SGMJ2019, March, Tokyo

# Population genomic studies of microbial recombination, phylogeny, and population structure

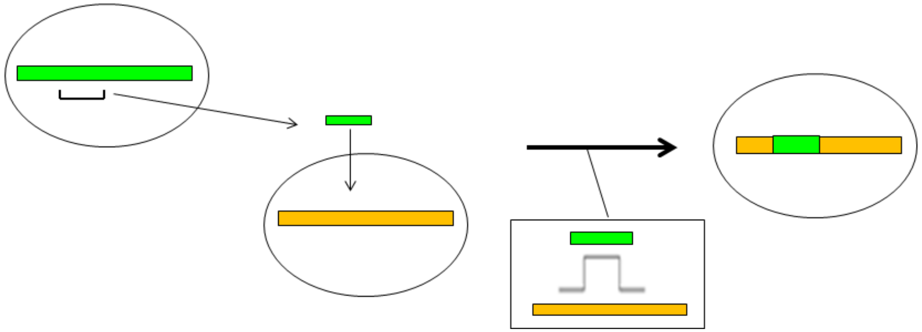
Koji Yahara

Senior Investigator

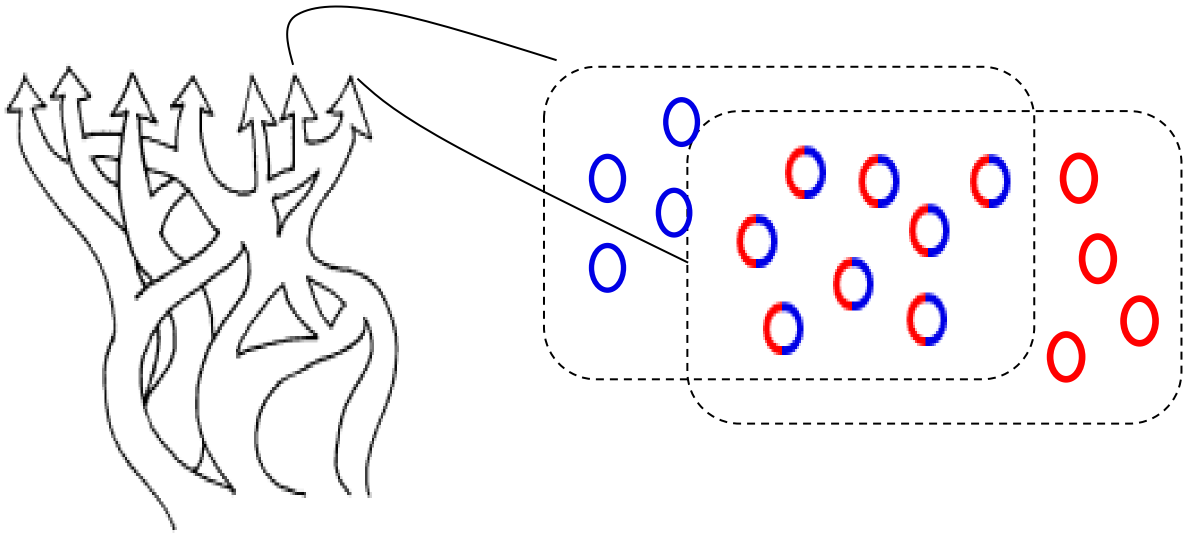
National Institute of Infectious Diseases

# Two main themes

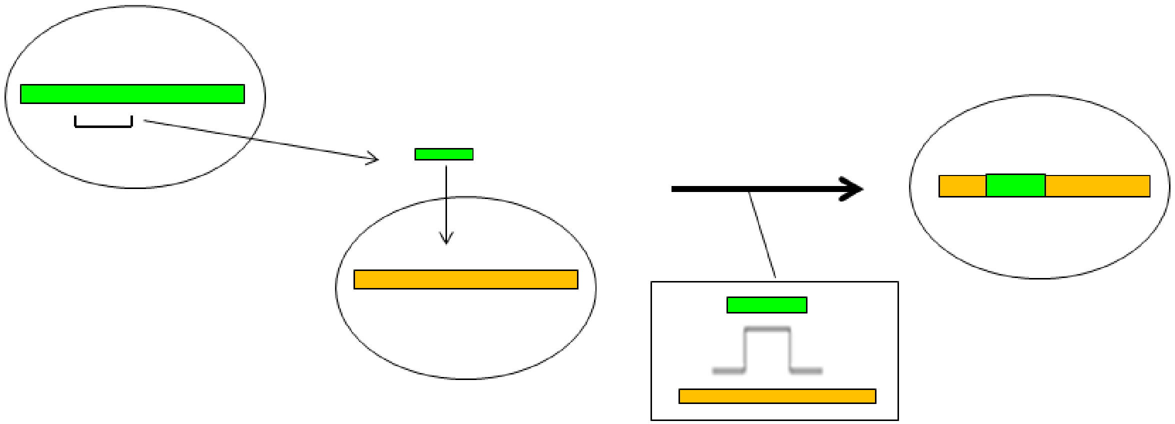
- Recombination



- Phylogeny and population structure

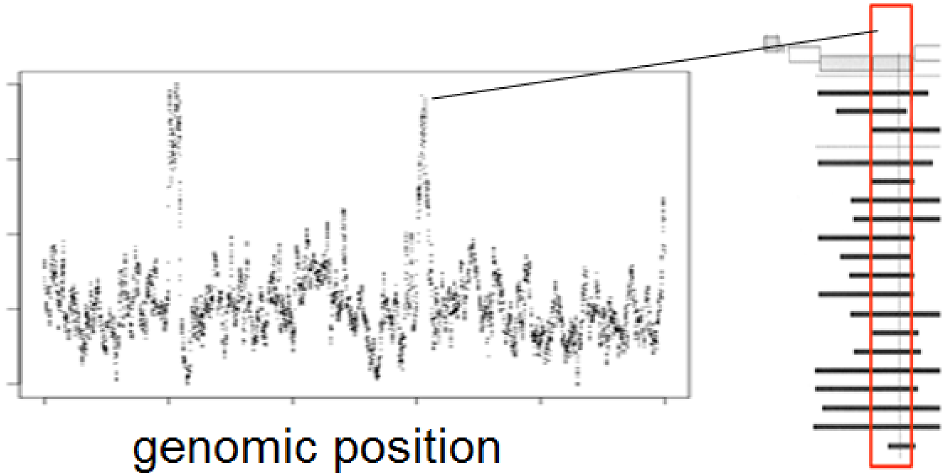


# Recombination



driving force of evolution

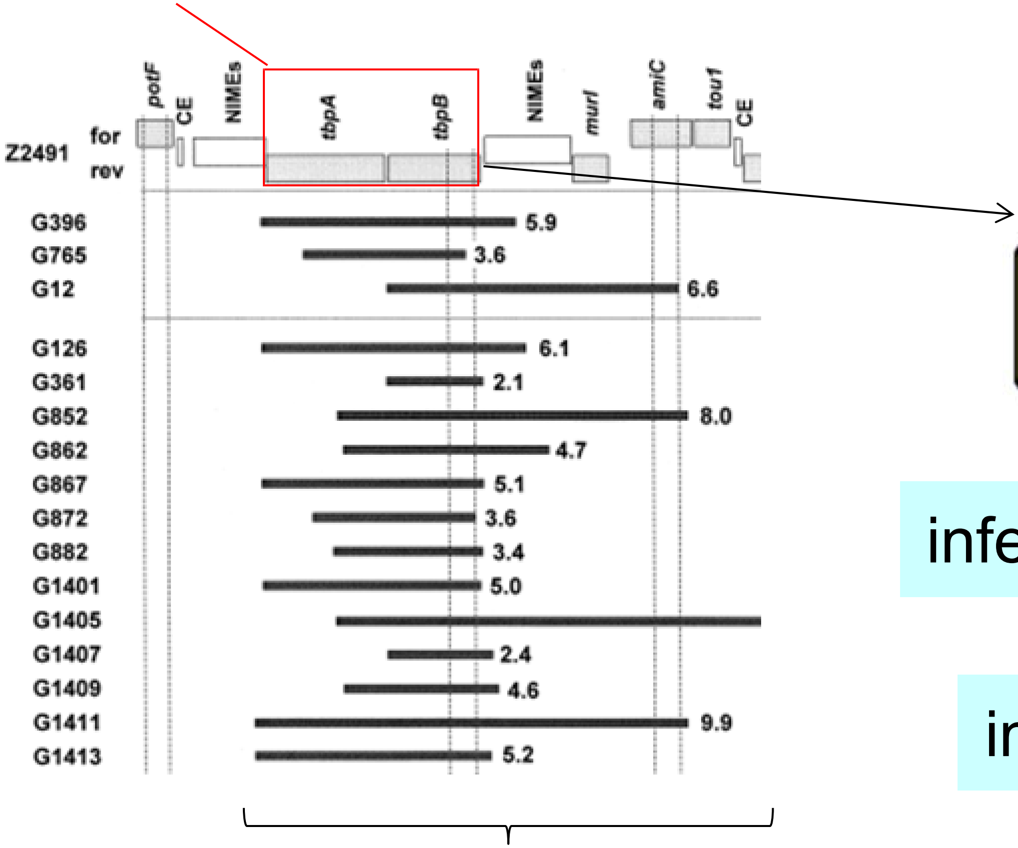
Variation of intensity of recombination across a genome?

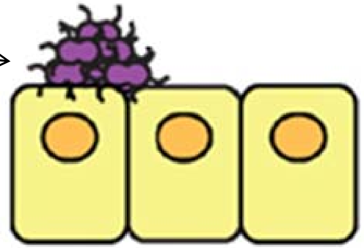



recombination-hot region

# For example

- *tbpB*, an outer membrane protein in *Neisseria*




infection & adaptation

immune selection

Imported DNA fragments

Linz (2000), Microbiology

**Largely unknown in most species**

# Recombination

Method

Relation to
diversifying selection

Application to
10 bacterial species

Bacteriophage
(virome)

# Based on “chromosome painting”

- reconstructs a ‘recipient‘ haplotype as recombination-derived mosaic of all the other donors


Lawson et al (2012), *PLoS Genetics*, for human
→ Yahara et al (2013), *MBE*, applied to bacteria

- Limitation: only for the most recent recombination

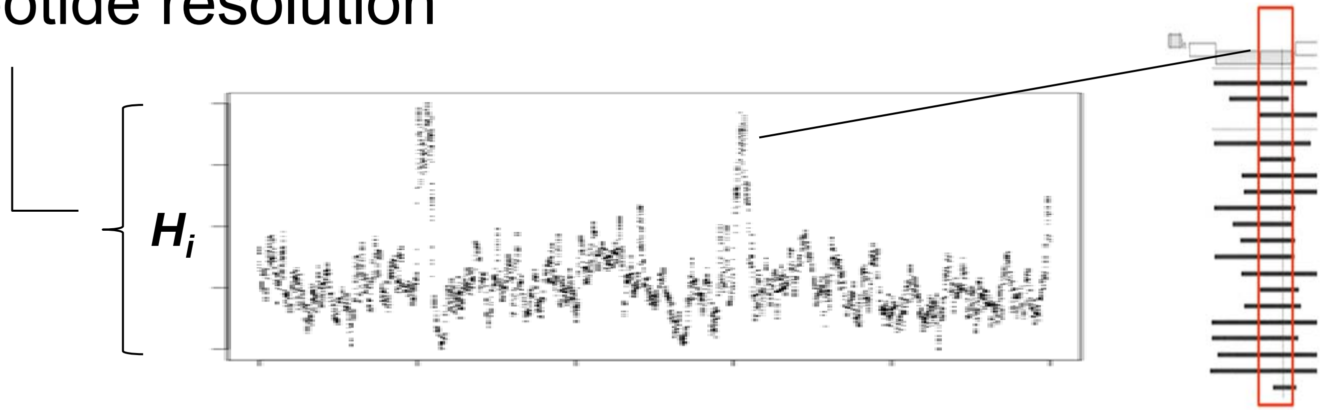
⇩

Improved to infer **intensity or frequency of recombination at a nucleotide**

Yahara et al (2014), *Mol. Biol. Evol.*

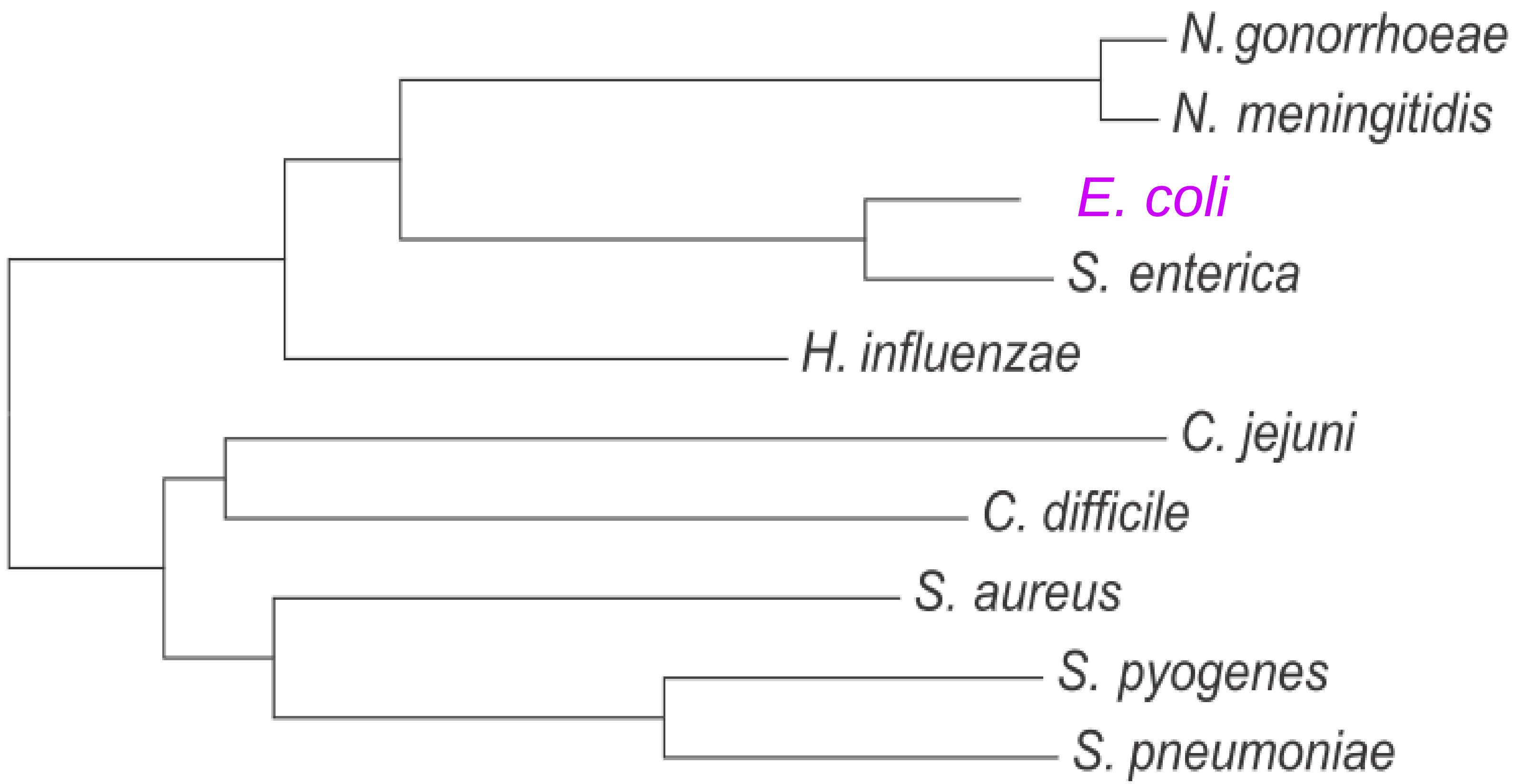
# Points of the method ("ordered painting")

- intensity of recombination along a genome at single-nucleotide resolution




  - highly correlated with local recombination rate
  - applicable to both clonal and highly recombining species
  - realized at population genetic level
    - <u>influenced by selection</u>
  - normalized for between-species comparison

# 10 species of public health importance

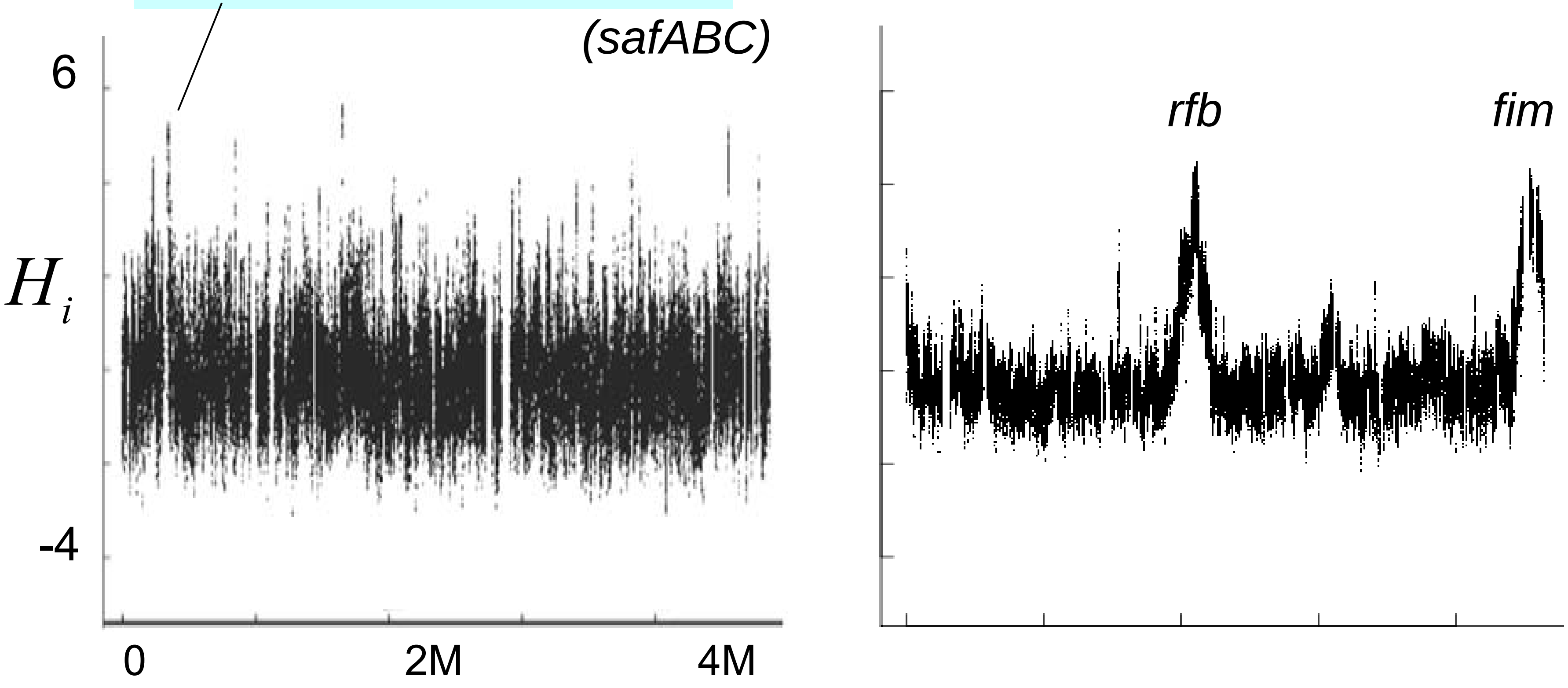



hundreds or thousands of genomes in Oxford
→ broadly selected 50 strains per species

Yahara et al (2016), *Mol. Biol. Evol.*

# *S. enterica* vs *E. coli*

intestine inflammation

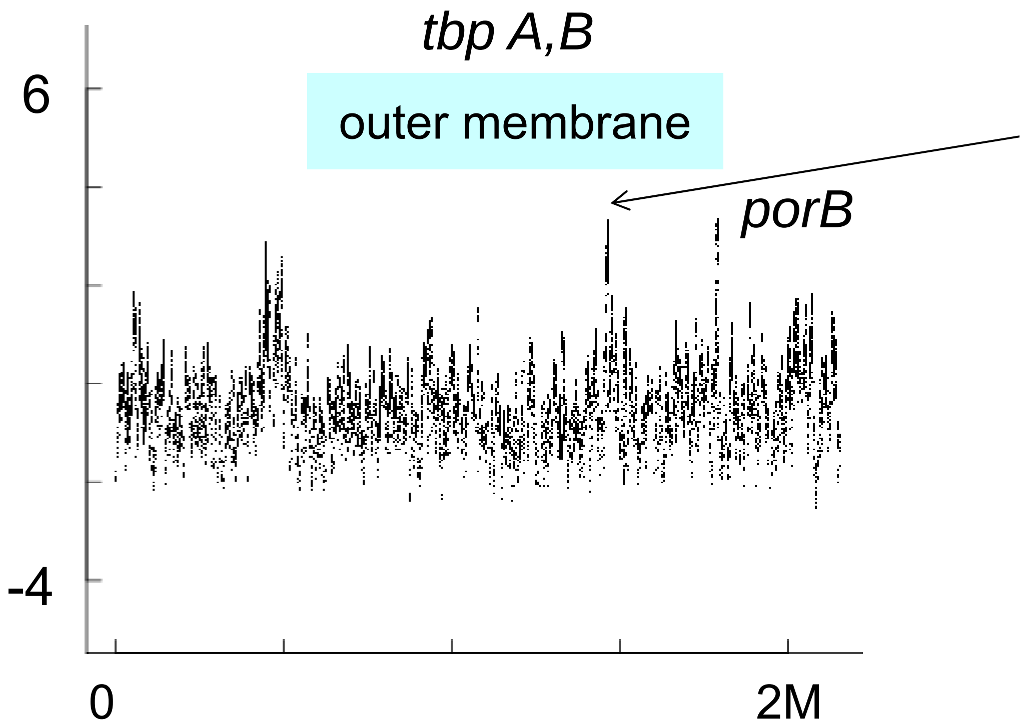


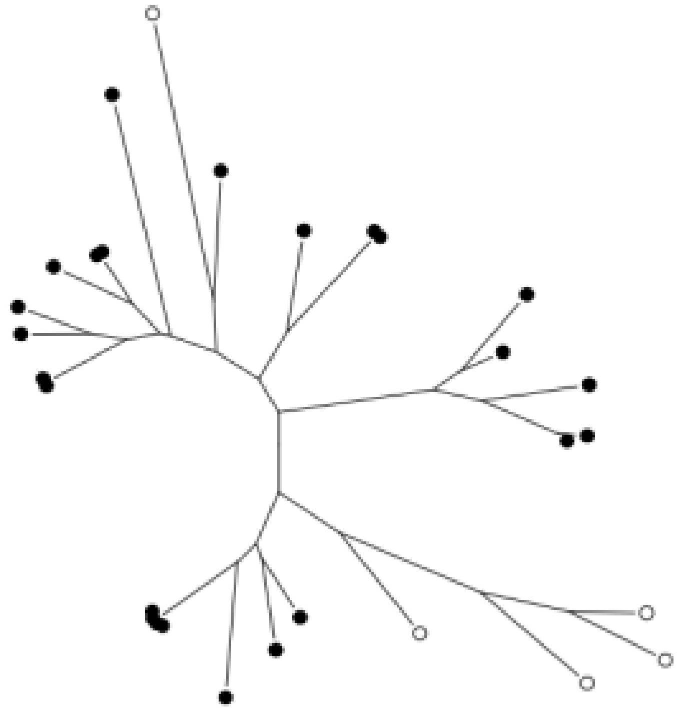

smaller hot regions than *E. coli*

highly variable even between related species

# *Two Neisseria spp.*

○ *N. gonorrhoeae* ● *N. meningitidis*



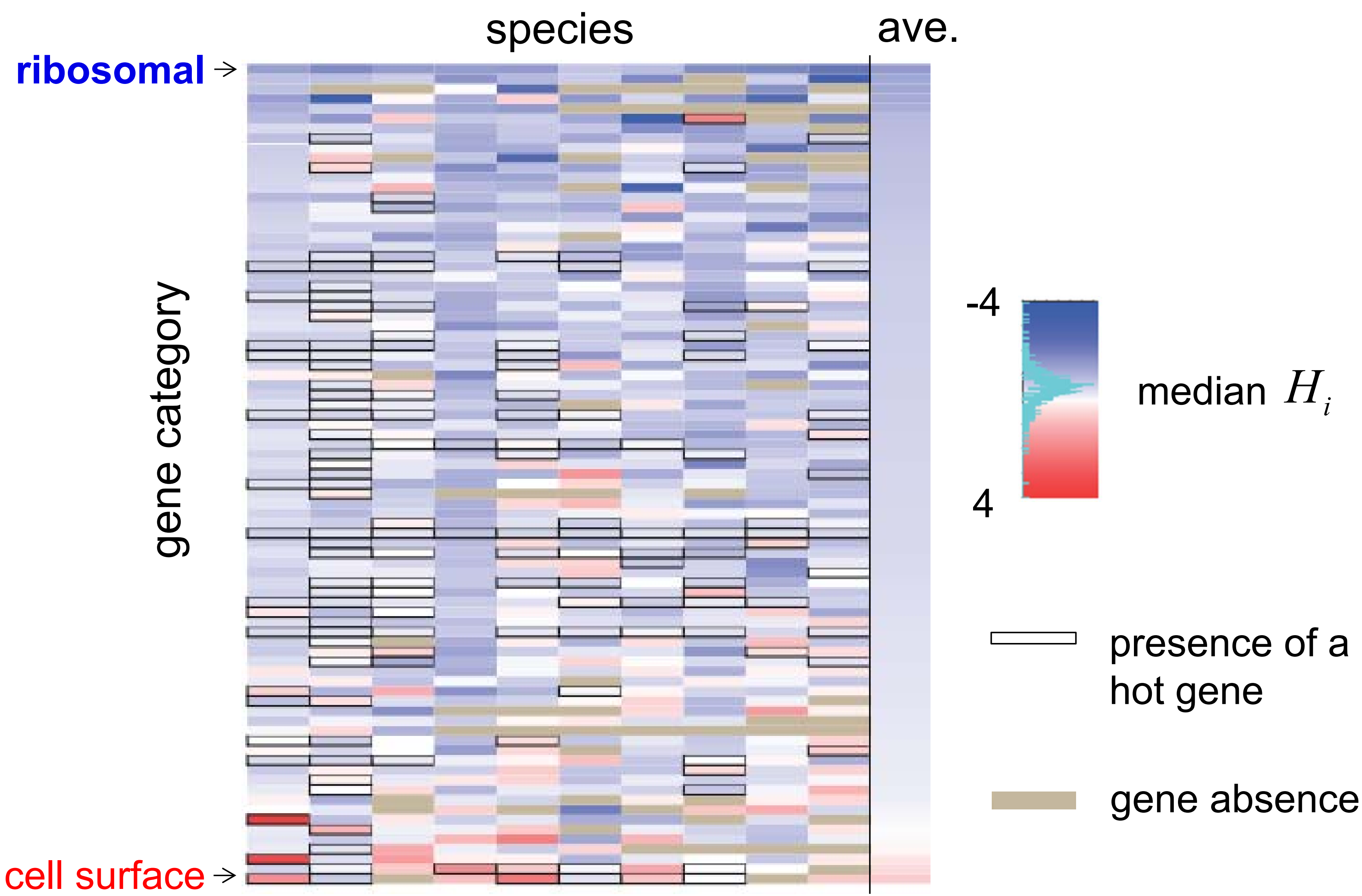



shared hot genes

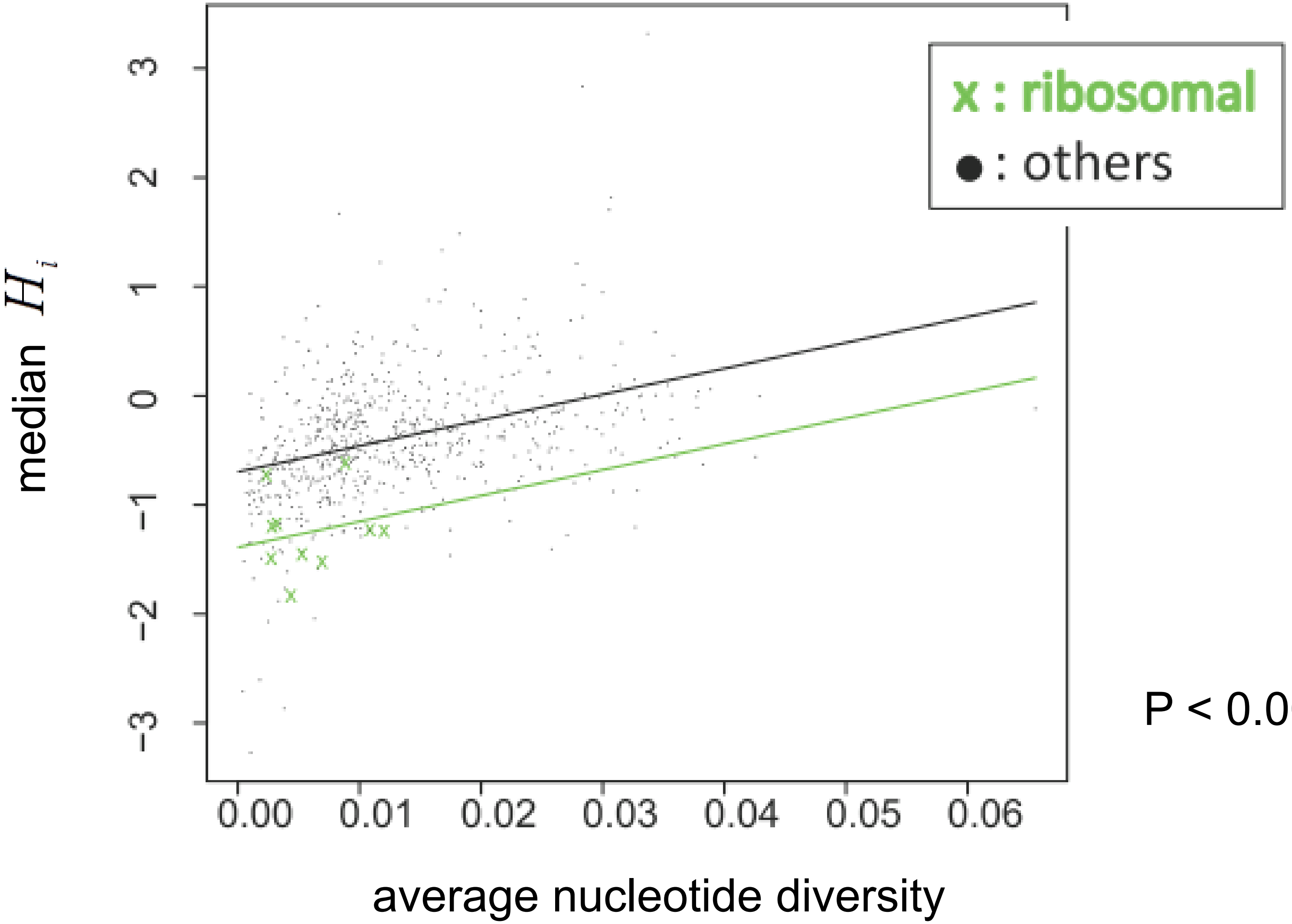
also in *H. influenza*
($P_{bonf} < 0.05$)

inter-species recombination

# Universal pattern across functional gene categories?



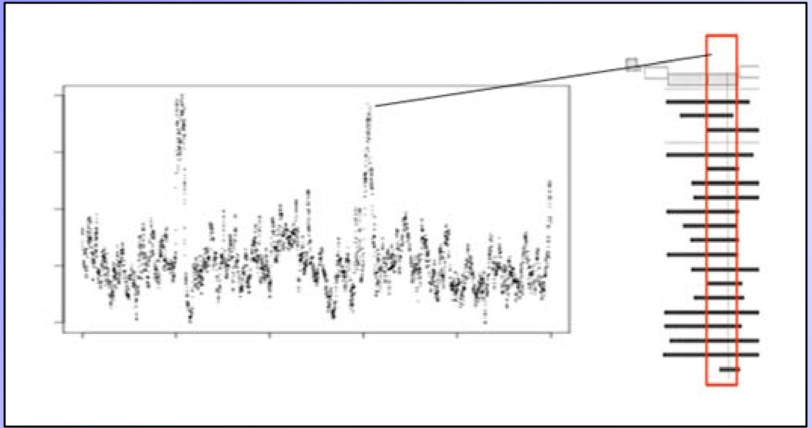
# Universal pattern across functional gene categories?



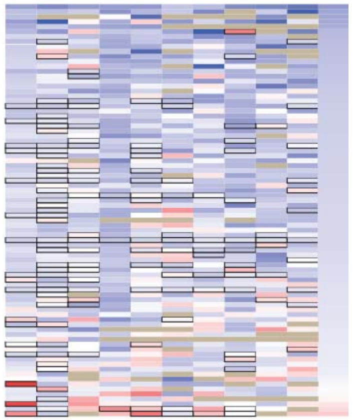

P < 0.0001

Yahara et al (2016), *Mol. Biol. Evol.*

# Recombination



Yahara et al (2014), *MBE*



Yahara et al (2016), *MBE*

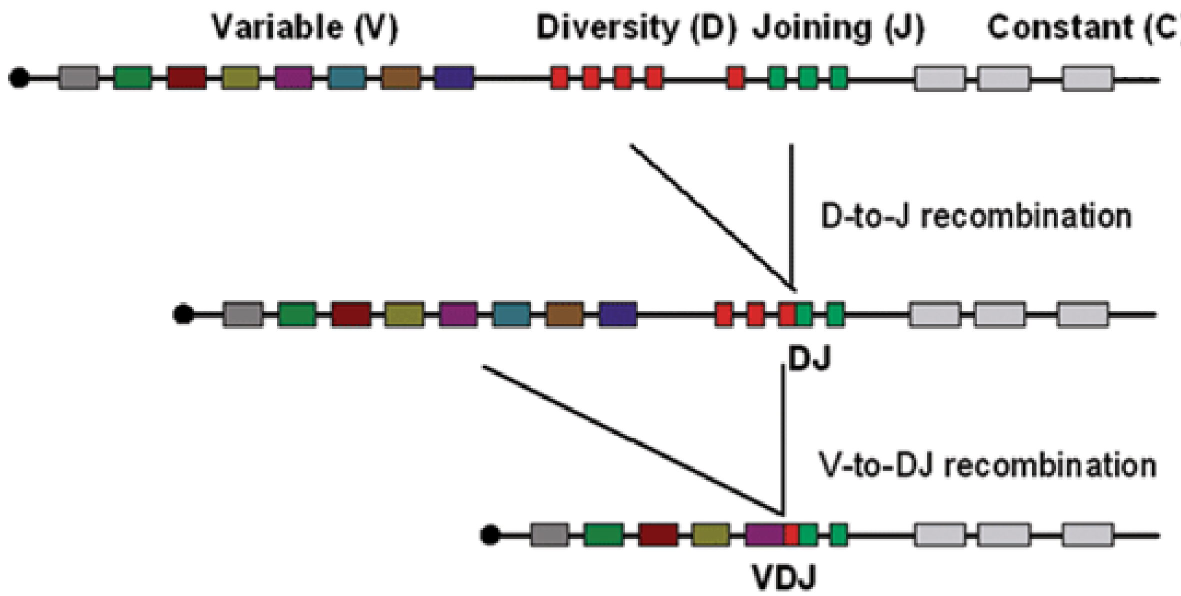
Relation to diversifying selection

Bacteriophage (virome)

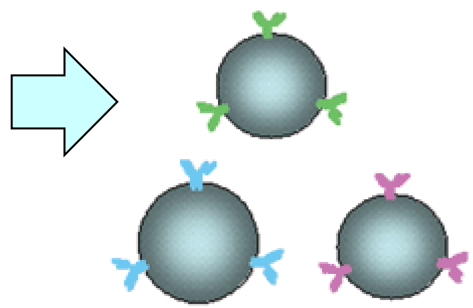
# Relation between recombination and selection

- A central problem in evolutionary biology

- Popular theory
  - recombination facilitates selection by creating advantageous genetic combination
    - some evidence (e.g. eukaryotic immune system)




Diverse antibody



# However, throughout a genome,

- Controversial in eukaryotes
  - “No effect of recombination on the efficacy of selection” measured by dN/dS in primates
    - Bullaughey, Przeworski et al (2008), *Genome Res*
- no quantitative study in bacteria

utilizing the high genomic diversity and recombination rate of *H. pylori*

1) How are codons under diversifying selection (dN/dS > 1) distributed in the genome?

2) Do such codons appear to be more frequently subject to recombination ?

# Difficulties in recombining genome

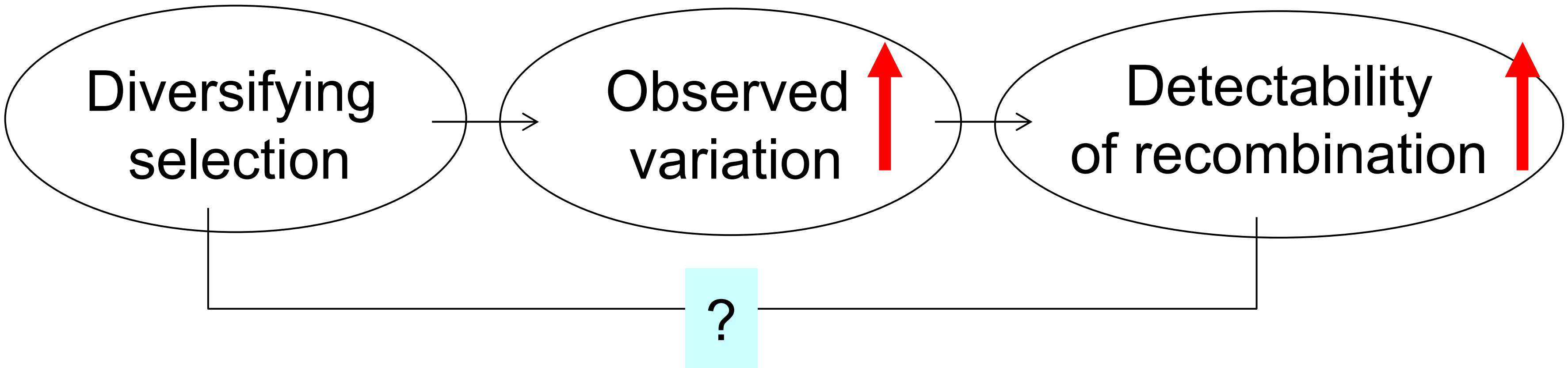
- Inference of dN/dS assumes <u>a single specific tree,</u>
  but recombination can change tree topology
  
  - a solution is to average it over

$$P(\mathrm{H} \mid \Theta) = \int P(\mathrm{H} \mid T, \Theta) P(T)\, dT$$

$\mathrm{H}$ : sequence data

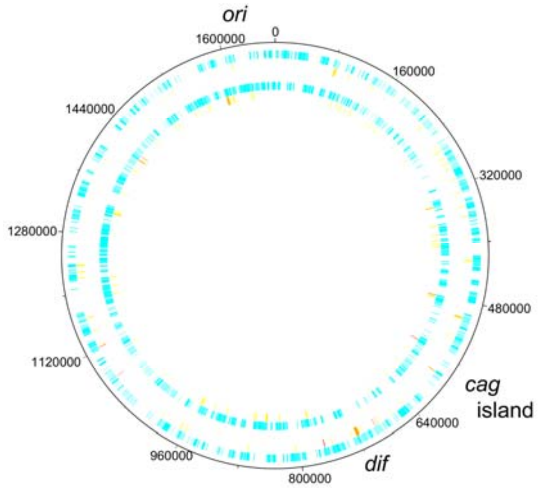
$\Theta$ : model parameters (e.g.. dN/dS)

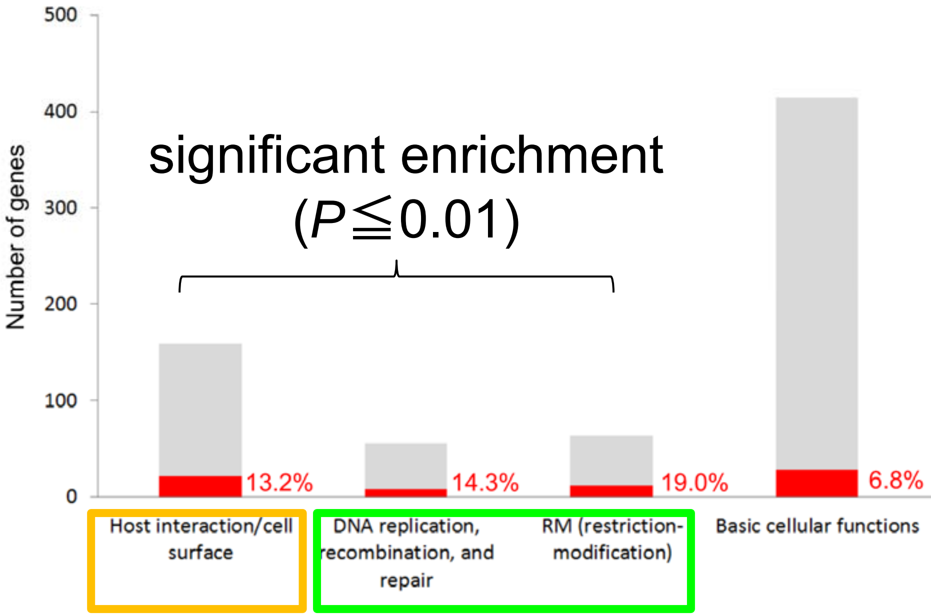
- Natural correlation between signatures of selection and recombination

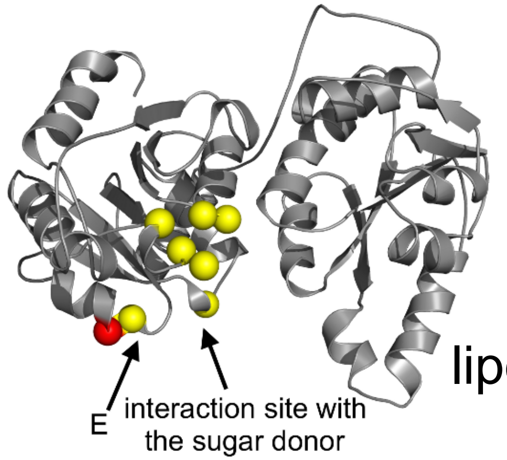


# Codons and genes under diversifying selection

- ~0.2% of all the codons



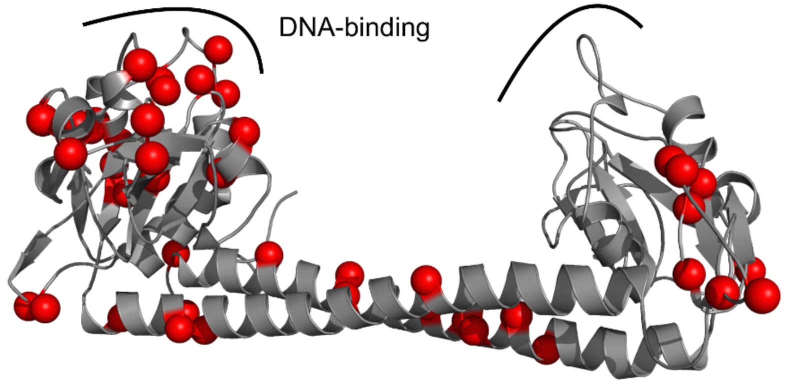




WaaC

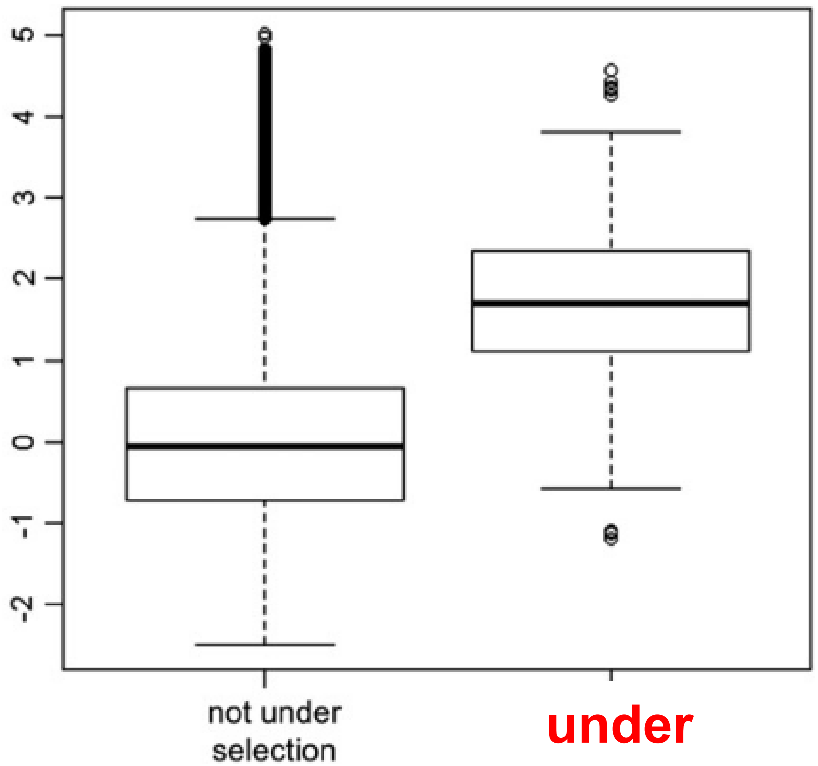
lipopolysaccharide synthesis

specificity subunit of restriction enzyme



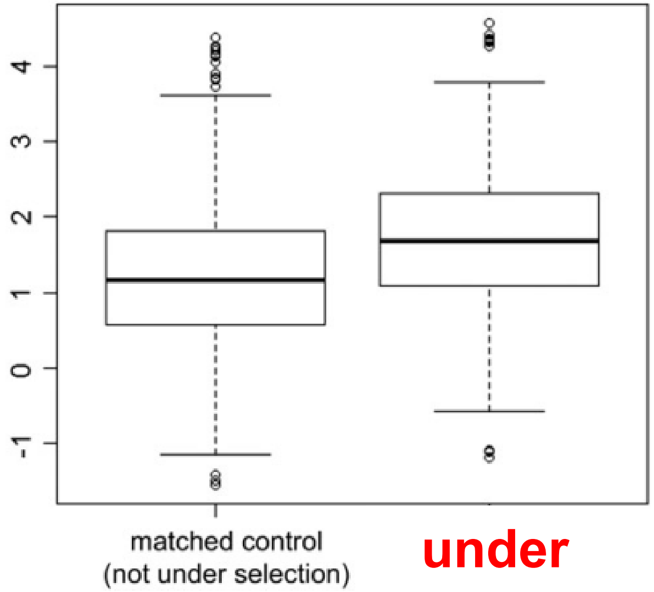
# Comparison of the intensity of recombination ($H_i$)

significantly higher



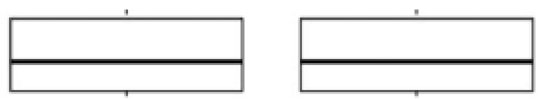

**under selection**

adjust



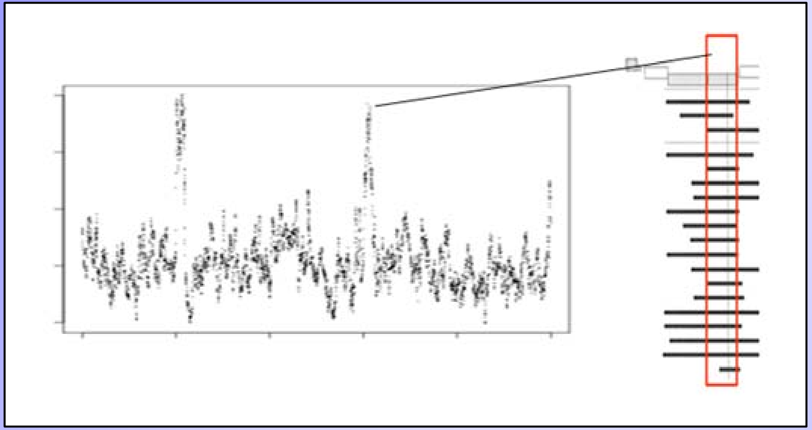

**under selection**

for "matched" nucleotides
carrying the closest level of diversity

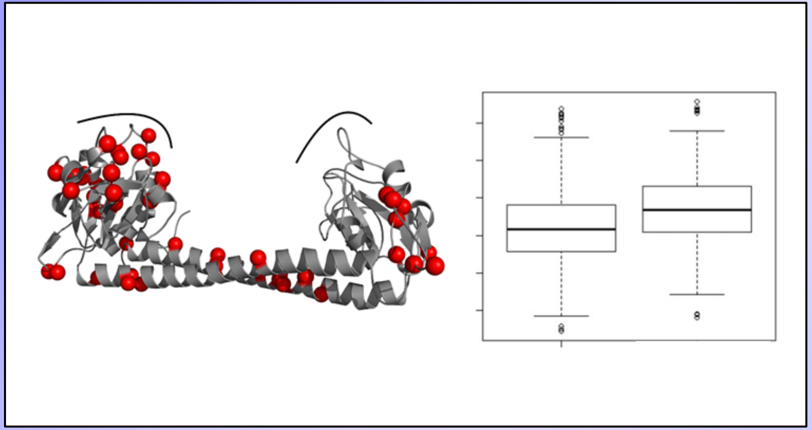


Yahara et al (2016), *DNA Research*

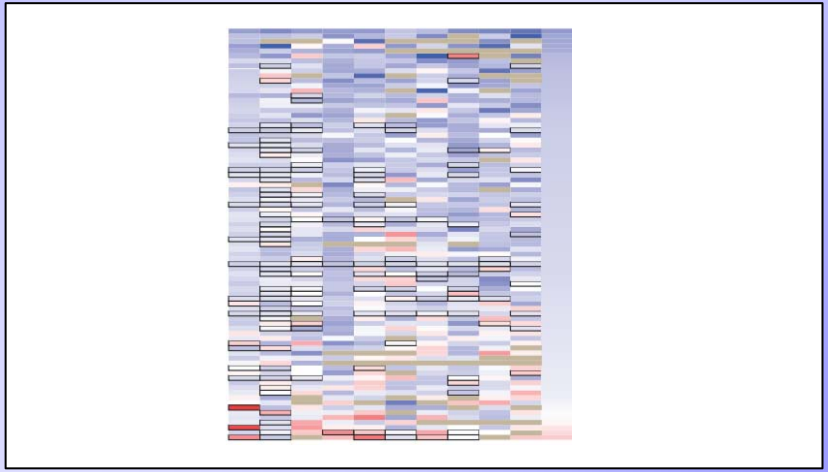
# Recombination



Yahara et al (2014), *MBE*



Yahara et al (2016), *DNA Res.*

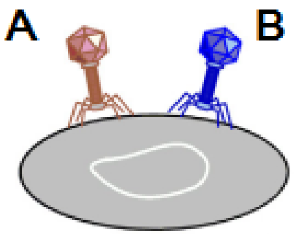


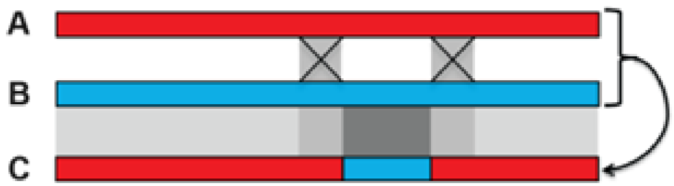
Yahara et al (2016), *MBE*

Bacteriophage
(virome)

# Phage and recombination (昨日の口頭発表)

- Most abundant and diverse biological entities
- Recombination occurs between co-infecting strains





but recombination does not necessarily increase the average fitness of offspring

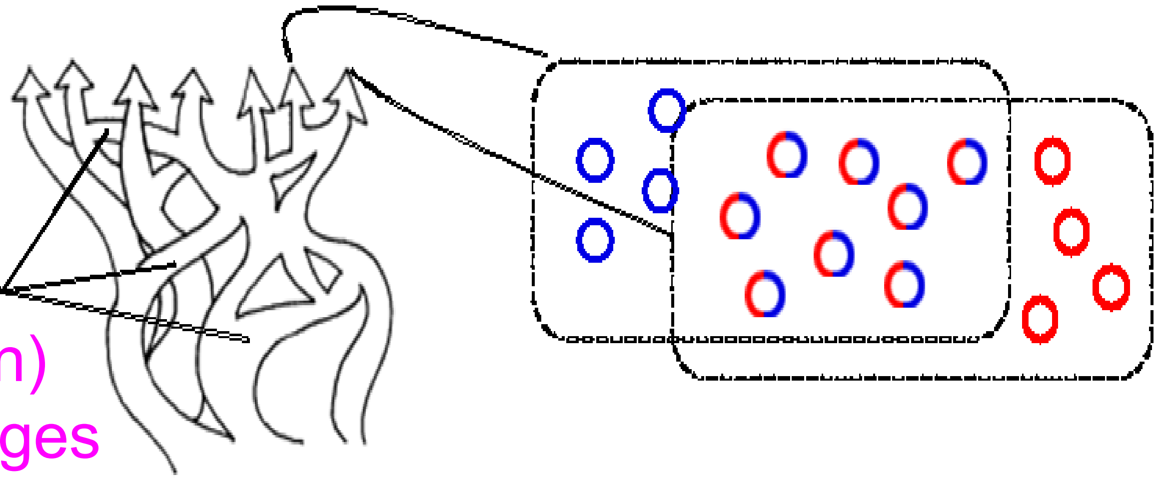
**Are signatures of recombination observed across various phylogenetic groups of phages?**

Investigation using Earth's virome data (Paez-Espino 2016, *Nature*)

Meier-Kolthoff et al (2018), *Sci. Rep.*

# Phylogeny & population structure

gene flow
(recombination)
between lineages



phenotype A

B

compare
genetic polymorphisms
(GWAS)

# Phylogeny & population structure

## gene flow (recombination)

*H. pylori*

in Americas

non-*H. pylori*

*Helicobacter* species

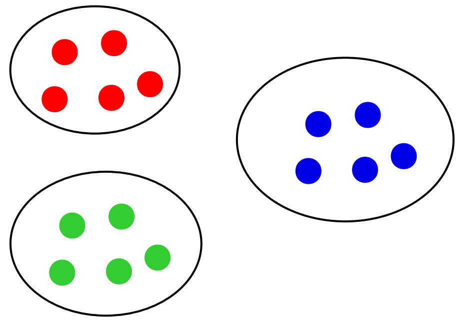
## GWAS

antimicrobial resistance
(*Acinetobacter*)

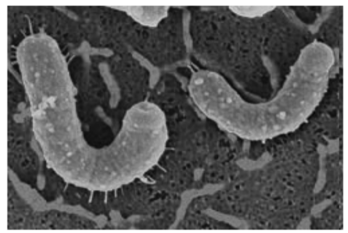
food-borne disease
(*Campylobacter*)

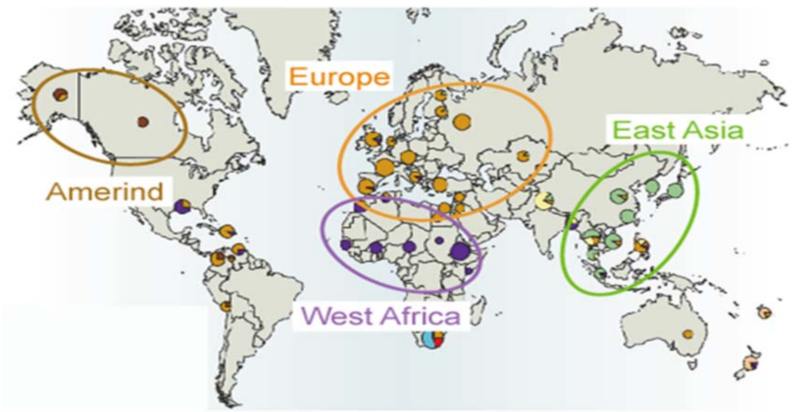
# Phylogeny & population structure

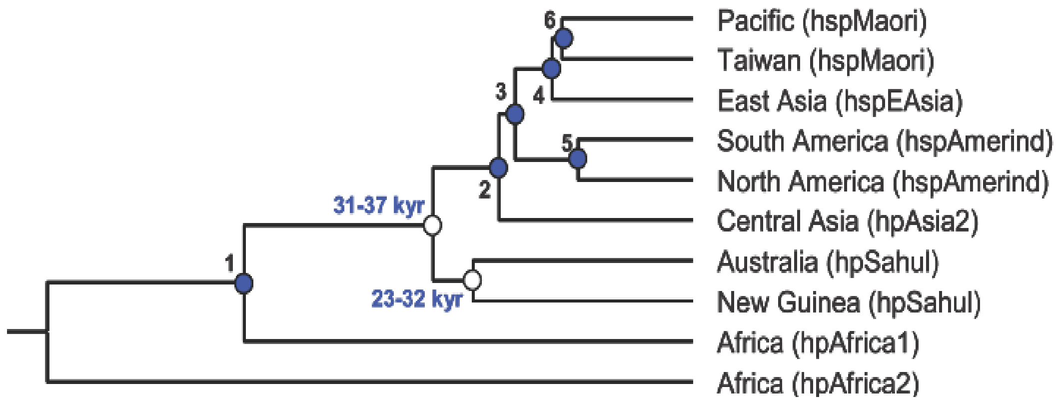
- Basis for various studies
  - population history & differentiation
  - selection
  - GWAS …



- *H. pylori* : interesting material
  - <u>phylogeographically differentiated</u>
    - Falush (2003) *Science*, Moodley (2009) *Science*, Moodley (2012) *PLoS Pathogens* …

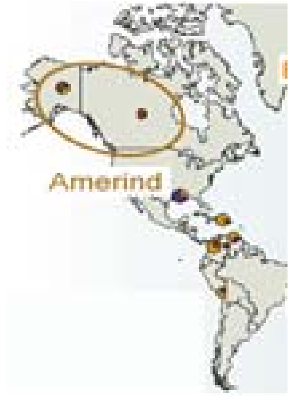




# Largely unexplored: Americas

- Common view: American *H. pylori* are basically European
- A known subpopulation: hspAmerind (Native Americans)
  - but rare
- Latin America: high mortality rate
  - associated with genotypes of *H. pylori*
    - but by only several genes

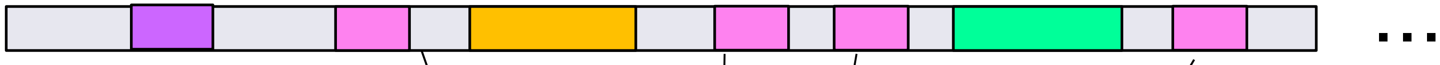



Kodaman (2014) *PNAS*

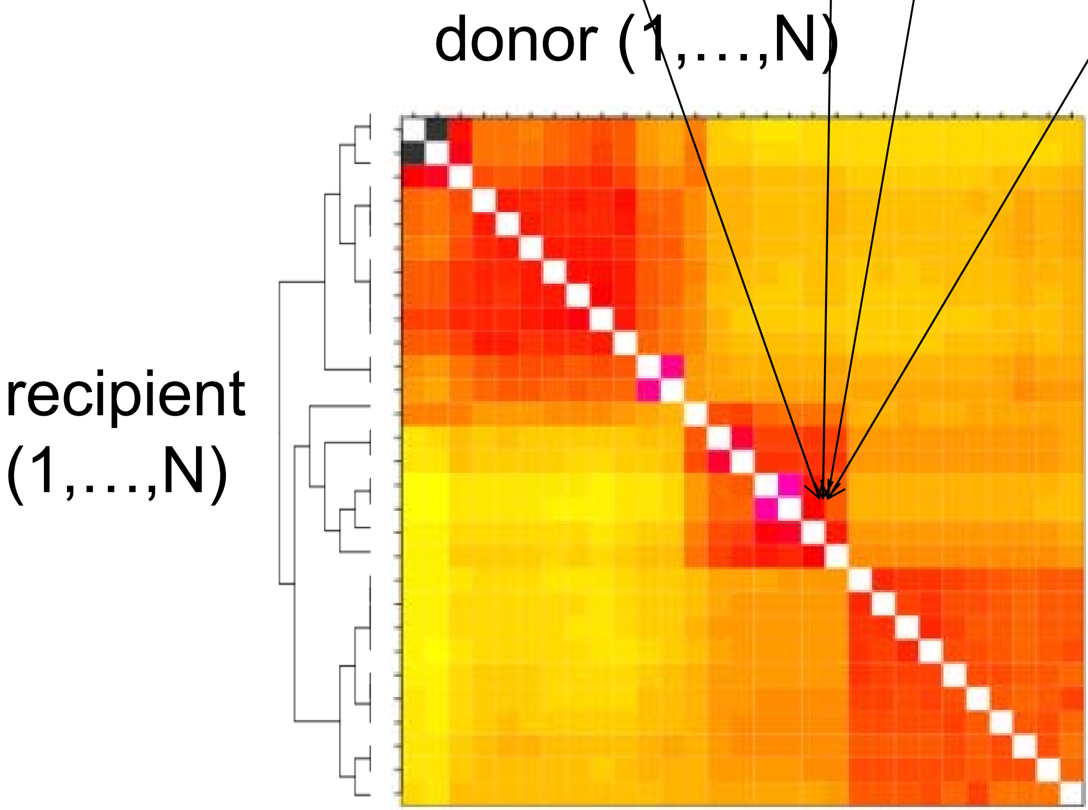
**American *H. pylori* differentiated from those in Old World?**

**Analysis of >400 genomes in various countries by the chromosome painting and another improvement**

# Chromosome painting for population structuring



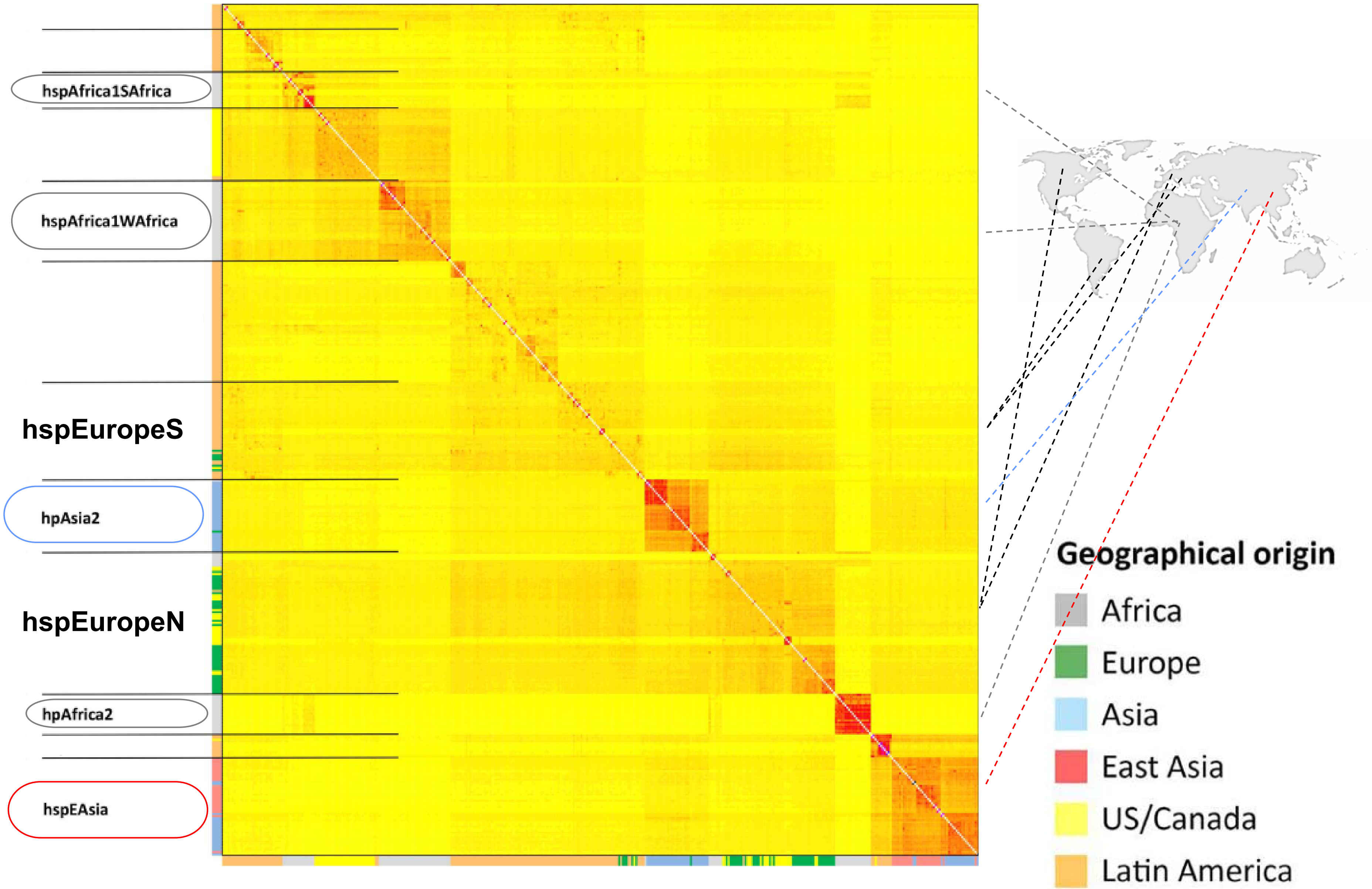
- into a “co-ancestry matrix”
  - counts the number of chunks



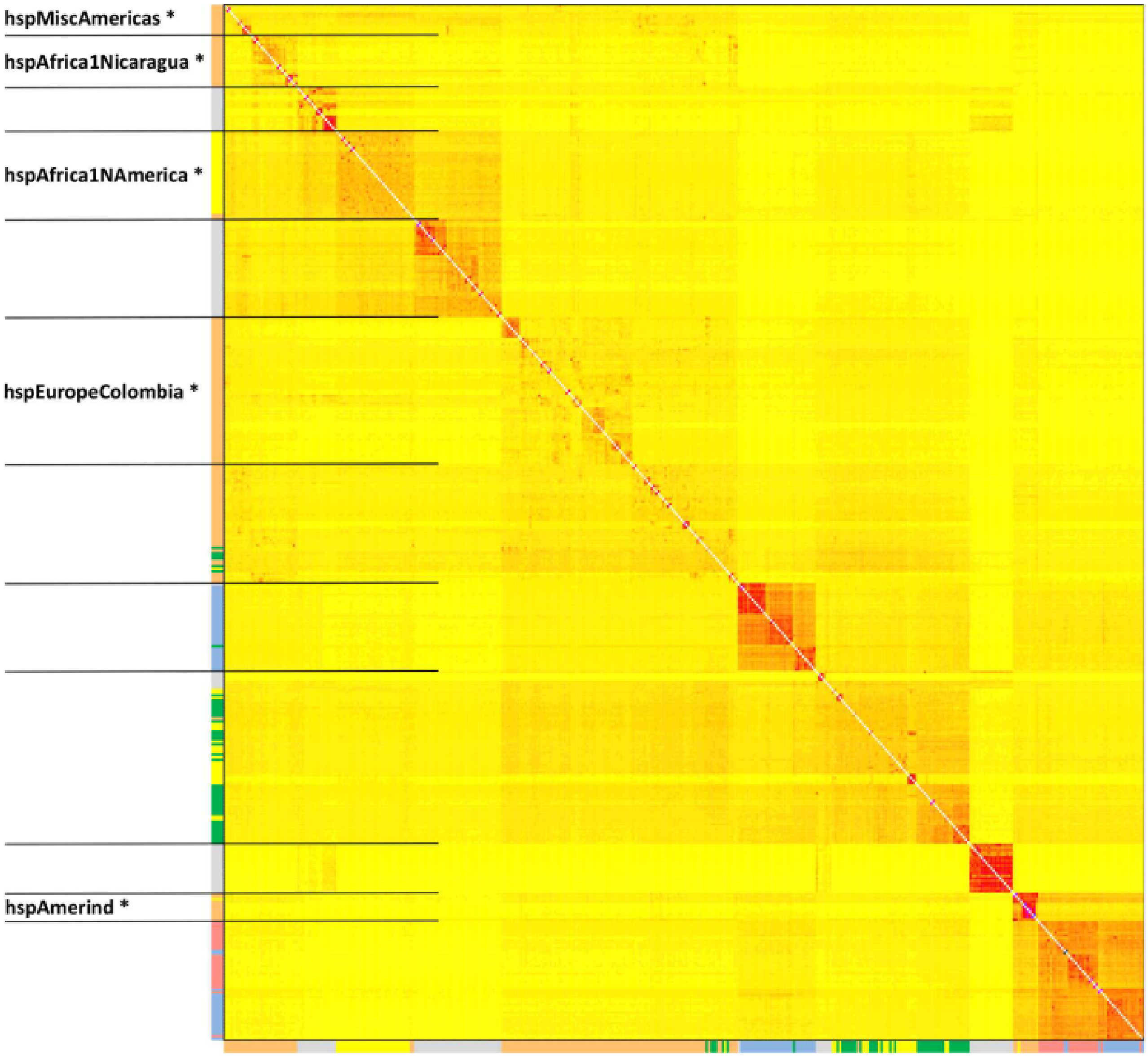

data reduction

fineSTRUCTURE captures more subtle population structure

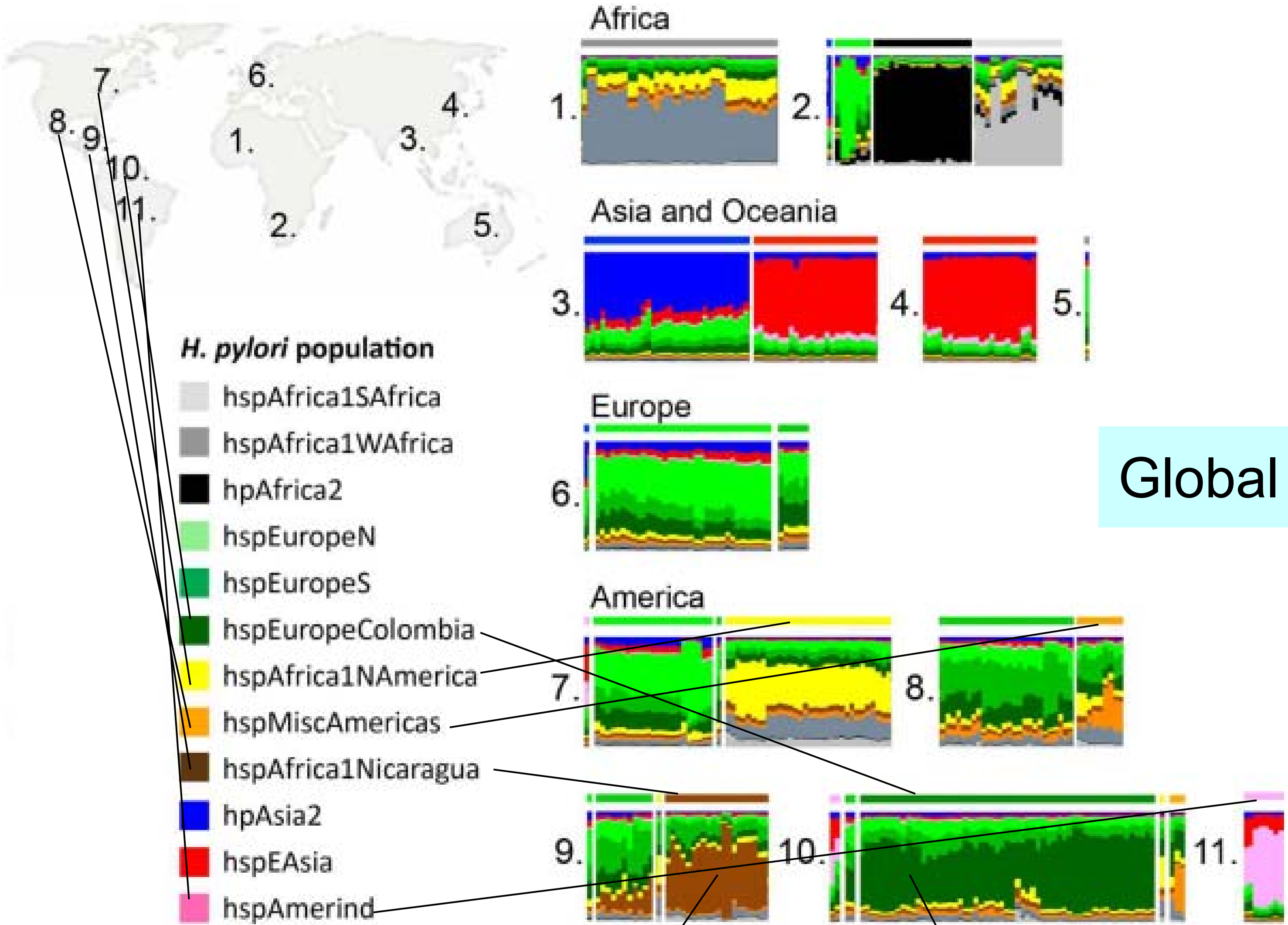
# Substructures in the Old World populations



# Distinct subpopulations in Americas



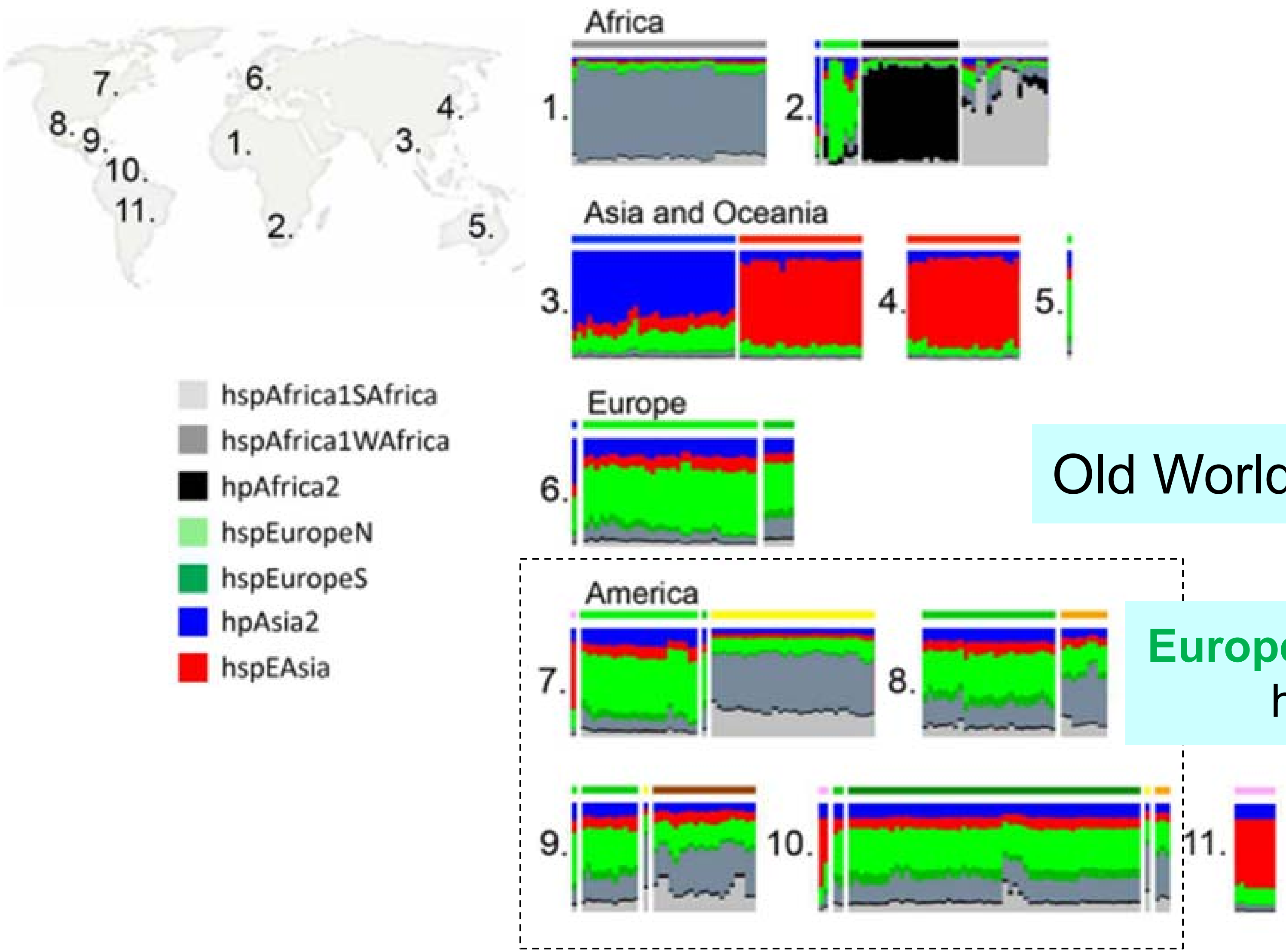

# Visualization of ancestry profile of each strain




Global painting

national gene pool in Nicaragua and Columbia

# Visualization using only Old World strains as donors



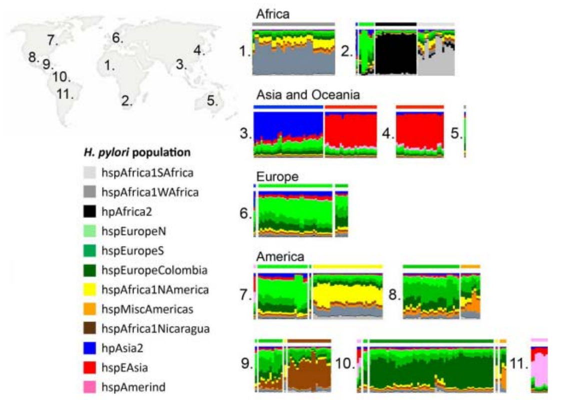

Old World painting

European/African hybrids

each column ("palette"): each strain

# Phylogeny & population structure

## gene flow (recombination)




Thorell*, Yahara* et al (2017), *PLoS Gen.*

non-*H. pylori*

*Helicobacter* species

## GWAS

antimicrobial resistance
(*Acinetobacter*)

food-borne disease
(*Campylobacter*)

# Various other *Helicobacter* species

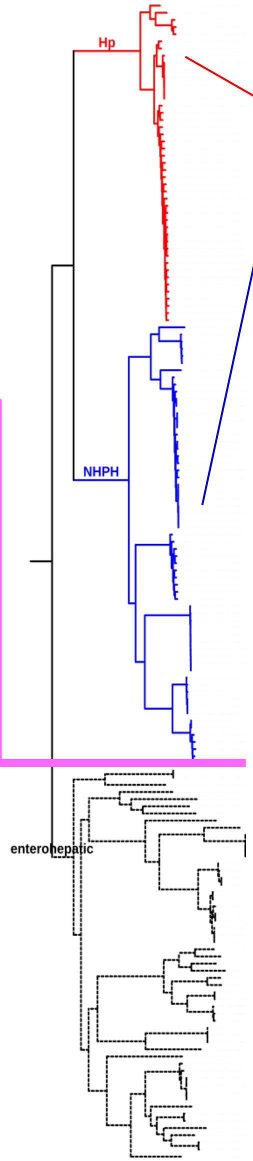
- in the stomach of domesticated and wild mammals

*H. pylori*
*H. acinonychis*
*H. cetorum*
*H. baculiformis*
*H. bizzozeronii*
*H. ailurogastricus*
*H. cynogastricus*
*H. felis*
*H. heilmannii*
*H. salomonis*
*H. suis*

gastric

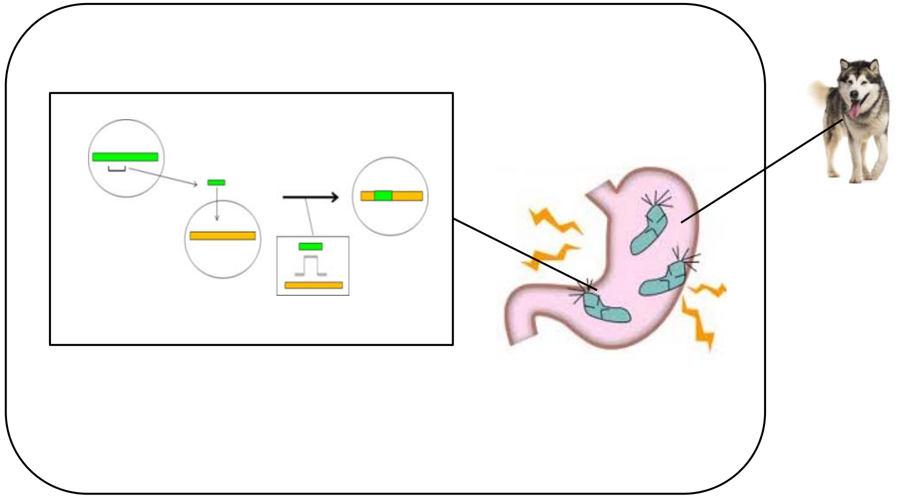
*H. cinaedi*
*H. fennelliae*
*H. muridarum*
*H. canis*
*H. hepaticus*
*H. pullorum*
*H. bilis*
*H. cholecystus*
*H. trogontum*
*H. rodentium*
*H. typhlonius*
*H. mesocricetorum*
*H. pametensis*
*H. canicola*
*H. canadensis*
*H. apodemus*
*H. aurati*
*H. marmotae*
*H. winghamensis*
*H. equorum*
*H. macacae*
*H. magdeburgensis*
*H. sanguini*

entero
hepatic

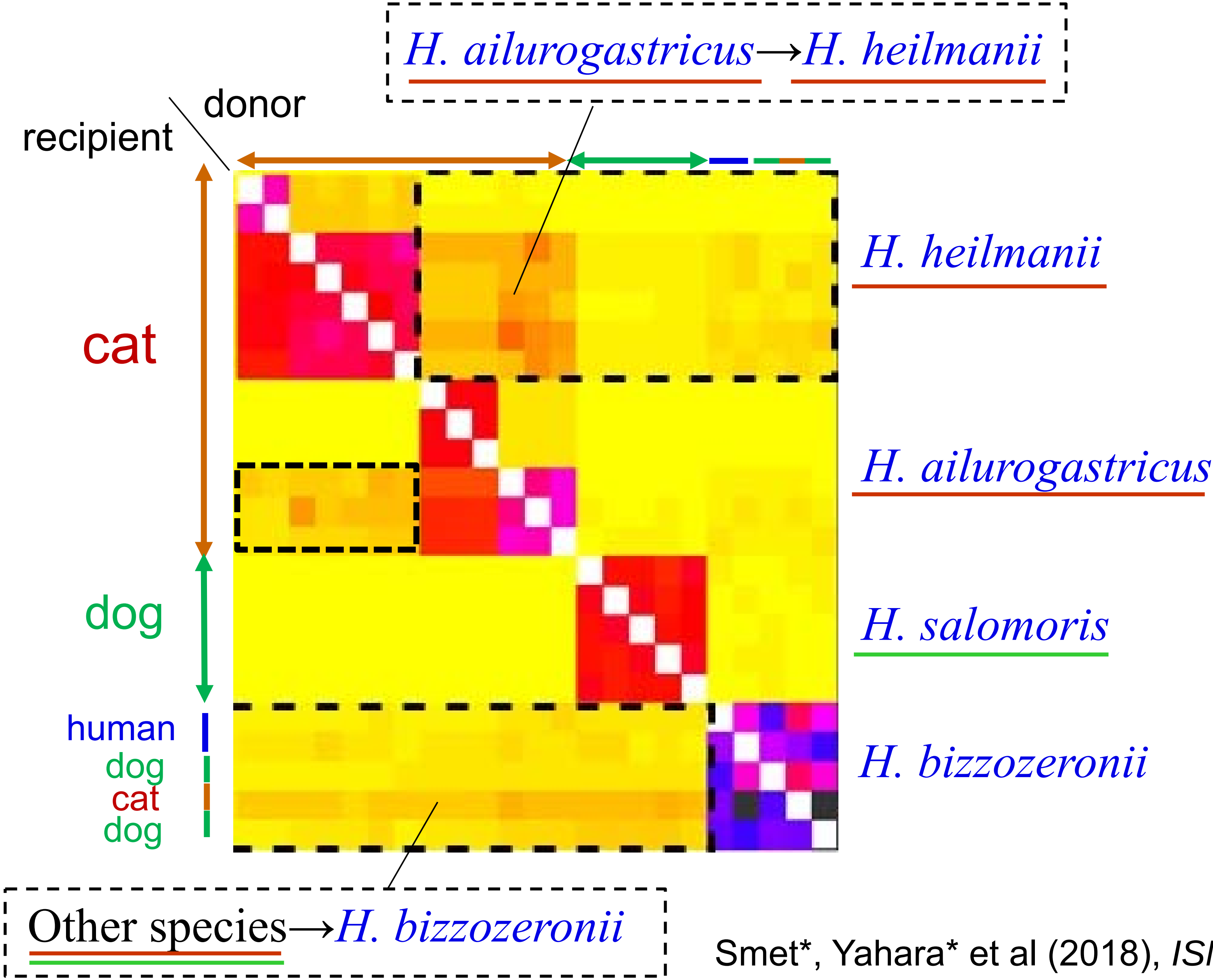



**diverged at > 1m years ago**

**Inter-species recombination among different species co-infecting to a pet**



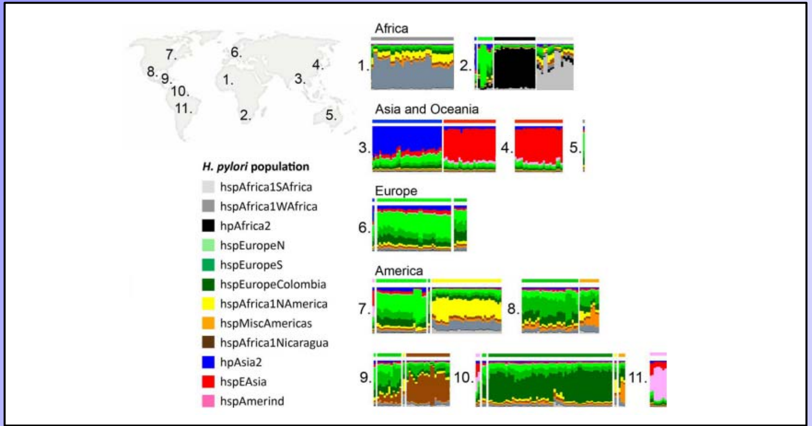
# Co-ancestry matrix by the chromosome painting



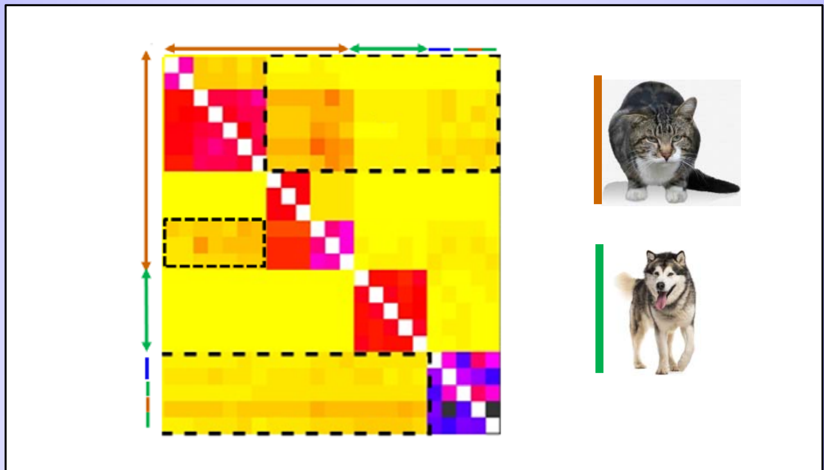

Smet*, Yahara* et al (2018), *ISME*

# Phylogeny & population structure

## gene flow (recombination)




Thorell*, Yahara* et al (2017), *PLoS Gen.*



Smet*, Yahara* et al (2018), *ISME*

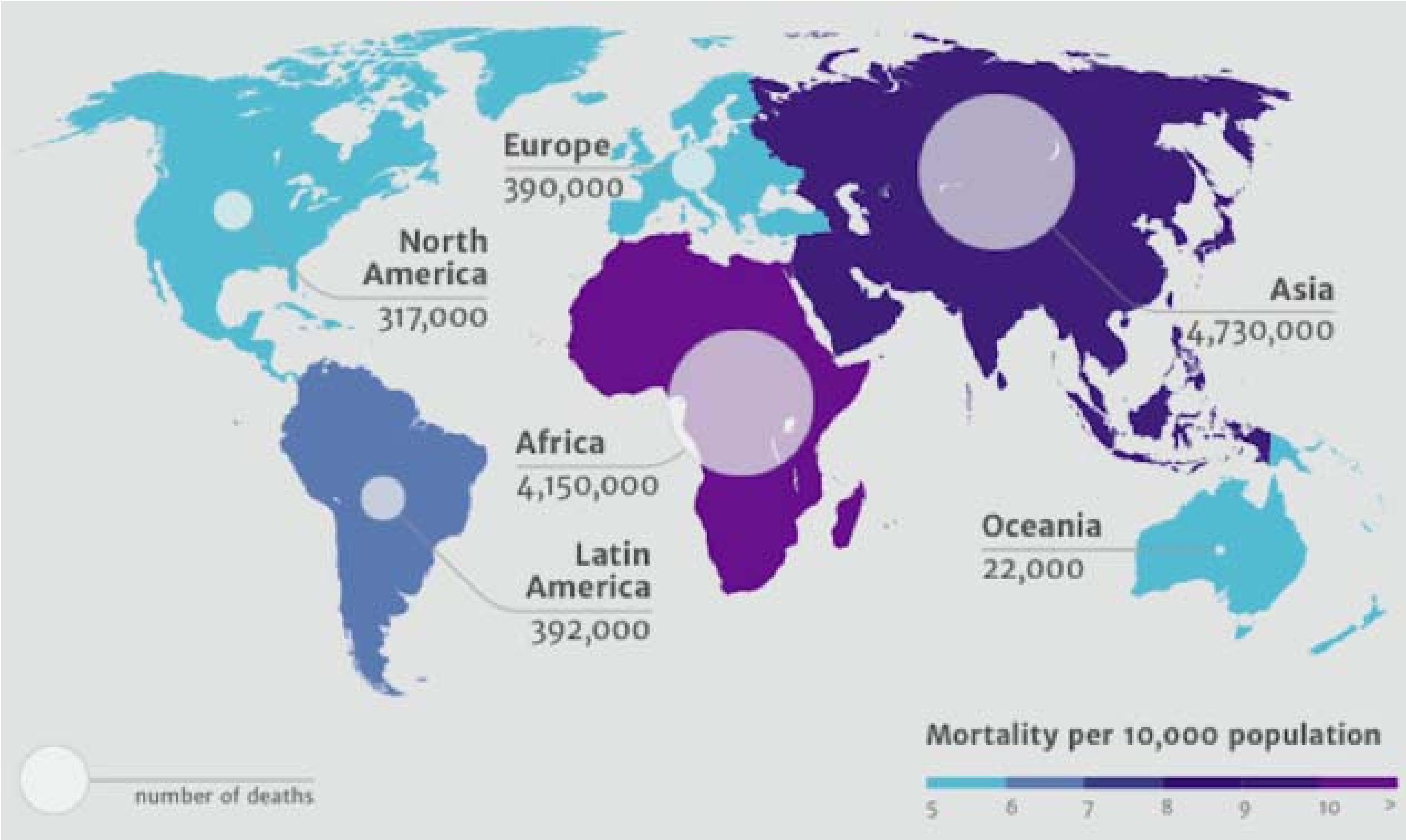
## GWAS

antimicrobial resistance
(*Acinetobacter*)

food-borne disease
(*Campylobacter*)

# Antimicrobial resistance

- 10 million deaths per year by 2050
  - > cancer deaths!!



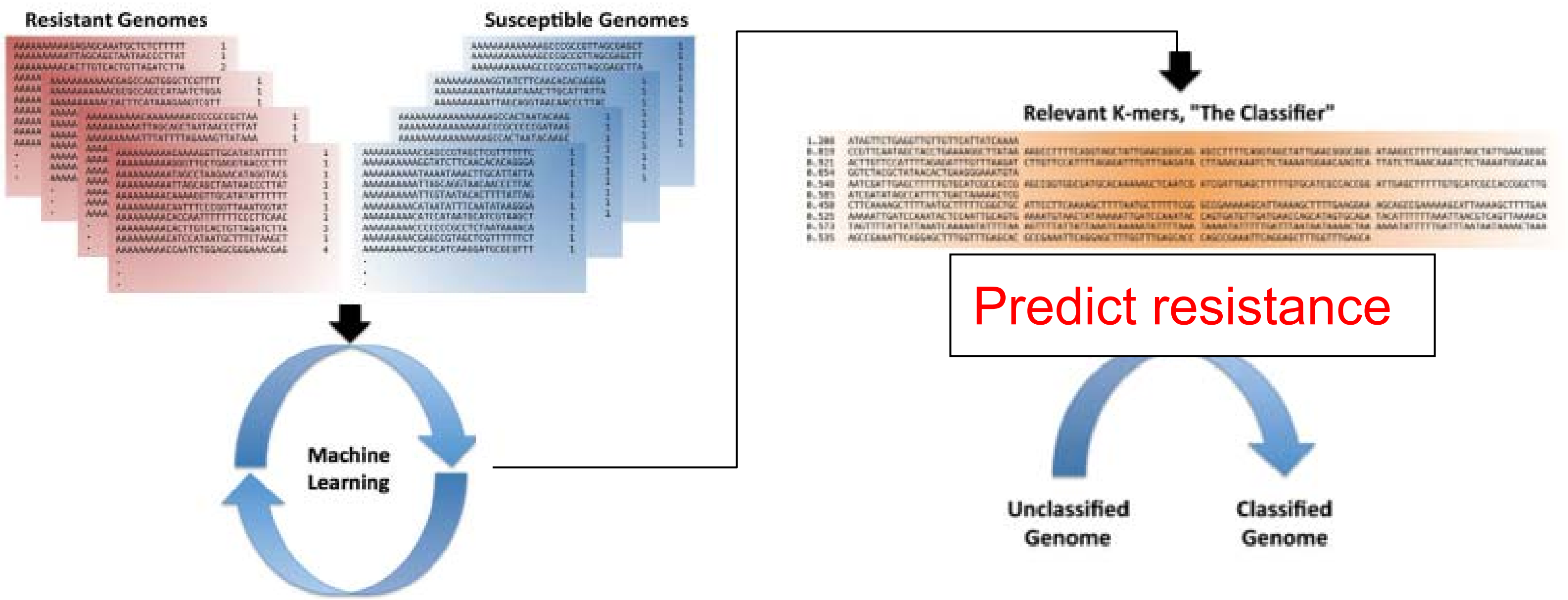

AMR-Review.org (2014)

# Carbapenem & *Acinetobacter*


- last-resort antibiotic
  - broad spectrum, critically important in medicine
- carbapenem-resistance
  - *Enterobacteriaceae* (腸内細菌科細菌)
  - *Pseudomonas aeruginosa*
  - <u>*Acinetobacter baumannii*</u>
    - surviving in a wide range of environments
    - notoriously difficult to control in hospitals

# A recent genomic study

- 122 carbapenem-resistant and 110 carbapenem-susceptible *A. baumannii* strains
  - first dataset of > 100 genomes + AMR metadata
    - available in PATRIC database

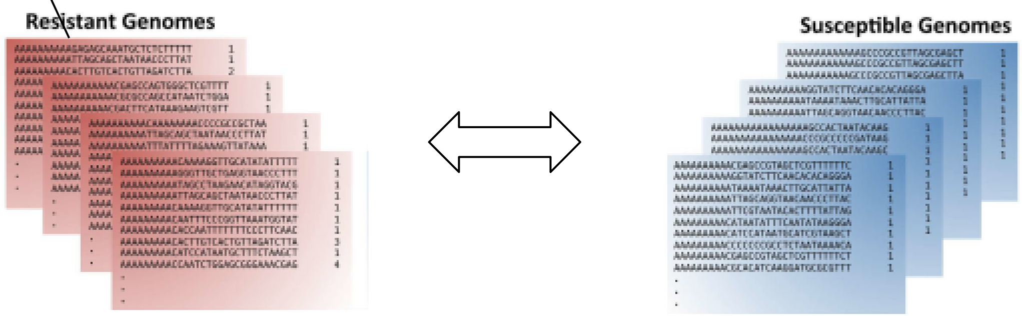



  - built a machine-learning classifier to predict resistance of a strain, with accuracy approximately 95%

Davis et al (2016) *Sci. Rep.* modified

# My scope

- The data included the commonly known features (e.g. *bla*$_{OXA-23}$)
  - prediction of resistance is obviously possible
- More interesting: exploration of novel genetic elements
  - in strains lacking the commonly known resistance features




Genome-wide association study (GWAS)

identifying any kind of genetic variation (SNP, indel, gene) enriched in resistant population

method in *bugwas* package, Earle et al (2016), *Nature Microb.*

# Bacterial GWAS methods

1) Breakdown genomes into words

to capture any kind of variation (SNP, indel, gene)

genomes

AAAAAAAAATTATACAAGCTGTGTATACATA
AAATTGAGTAACTTTATCGCCAATCGCTTCT
CAGATTTACTCGATAAAAGAAACAAACAAAG
AAATATTGAAAAAATCCTTGTTCCAAGCGAA
...

k-mer
(e.g, 31 bp)

(different start positions)

| | i=1 | i=2 | ... | i=N |
|---|---|---|---|---|
| word1 | 1 | 0 | | 0 |
| word2 | 0 | 1 | | 1 |
| word3 | 1 | 1 | | 0 |
| word4 | 0 | 0 | | 0 |
| word5 | 1 | 0 | | 1 |
| word6 | 1 | 1 | | 1 |
| ... | | | | |
| R or S | R | S | ... | R |

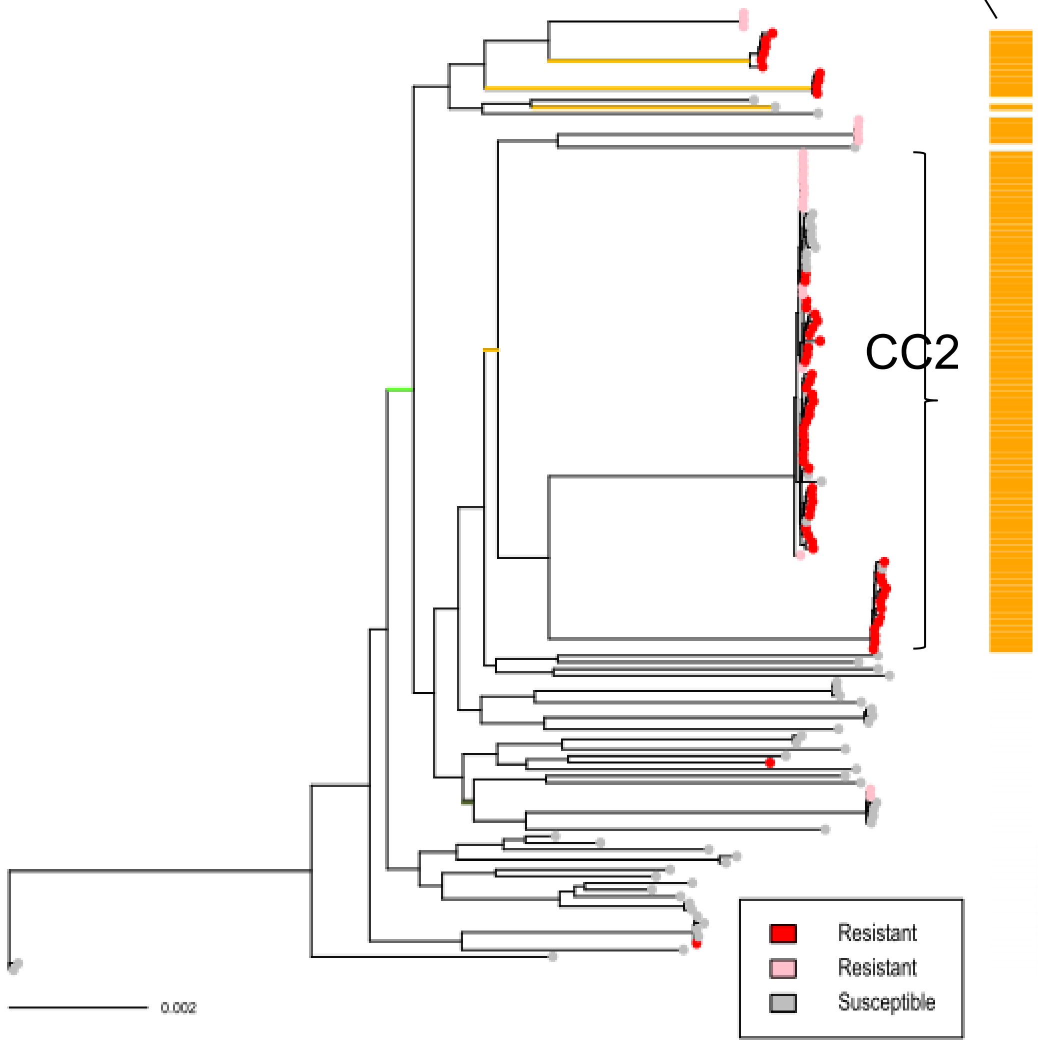
individuals

2) Statistical test of significance

after accounting for phylogenetic relatedness of the strains

(not independent!!)

# Top hit: three overlapping words that were >70% more frequent in the resistant strains

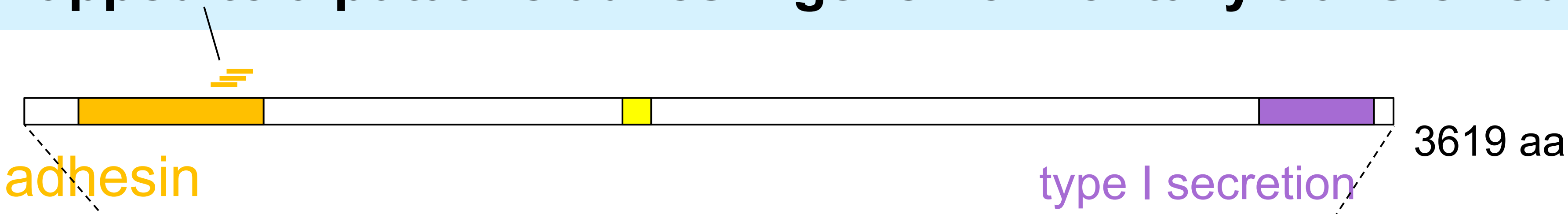



across multiple lineages

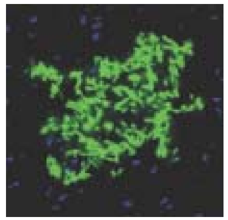
10% higher than Sheppard (2013), *PNAS*

$P<10^{-4}$

# Mapped to a putative adhesin gene horizontally transferred

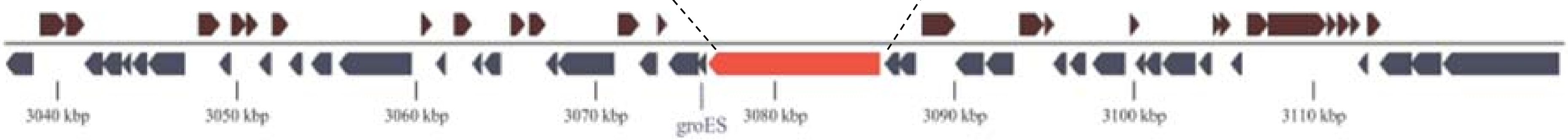



mediates self-association and biofilm formation



Sherlock (2004) *J. Bac.*

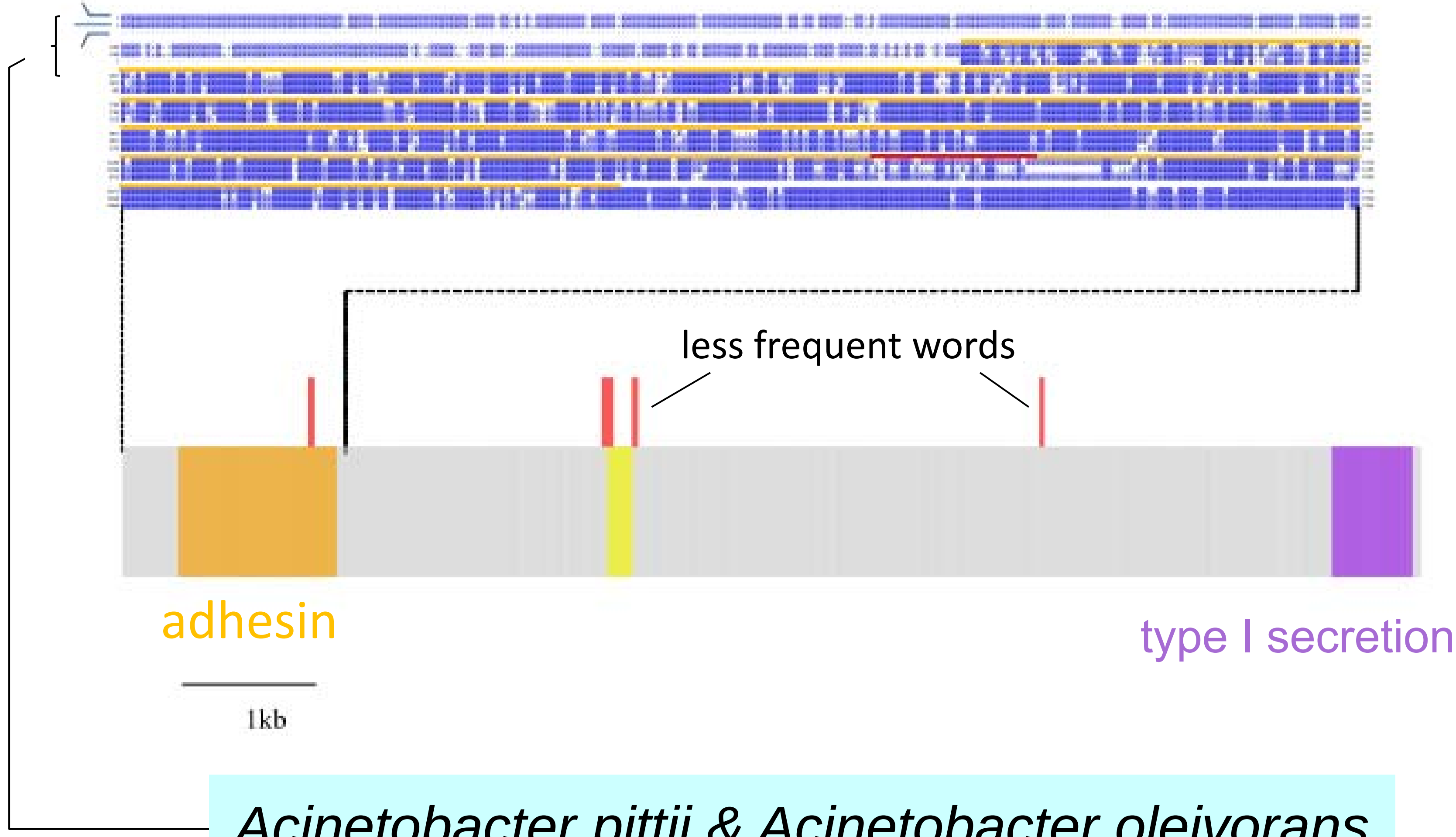
resistant genome




**atypical** nucleotide composition & gene tree topology

# Also found in other *Acinetobacter* species

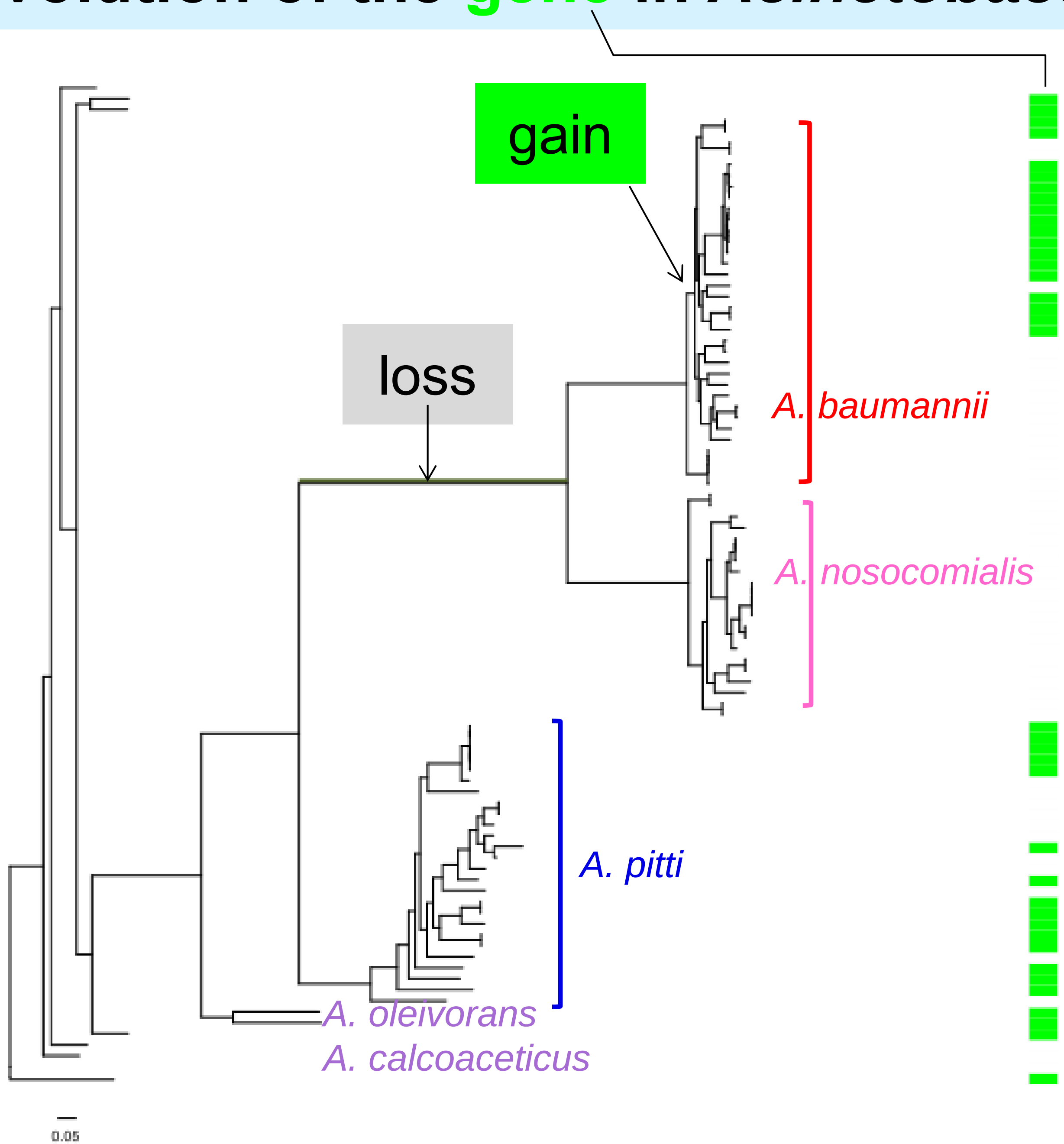
*A. baumanni* ACICU




*Acinetobacter pittii & Acinetobacter oleivorans*

89% and 88% sequence identity
over 100% and 95% of the alignment length of the locus

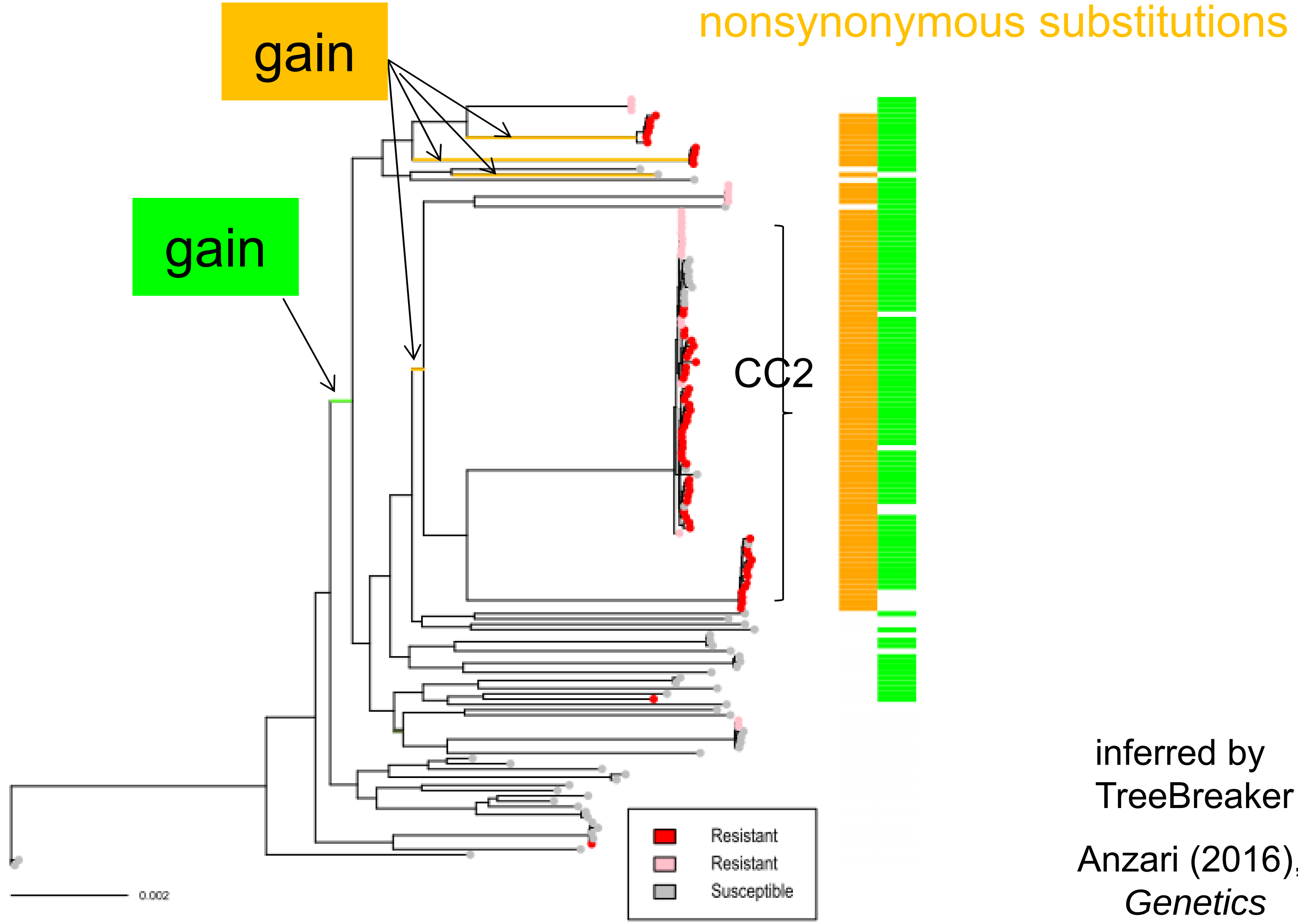
# Evolution of the gene in *Acinetobacter* spp.




inferred by TreeBreaker

Anzari (2016), *Genetics*

# Evolution of the gene and word in *A. baumanni*



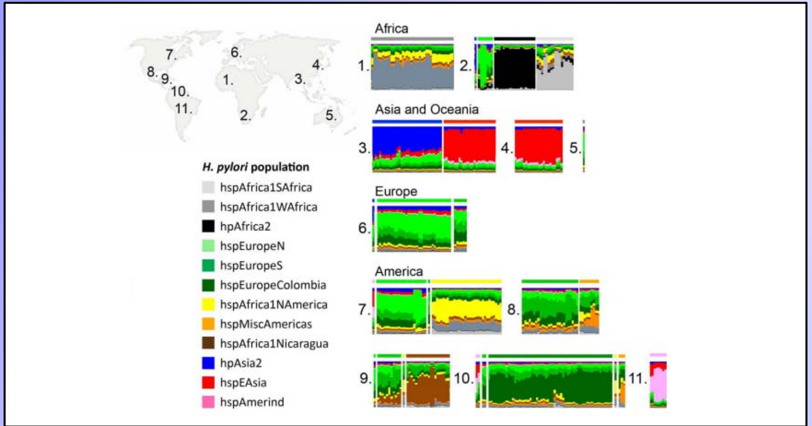

inferred by TreeBreaker

Anzari (2016), *Genetics*

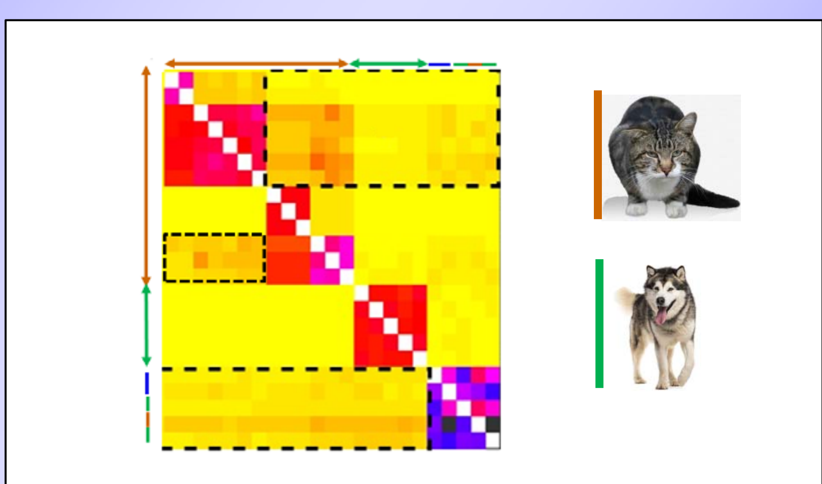
recurrent evolutionary signals across different lineages

# Phylogeny & population structure

## gene flow (recombination)

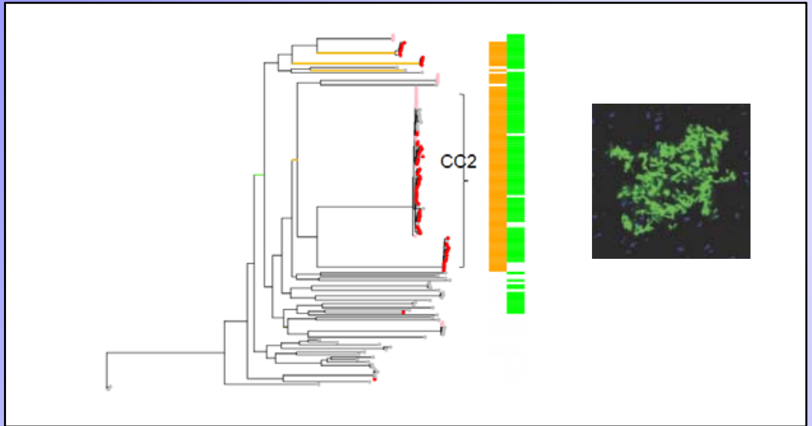



Thorell*, Yahara* et al (2017), *PLoS Gen.*



Smet*, Yahara* et al (2018), *ISME*

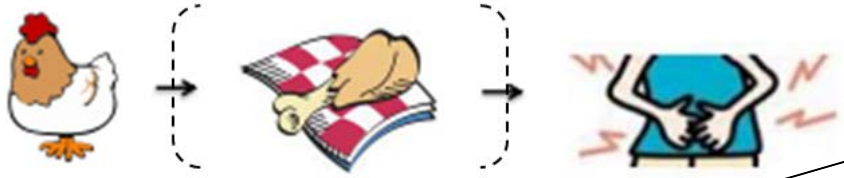
## GWAS




Suzuki, … ,Yahara (2016), *Sci. Rep.*

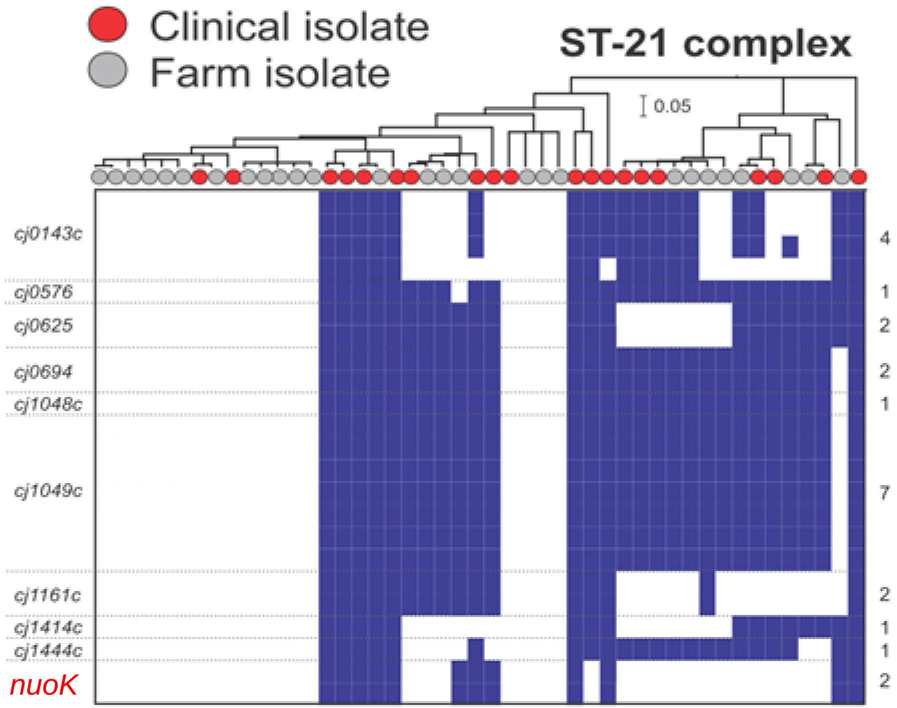
food-borne disease
(*Campylobacter*)

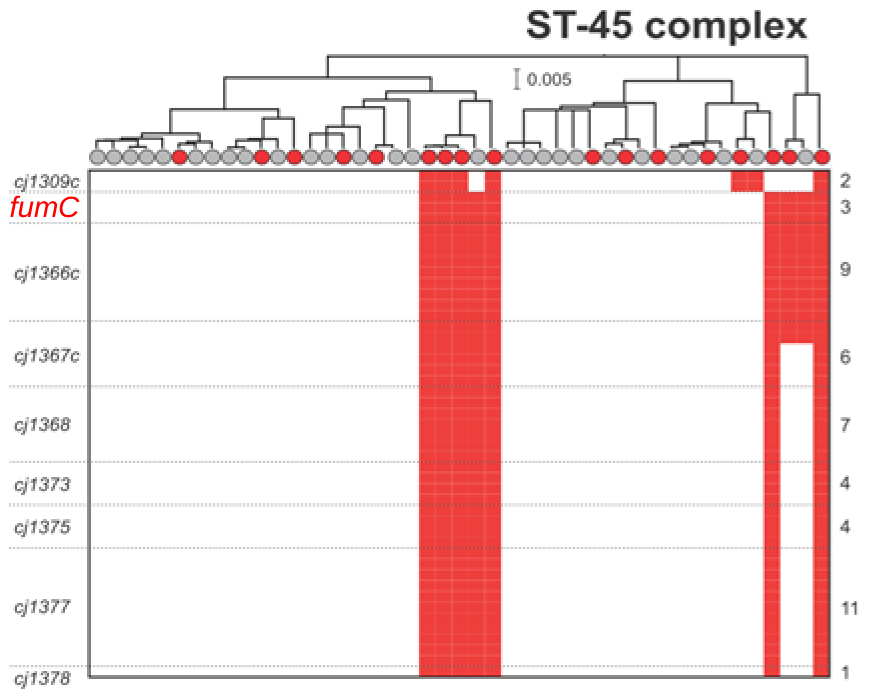
# How do *Campylobacter* genomes change to cause diseases in human?



Previously, “similar”

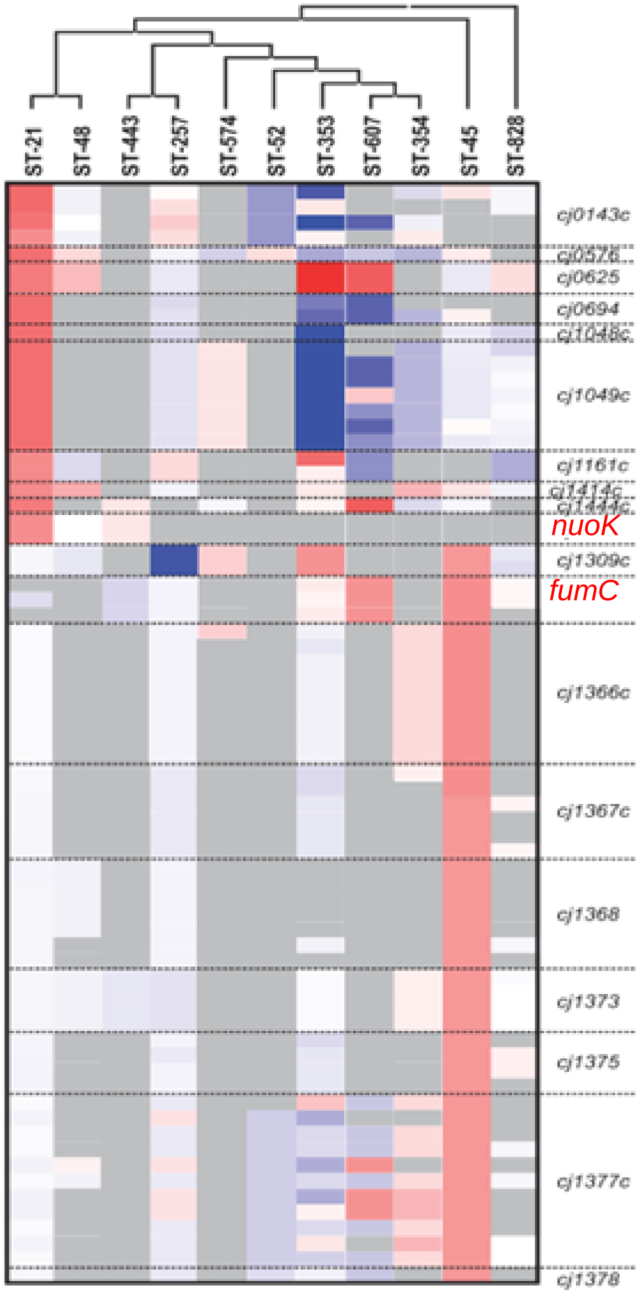
Discovery in two major lineages

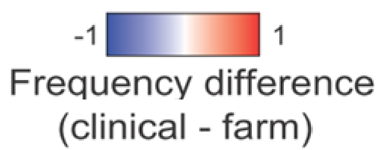


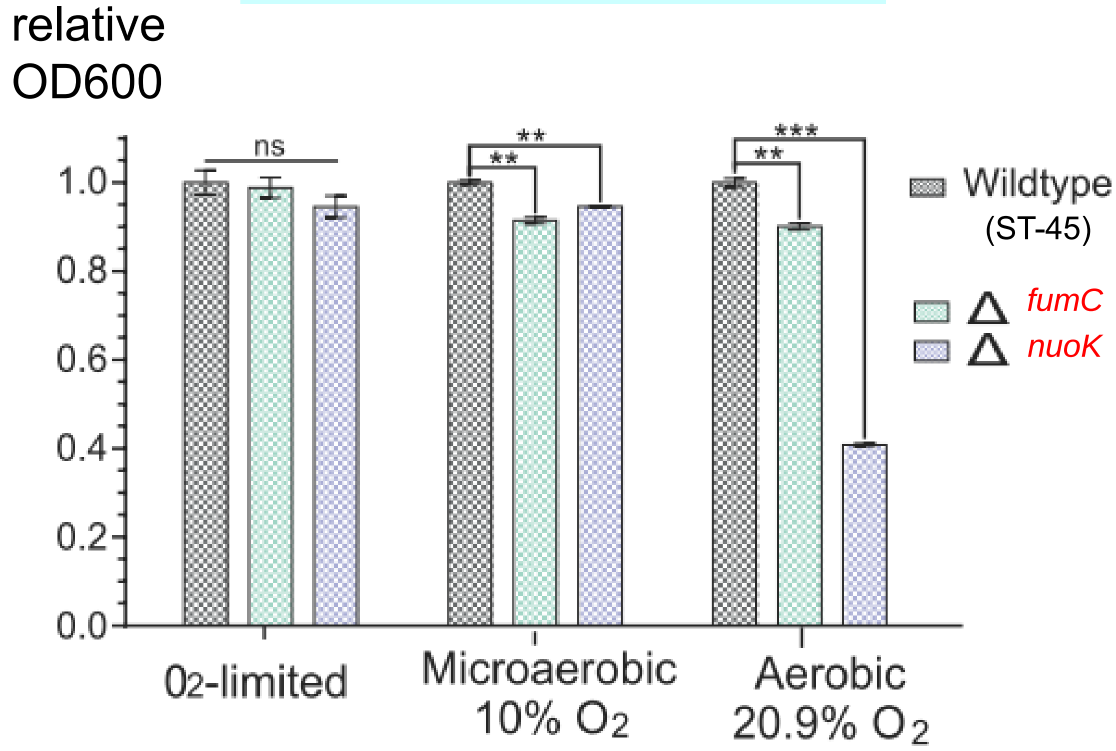

Disease-associated SNPs: 32-46% frequency increase ($P < 5x10^{-4}$)

# Validation: function revealed by knock-out



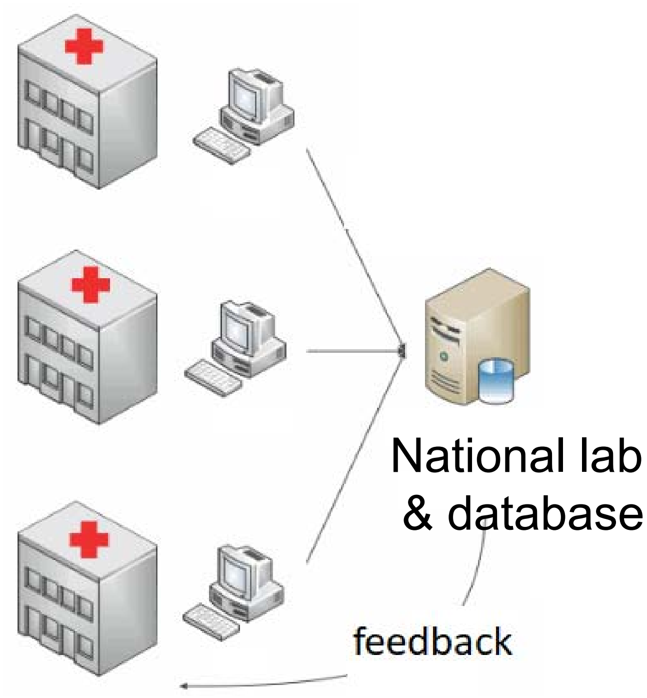

For aerobic survival

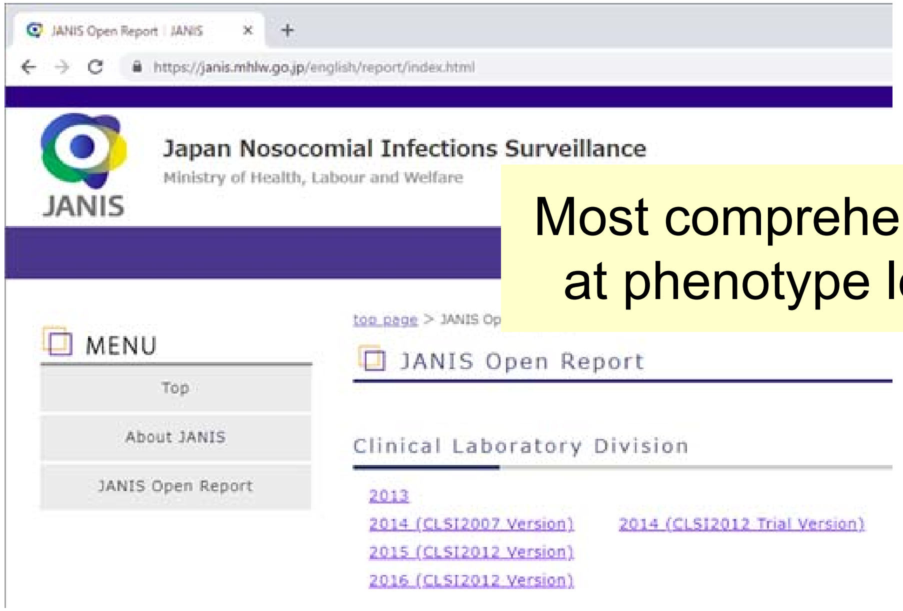



Yahara*, Meric* et al (2017),
*Env. Microbiology*

# Next direction

# National surveillance of antimicrobial resistance




Most comprehensive at phenotype level

all data of bacterial culturing and drug susceptibilities form > 2000 hospitals

Collection and genome sequencing of **isolates** satisfying specific criteria

genome — phenotype — patient's prognosis

# Hazard level in CDC and WHO priority list

HAZARD LEVEL
URGENT

*Clostridium difficile*
Carbapenem-R Enterobacteriaceae (CRE)
Cephalosporin-R *Neisseria gonorrhoeae*

1st (2009) &
a new type (2016)
in Japan

## Priority 1: CRITICAL

- *Acinetobacter baumannii*, carbapenem-resistant
- *Pseudomonas aeruginosa*, carbapenem-resistant
- *Enterobacteriaceae*, carbapenem-resistant, ESBL-producing

## Priority 2: HIGH

- *Enterococcus faecium*, vancomycin-resistant
- *Staphylococcus aureus*, methicillin-resistant, vancomycin-intermediate and resistant
- *Helicobacter pylori*, clarithromycin-resistant
- *Campylobacter* spp., fluoroquinolone-resistant
- *Salmonellae*, fluoroquinolone-resistant
- *Neisseria gonorrhoeae*, cephalosporin-resistant, fluoroquinolone-resistant

# Determinant of cephalosporin resistance

- Altered penicillin binding protein
  - encoded by mosaic *penA* arising from recombination

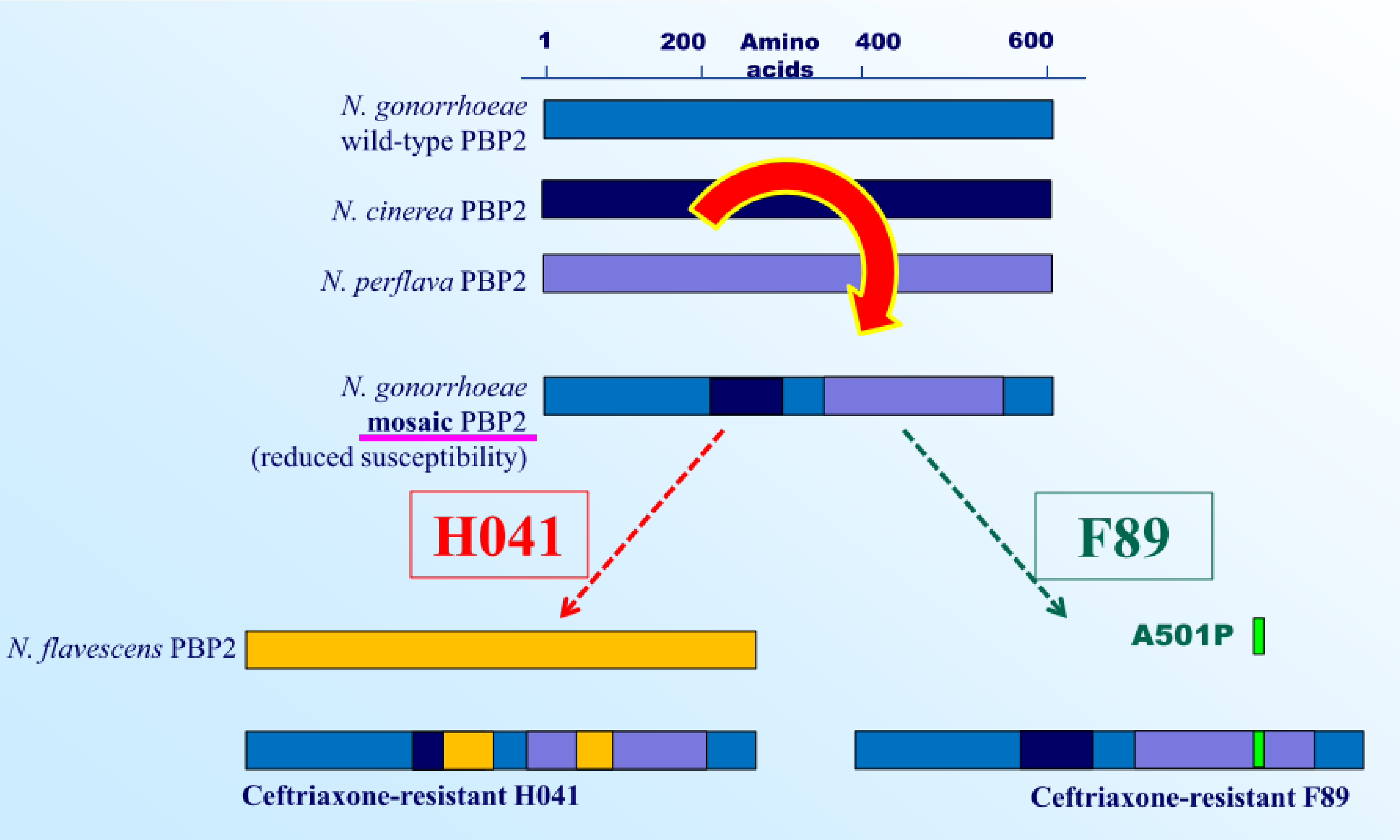
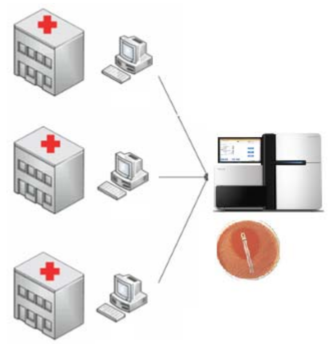



Figure from David Whiley

# How is it spreading and evolving in Japan as a region of global health concern?

- Surveillance has <u>not been based on genome</u>



- Unanswered questions:
  - type & distribution of resistance determinants?
  - the extent to which these determinants can explain the observed phenotypic resistance?
  - population structures at the genomic level?
    - sub-lineage exhibiting unusual drug susceptibility?

Yahara et al (2018), *Microbial Genomics*

# Acknowledgments 1/2

- Univ. Bath
  - Daniel Falush
  - Sam Sheppard
  - Sheppard Lab
- Univ. Warwick
  - Xavier Didelot
- Univ. Oxford
  - Azim Anzari
  - Martin Maiden
  - Maiden Lab
- Karolinska Institutet
  - Kaisa Thorell
- Univ. Antwerp
  - Annemieke Smet
- Univ. Tokyo & NIBB
  - Ichizo Kobayashi
  - Kobayashi Lab
  - Ikuo Uchiyama
- Ohita Univ.
  - Yoshio Yamaoka
- International Institute of Molecular and Cell Biology
  - Janusz M. Bujnicki
  - Bujnicki Lab
- Leibniz Institute
  - Jan P. Meier-Kolthoff
- Joint Genome Institute
  - David Paez-Espino

# Acknowledgments 2/2

- National Institute of Infectious Diseases
  - Makoto Ohnishi
  - Ohnishi Lab
  - Masato Suzuki
  - Keigo Shibayama
  - Motoyuki Sugai
  - Colleagues in Antimicrobial Resistance Research Center

- Antibiotic-Resistant Gonorrhea Study Group
- All contributors to the national surveillance of antimicrobial resistance

- Super computing systems
  - Univ. Tokyo, HPC Wales, and National Institute of Genetics

- JSPS Fellowships & Grants-in-Aid (KAKENHI)