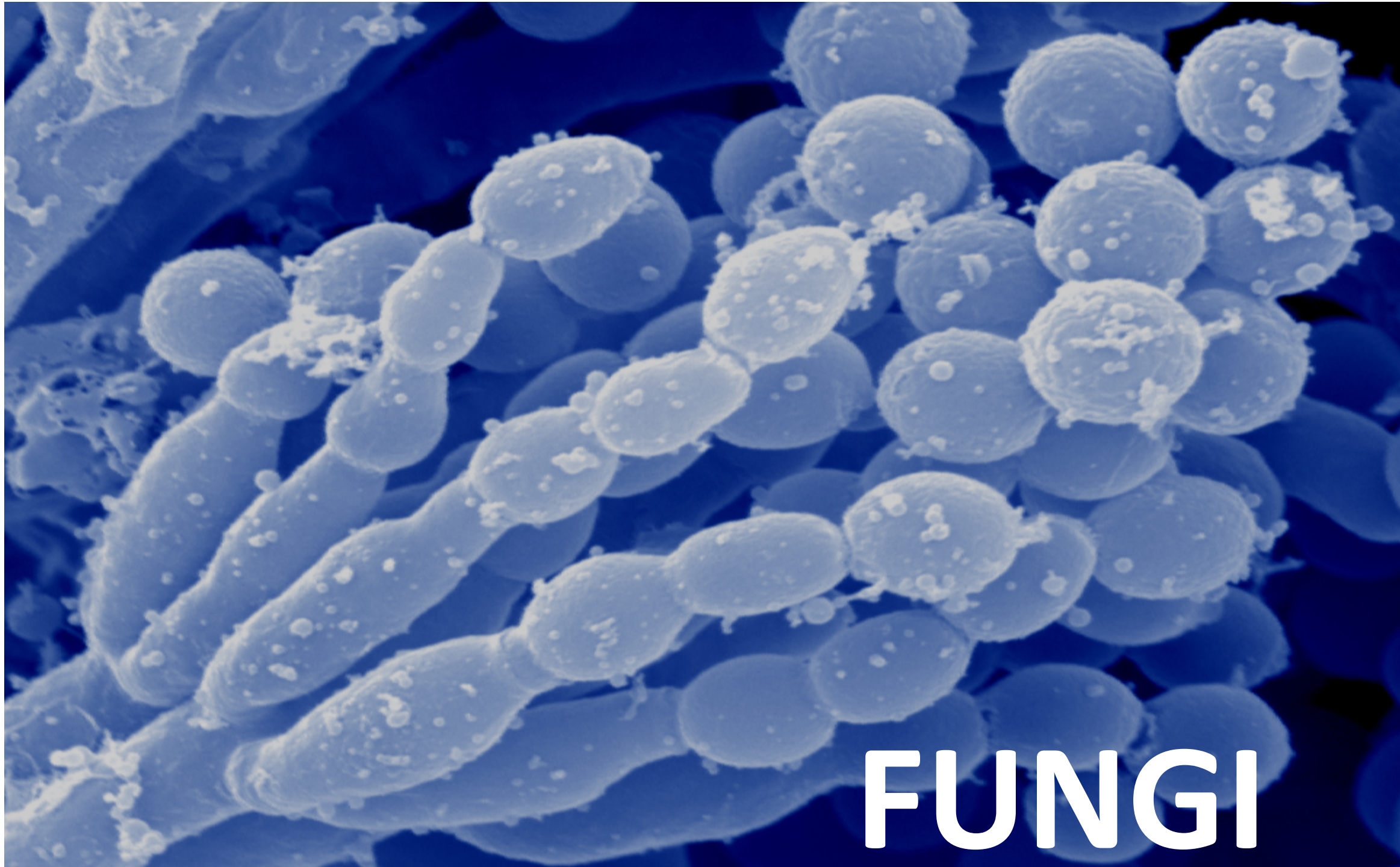
http://www.mgc.ac.cn/VFs/main.htm

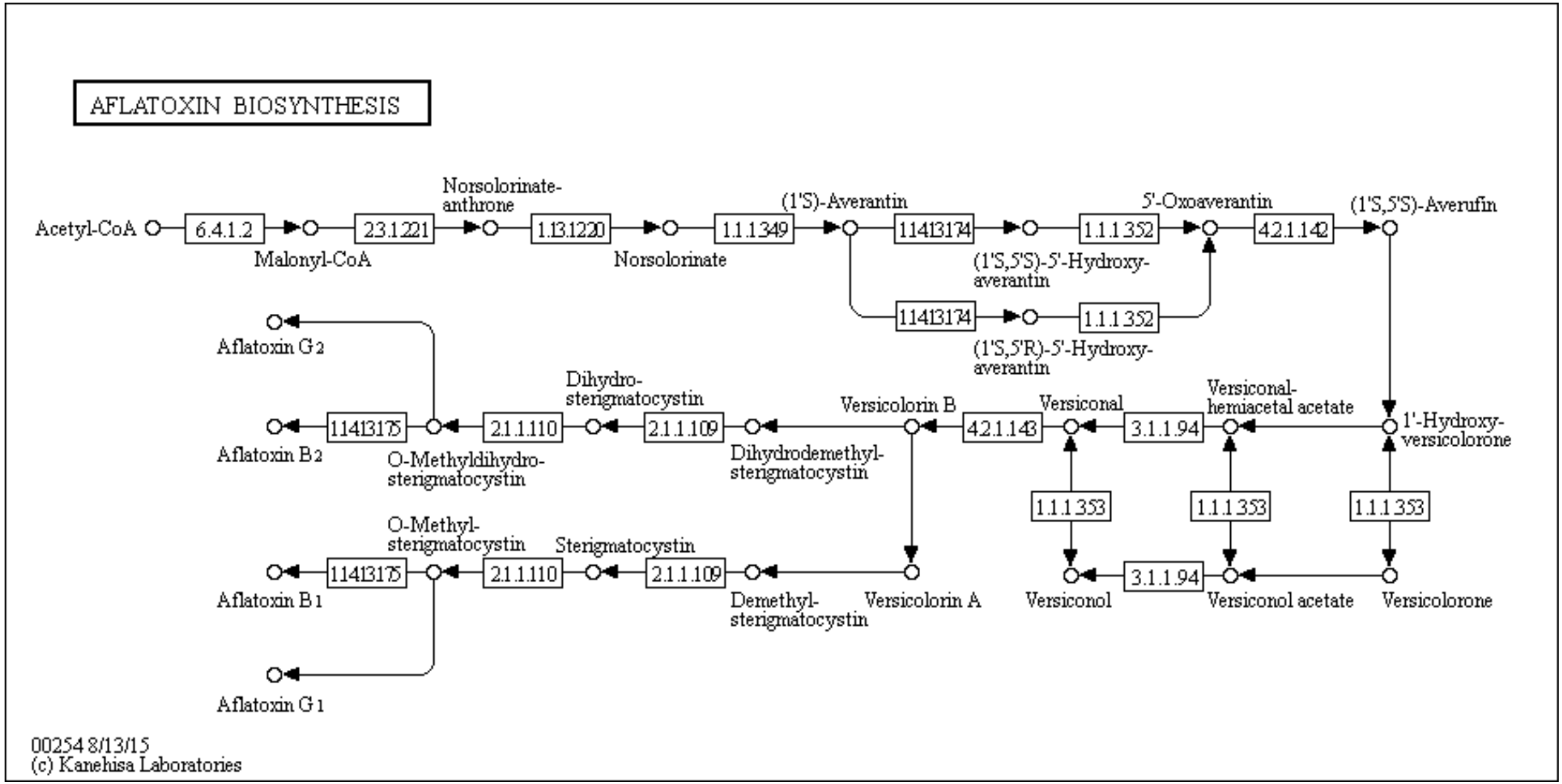
http://www.paidb.re.kr/about_paidb.php

http://mvirdb.llnl.gov

UNIVERSITÀ CATTOLICA del Sacro Cuore

Pathways reconstruction


FUNGI


AFLATOXIN BIOSYNTHESIS
Acetyl-CoA
6.4.1.2
Malonyl-CoA
2.3.1.221
Norsolorinate-anthrone
1.13.1220
Norsolorinate
1.1.1.349
(1'S)-Averantin
1.1.1.352
(1'S,5'S)-5'-Hydroxy-averantin
5'-Oxoaverantin
4.2.1.142
(1'S,5'S)-Averufin
1.1.1.352
(1'S,5'R)-5'-Hydroxy-averantin
Aflatoxin G2
Dihydro-sterigmatocystin
2.1.1.109
Versicolorin B
4.2.1.143
Versiconal
3.1.1.94
Versiconal-hemiacetal acetate
1'-Hydroxy-versicolorone
1.1.1.353
1.1.1.353
1.1.1.353
Aflatoxin B2
2.1.1.110
O-Methyldihydro-sterigmatocystin
Dihydrodemethyl-sterigmatocystin
O-Methyl-sterigmatocystin
Sterigmatocystin
2.1.1.109
Demethyl-sterigmatocystin
Versicolorin A
Versiconol
3.1.1.94
Versiconol acetate
Versicolorone
Aflatoxin B1
2.1.1.110
Aflatoxin G1
00254 8/13/15
(c) Kanehisa Laboratories

UNIVERSITÀ CATTOLICA del Sacro Cuore

WGS and MICs

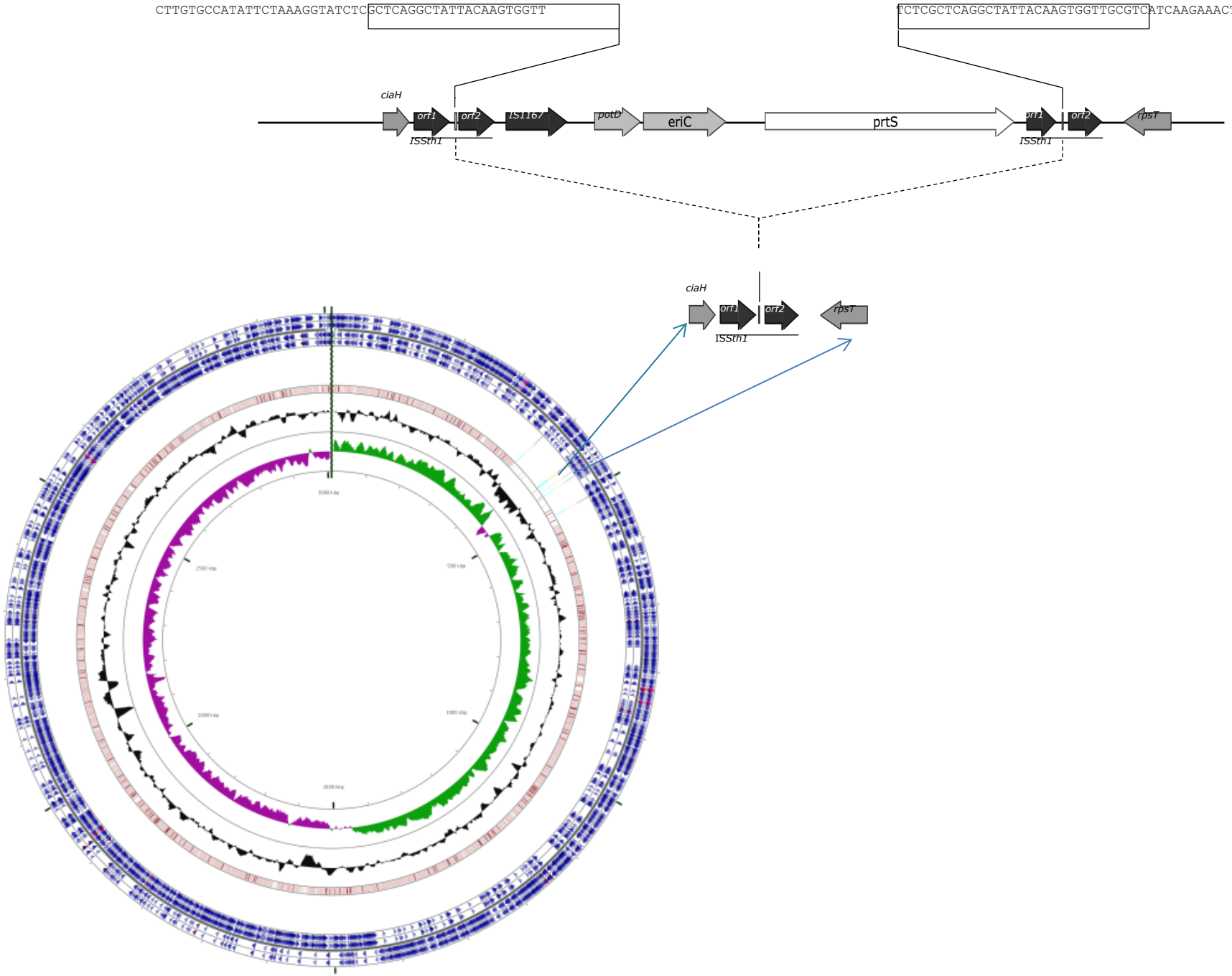
- MIC > cut-off value for one or more antimicrobials requires further investigation using genomic data :
  - If no known AMR gene is identified that can be linked to the phenotype: OK
  - If the phenotypic resistance related to the presence of a known AMR gene: HAZARD
- If the genetic analysis reveals AMR genes for antimicrobials considered to be CIAs or HIAs (WHO, 2016), the MIC values should be determined and compared with values in the literature:
  - If MIC≤ (reference values), the likelihood of the AMR gene to become active should be assessed (e.g. based on sequence comparison with active genes)
  - If MIC> (reference values), this is considered as a hazard.

UNIVERSITÀ CATTOLICA del Sacro Cuore

GMM: assessment of genetic modification by WGS analysis

- Detailed information, including a map or graphic presentation of all genomic regions (chromosome, contig or plasmid) harbouring genetic modifications, indicating:
- the open reading frames (ORF) actually inserted, modified or deleted. the non-coding sequence(s) inserted/deleted/modified. The role and function of these sequences (e.g. promoters, terminators) should be indicated
- Comparison the WGS of the GMM with that of the non-modified parental or recipient strain.

UNIVERSITÀ CATTOLICA del Sacro Cuore



UNIVERSITÀ
CATTOLICA
del Sacro Cuore

Conclusions