

Part 3: Metagenomics and its applications to identification of antibiotic resistance

MVE 360 – Bioinformatics, 2012

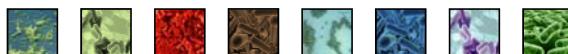
Erik Kristiansson, erik.kristiansson@chalmers.se

Agenda

- Part 3: Metagenomics
 - Metagenomics
 - Binning and annotation
 - Statistical Analysis
 - Applications to detection of antibiotic resistance
 - Case study: Antibiotic pollution
 - Case study: Antibiotic resistance genes in the human gut microbiome

Some facts about microbes

Number of microbes on Earth	5×10^{30}
Number of microbes in all humans	6×10^{23}
Number of stars in the universe	7×10^{21}
Number of bacterial cells in one human gut	10^{14}
Number of human cells in one human	10^{13}
Number of bacterial genes in one human gut	3,000,000
Number of genes in the human genome	21,000

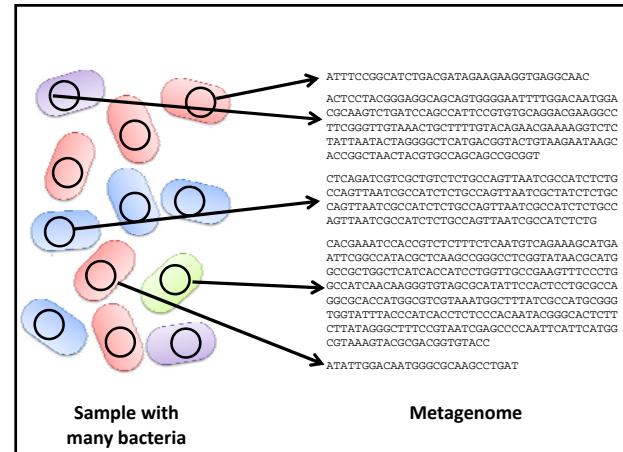
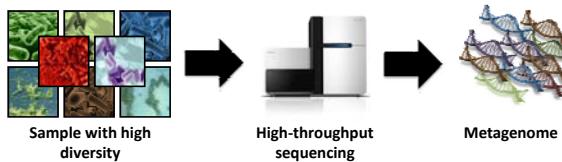


Microbial diversity

- Bacteria are ubiquitous in every habitat on Earth
- There are up to 100 million bacterial species – only a small fraction of these are known
- More than 99% of all bacteria are uncultivable and can not be grown in laboratories

Shotgun metagenomics

- Analysis of all DNA in a sample – *the metagenome*



The advantages of metagenomics

- An entire community of microorganisms can be studied as a whole.
- No cultivation of individual strains is required,
 - ~ 50% of human gut bacteria are unculturable,
 - > 99% of environmental bacteria are unculturable.
- Efficient,
 - Vast amount of nucleotides can be read fast and cheap using next generation DNA sequencing.

How much data do we need?

- 1 gram of soil
 - Can contain up to 100,000 different bacteria
 - Can contain up to 100,000,000,000,000 bases (100 terabases)
- The total amount of metagenomic data today is less than the DNA content in a liter of ocean water



Metagenomic data

- Metagenomes are often heavily undersampled
- Fragmented and error prone
 - Short read length (50-700 bases)
 - Error prone (usually > 1% error rate)
- Lack of reference data
 - There exists millions of bacterial species but less than 10.000 have been sequenced so far.

Metagenomics

Qualitative metagenomics

- Descriptive analysis of species
- Identification of novel genes

Quantitative (comparative) metagenomics

- Analysis of differences between metagenomes
 - Taxonomy (species)
 - Function (genes)

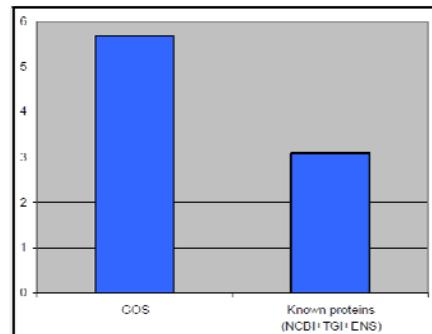
The Global Ocean Sampling

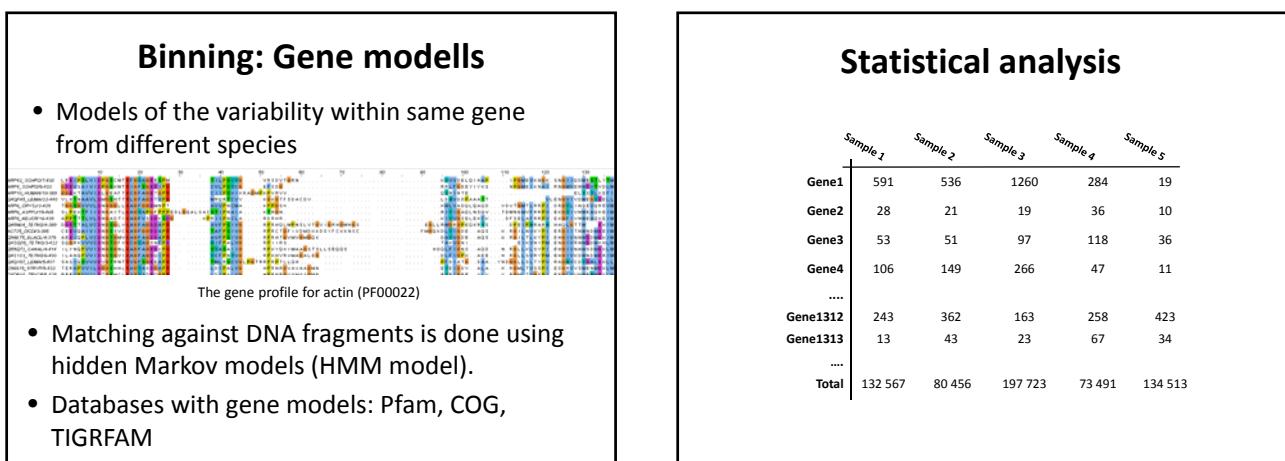
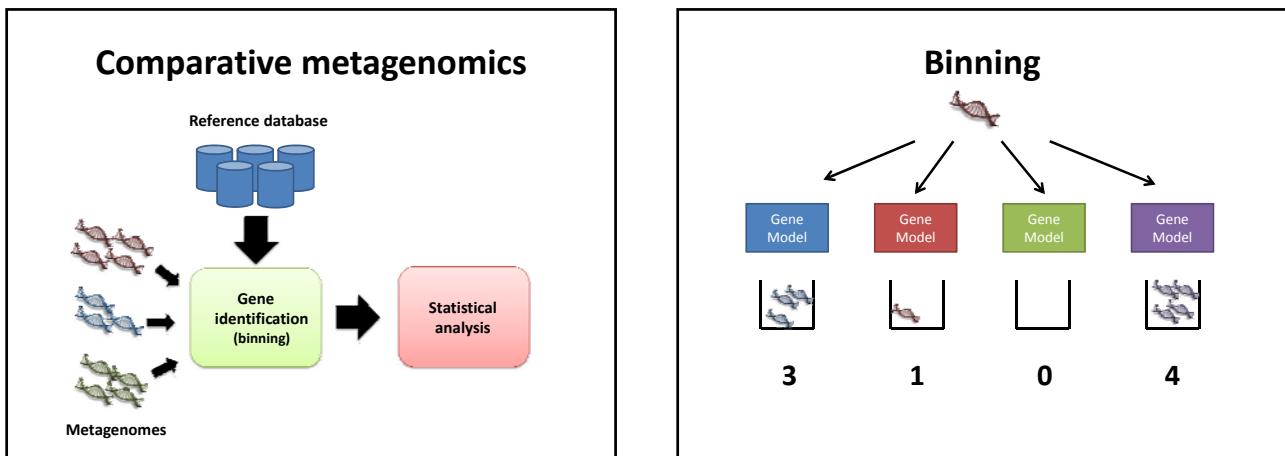
- Investigating microbial diversity in the worlds ocean
- A sailing boat equipped with a sequencer



<http://www.jcvi.org/cms/research/projects/gos/overview/>

The Global Ocean Sampling





ShotgunFunctionalizeR
- an R-package for comparative metagenomics

- A statistical toolbox for comparative metagenomics
- Includes methods for data exploration and estimation of biodiversity
- Multiple annotations (COG, Pfam, etc)
- Available at <http://shotgun.math.chalmers.se>

Kristiansson, E., Hugenholz, P., Dalev, D. (2009). ShotgunFunctionalizeR – an R-package for functional analysis of metagenomes. Bioinformatics 25(20).
Kristiansson, E., Bouland, F., Jonsson, V., Pereira, M., Johnnig, A., et al. (2012). Analysis of antibiotic resistance genes in uncultured bacterial communities. In preparation.

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Gene1	591	536	1260	284	19
Gene2	28	21	19	36	10
Gene3	53	51	97	118	36
Gene4	106	149	266	47	11
...					
Gene1312	243	362	163	258	423
Gene1313	13	43	23	67	34
...					
Total	132 567	80 456	197 723	73 491	134 513

Estimate abundance

$i \downarrow$

$j \rightarrow$

$Y_{i,j}$

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Gene1	591	536	1260	284	19
Gene2	28	21	19	36	10
Gene3	53	51	97	118	36
Gene4	106	149	266	47	11
...					
Gene1312	243	362	163	258	423
Gene1313	13	43	23	67	34
...					
Total	132 567	80 456	197 723	73 491	134 513

The Poisson model

- Generalized linear model (log-link)
- Regression with the relative abundance as the dependent variable

$$\log E[Y_{i,j}] = \log(\text{OFFSET}_j) + \sum \alpha_k y_{j,k}$$

Offset Covariates

The Poisson model

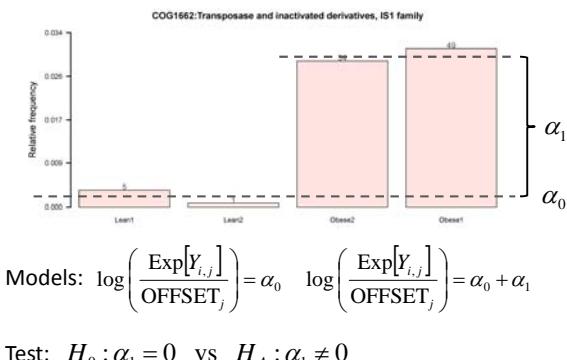
- Generalized linear model (log-link)
- Regression with the relative abundance as the dependent variable

$$\log\left(\frac{\text{Exp}[Y_{i,j}]}{\text{OFFSET}_j}\right) = \sum \alpha_k x_{j,k}$$

Covariates

- Possible offsets: (i) Total reads, (ii) Total amount of bacterial DNA, (iii) A reference gene

Comparison of two groups



The Poisson model

- Metagenomic data is overdispersed!
-
- $\text{Exp}[X] = \lambda$
- $\text{Var}[X] = \phi\lambda$
- Multiple testing correction: Benjamini-Hochberg's false discovery rate (FDR).

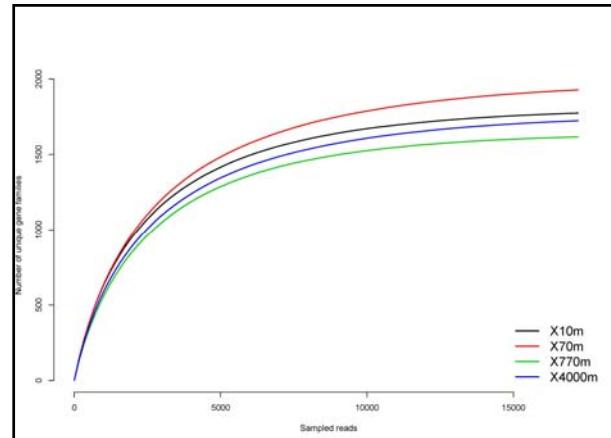
Estimation of biodiversity

- Rarefaction curve analysis
 - Estimation of diversity through resampling

N = Total number of observations (DNA fragments)
 M = Total number of bins, each have N_i observations

Algorithm

- Sample n items from the N observations
- Calculate the number of unique bins/genes
- Repeat for different values of n



Summary metagenomics

- Metagenomics is a way to study bacteria in their community without cultivation
- DNA sequencing is the main bottleneck for metagenomics
- Competence in DNA sequence analysis is essential for interpretation of metagenomic data

Case study 1: Antibiotic pollution

1. Promotion of resistance genes?
2. Transfer of resistance genes to human pathogens?



The History of Antibiotics

- Commercially introduced 1932
- Mass-production and mass-consumption started in the 40's
- Bacteria with acquired resistance discovered a few years later



Alexander Fleming



Penicillin-producing fungi

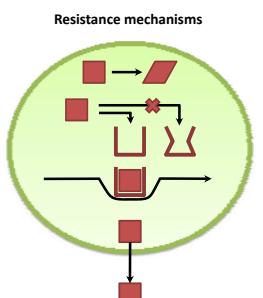
- Antibiotic resistance is caused by
 1. Mutations in pre-existing DNA
 2. Acquisition of resistance genes



It was on a short-cut through the hospital kitchens that Albert was first approached by a member of the Antibiotic Resistance.

Acquired Antibiotic Resistance

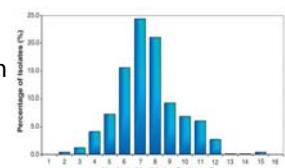
- Common mechanisms
 - Drug inactivation
 - Alteration of target site
 - Alteration of pathways
 - Efflux pumps
 - Changes in gene expression
- Causes
 - Chromosomal mutations
 - Mobile genes
- Mobility
 - Horizontal gene transfer



Picture by Anna Johnning, Gothenburg University

The resistome

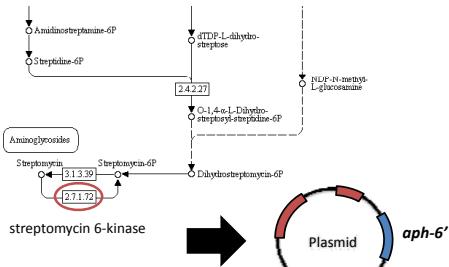
- The resistome contains
 - Thousands of mobile resistance genes
 - Resistance mechanisms for many clinically important antibiotics
- Environmental bacteria contain a large collection of resistance genes



D'Costa et al. (2006). Sampling the antibiotic resistome. Science 311 pp. 374-7

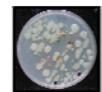
Example: Streptomycin resistance

- Produced naturally by *Streptomyces* sp.



Characterization of antibiotic resistance

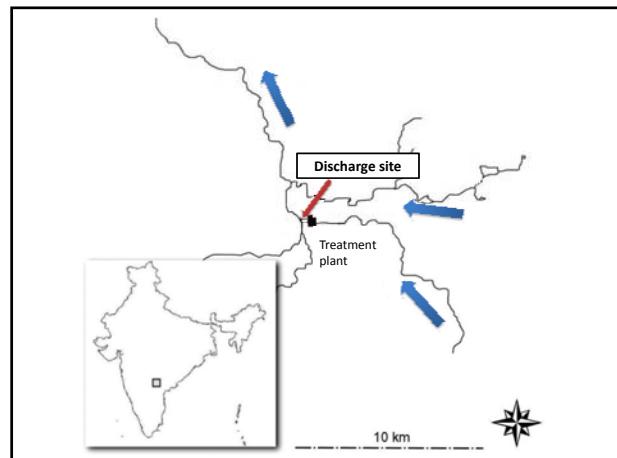
- Traditional approach:
 - Isolation of individual strains
 - Phenotypic and genotypic screening



- Not all bacteria can be cultivated
 - Human gut: 60 % unculturables
 - Soil: 99 % unculturables (!!)

- A vast number of resistance genes
 - PCR not time- and cost-efficient

Hugenholtz et al. Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity. *J. Bacteriol.* 180(18) 1998
Bapné & Giovannoni. The Uncultured Microbial Majority. *Annu. Rev. Microbiol.* 57 2003

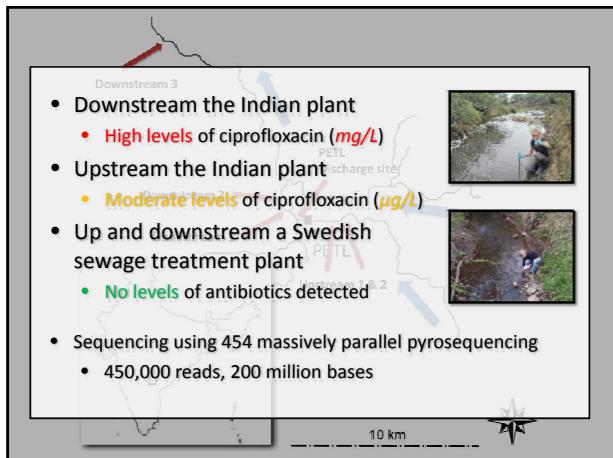


Antibiotic pollution

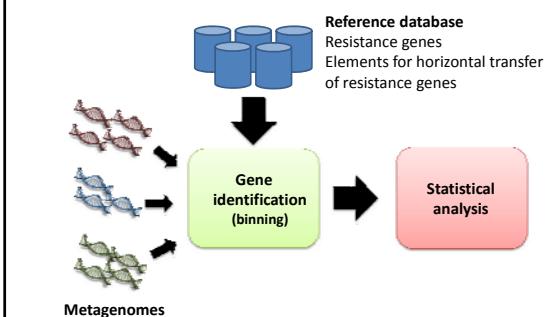
Active ingredient	Type of drug	Range ($\mu\text{g/L}$)
Ciprofloxacin	antibiotic-fluoroquinolone	28,000-31,000
Losartan	angiotensin II receptor antagonist	2,400-2,500
Cetirizine	H ₁ -receptor antagonist	1,300-1,400
Metoprolol	β_1 -adrenoreceptor antagonist	800-950
Enrofloxacin	antibiotic-fluoroquinolone (veterinary use)	780-900
Citalopram	serotonin reuptake inhibitor	770-840
Norfloxacin	antibiotic-fluoroquinolone	390-420
Lomefloxacin	antibiotic-fluoroquinolone	150-300
Enoxacin	antibiotic-fluoroquinolone	150-300
Oflloxacin	antibiotic-fluoroquinolone	150-160
Ranitidin	H ₂ -receptor antagonist	90-160

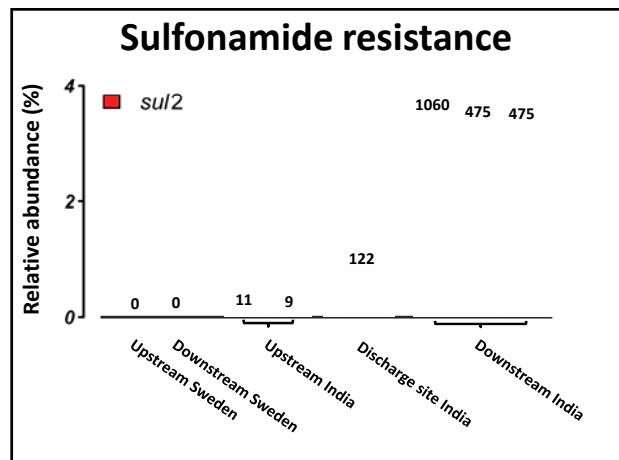
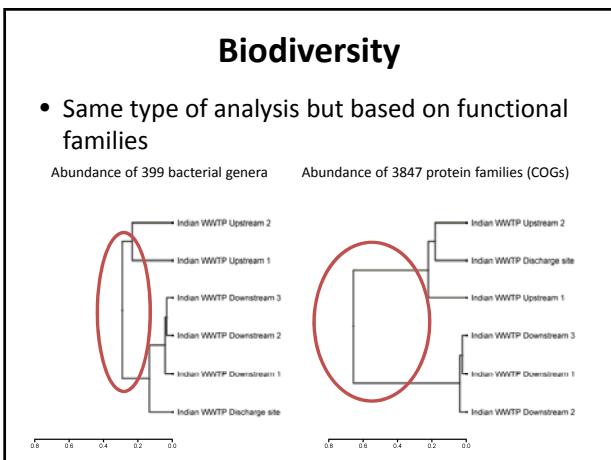
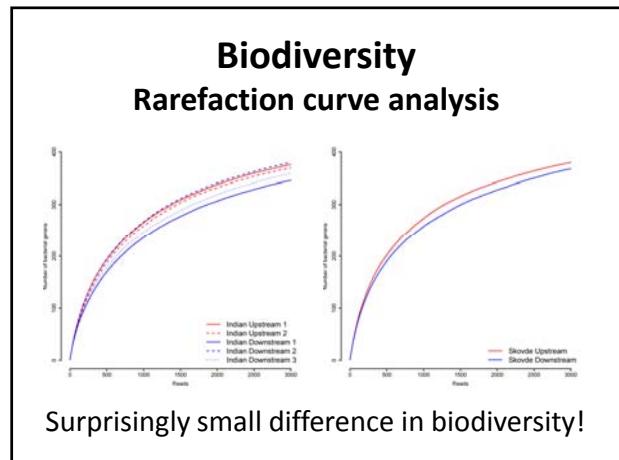
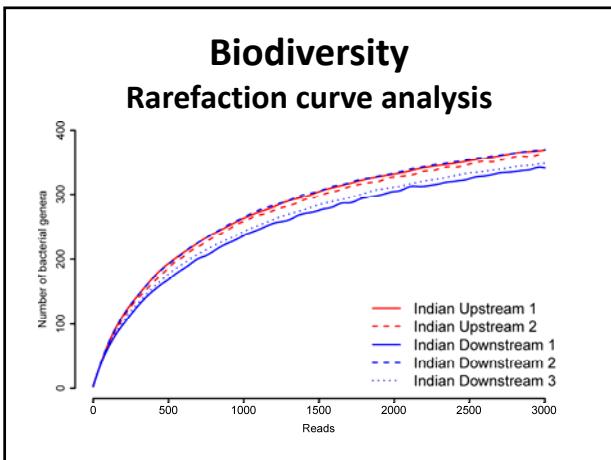
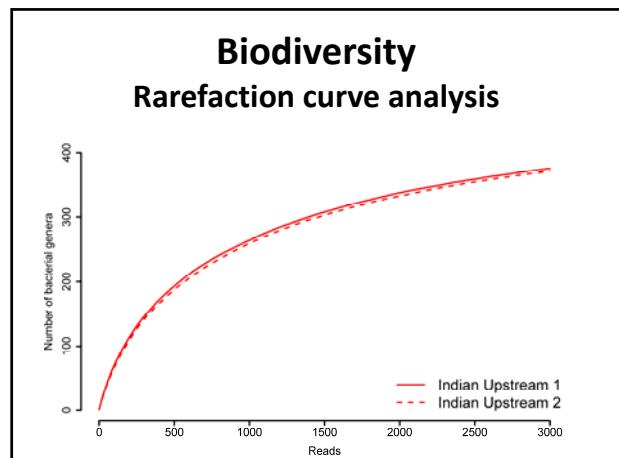
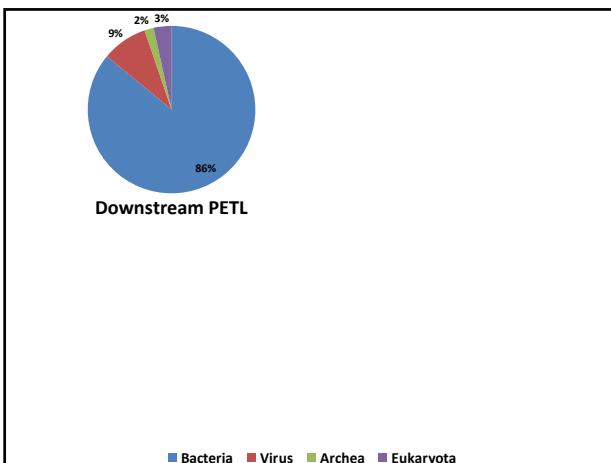
Larsson et al. (2007). Effluent from drug manufactures contains extremely high levels of pharmaceuticals. *J. of Hazard Mater.* 3 (30).

Erick et al. (2009). Contamination of surface, ground, and drinking water from pharmaceutical production. *Environ. Toxicol. and Chem.* 28 (12).

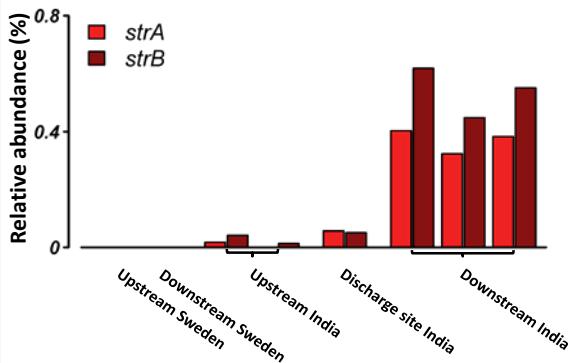


Comparative metagenomics





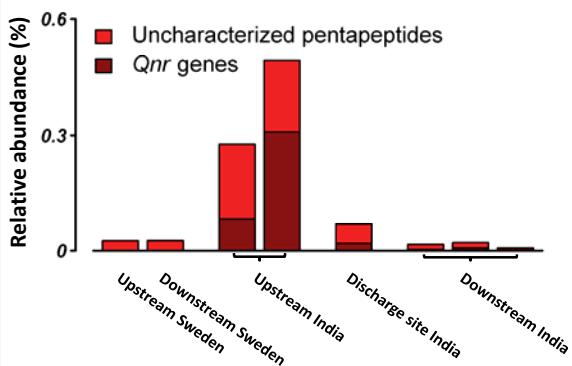
Aminoglycoside resistance



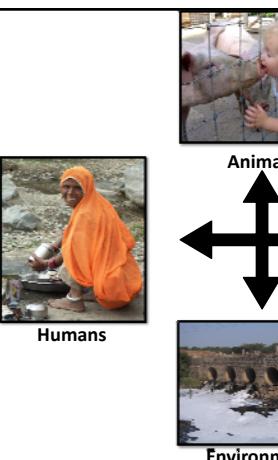
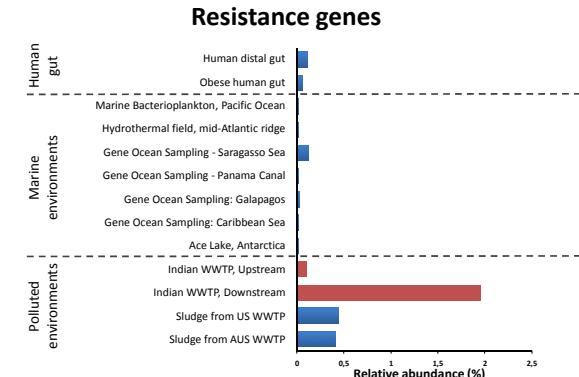
Mobile fluoroquinolone resistance

- Fluoroquinolones inhibit topoisomerase
-
- DNA
Qnr
- *QNR*-genes have a DNA-like structure that protects topoisomerase from fluoroquinolones.

Fluoroquinolone resistance

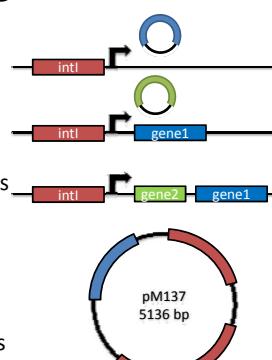


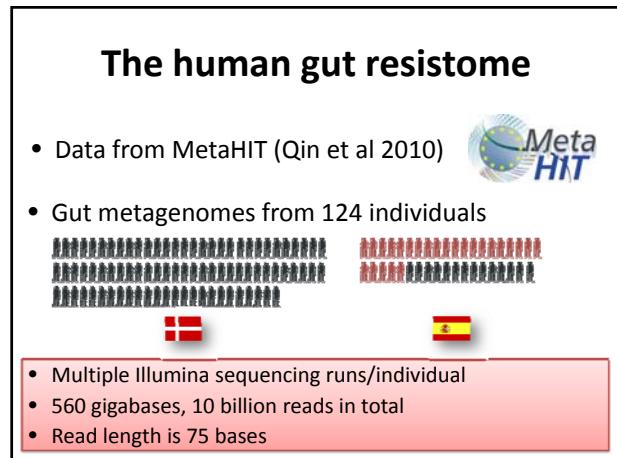
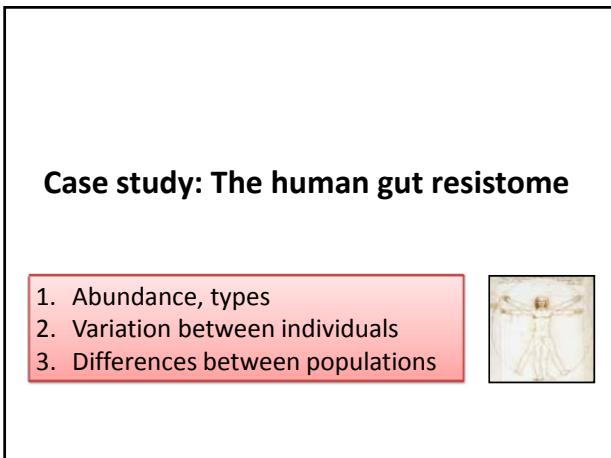
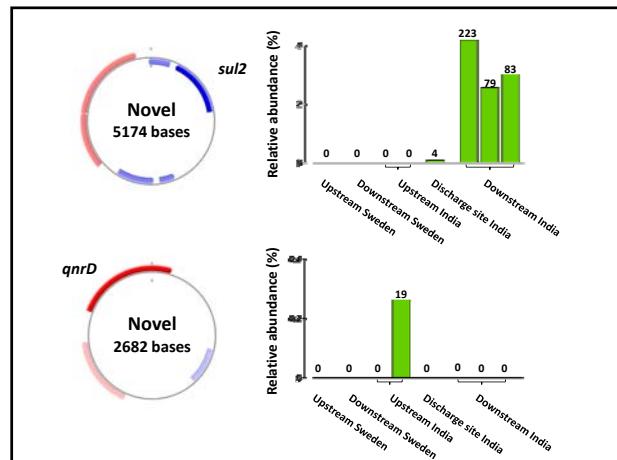
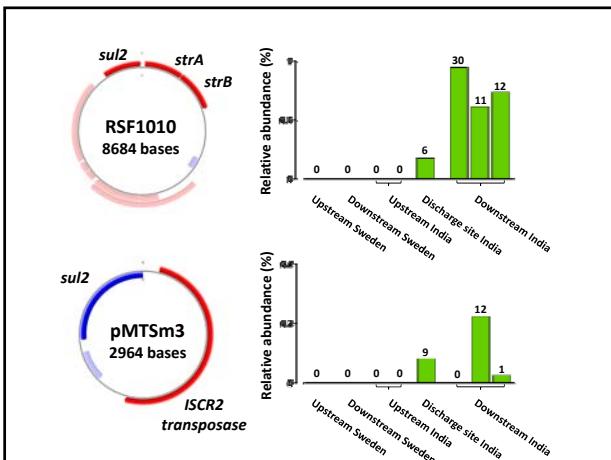
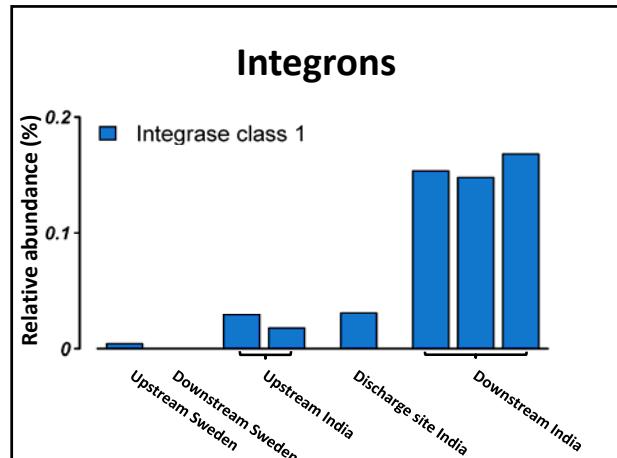
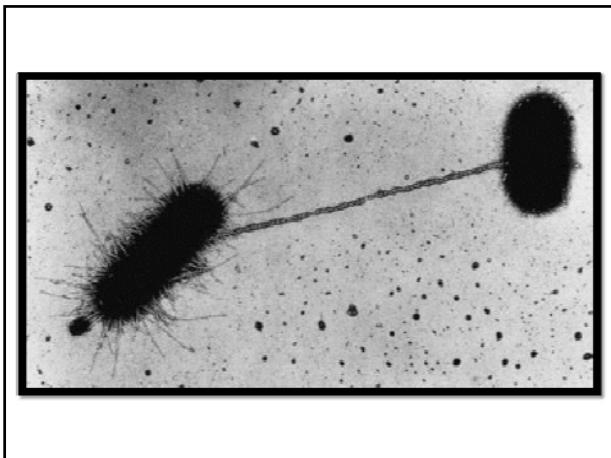
Perspective – How extreme is “extreme”?



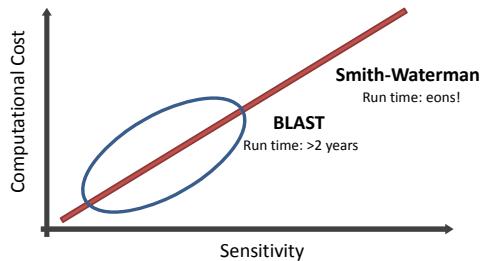
Horizontal gene transfer

- Integrons
 - Single genes
- Transposons
 - Small to medium regions
- Plasmids
 - Intercellular transfer of genetic material
 - Medium to large regions



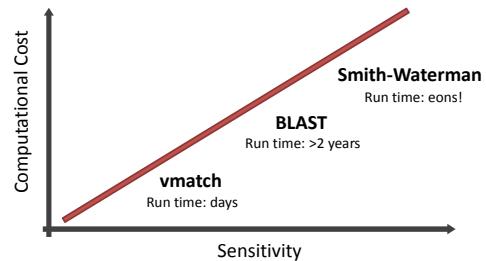


Annotation by sequence similarity



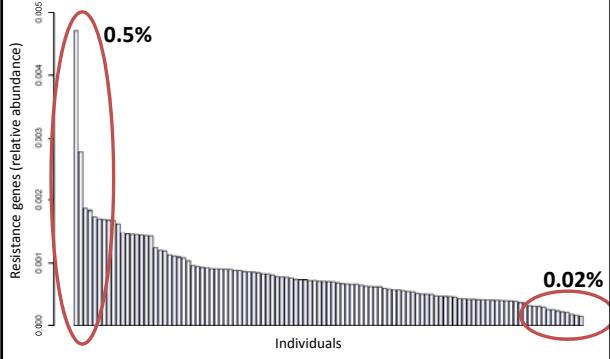
- Resistance genes are remarkably conserved!

Annotation by sequence similarity



- Resistance genes are remarkably conserved!

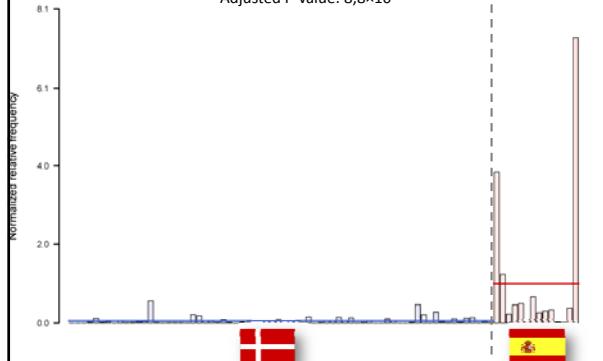
Overview of the gut resistome



Chloramphenicol resistance

Chloramphenicol Acetyltransferase (class A)

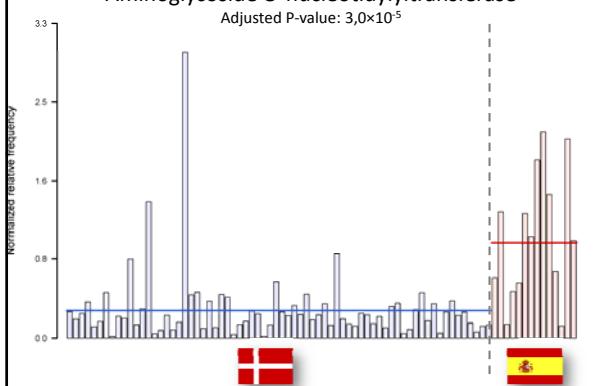
Adjusted P-value: $8,8 \times 10^{-10}$



Aminoglycoside resistance

Aminoglycoside O-nucleotidyltransferase

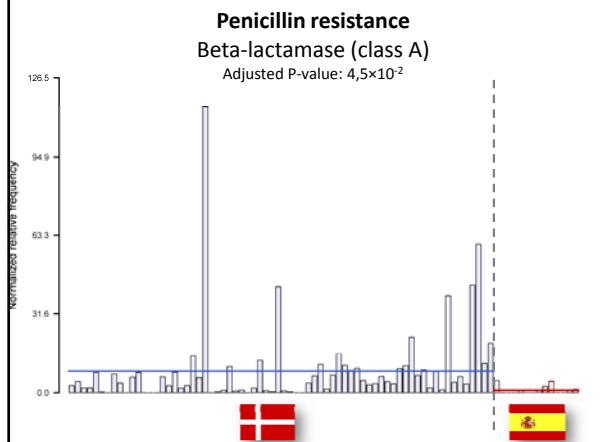
Adjusted P-value: $3,0 \times 10^{-5}$



Penicillin resistance

Beta-lactamase (class A)

Adjusted P-value: $4,5 \times 10^{-2}$



Summary – Antibiotic resistance

- Antibiotic resistance genes are ubiquitous
- Metagenomics provides a powerful way to
 - Identify resistance genes in unculturable bacteria
 - Identify mechanisms for their horizontal transfer
- Antibiotic pollution results in increased levels of resistance genes in the environment
- The risks for transfer to pathogens is not known

Summer Projects 2012

- Stipends for summer projects in bioinformatics
- Topics in metagenomics and DNA sequence analysis
- Interested? Contact me at erik.kristiansson@chalmers.se

