

Molecular Epidemiology

How Many Isolates Does it Take?

Dörte Döpfer¹, R.N. Zadoks², W. Buist¹, and B. Engel¹

- (1) Quantitative Veterinary Epidemiology Group, Animal Sciences Group of Wageningen UR, The Netherlands
- (2) Department of Food Science and Quality Milk Control Services, Cornell University, Ithaca/NY, USA



Introduction

- **genetic heterogeneity** among multiple isolates from one sample is a common problem in microbiology
 - the number of isolates pheno/genotyped is often determined by:
 - **convenience**
 - **habit**
 - **(lack of) resources**
- there is a **need to optimise the sample sizes** in order to make optimal usage of available resources.

- genetic heterogeneity in multiple isolates of one sample



genetic
homogeneity



genetic
heterogeneity

- How many isolates have to be genotyped in order to be 95% confident that all genotypes are found?

Examples of Applications

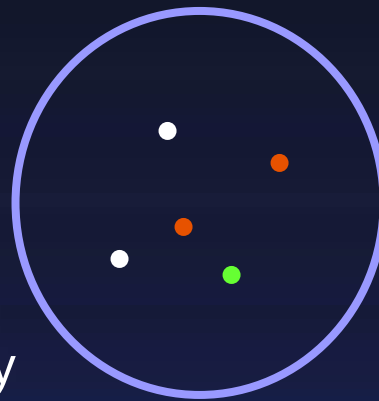
- Ribotyping of *Listeria monocytogenes* in smoked fish plants
- *rpoB* sequencing of sporeformers in dairy processing plants
- PFGE of *S. uberis* from environmental and milk samples
- serotype of *E. coli* in fecal pats
- Antimicrobial resistance of *S. aureus* in bulk tank milk

- Any combination of organism, source and typing technique

- the statistical problem: genetic heterogeneity in multiple isolates of one sample



genetic
homogeneity



genetic
heterogeneity

- a statistical “occupancy problem”:
 - how many ways are there to distribute 5 jackets over 10 hangers in a cloakroom?
 - (Johnson & Kotz)
- need for the probability of finding all genotypes
 - while analyzing a fraction of isolates from one sample

Bayesian approach:

- calculate the probability of finding j genotypes given that i genotypes are present and N isolates are characterized (Johnson & Kotz)

$$P_{N(j/i)} = \binom{i}{j} \sum_{r=0}^j (-1)^r \binom{j}{r} \left(\frac{j-r}{i} \right)^N$$

- use real-world data
- consult an expert...
 - “prior” knowledge:
 - expert expects an average of 5 genotypes in 5 isolates,
 - A uniform Dirichlet distribution was specified as a prior distribution for the marginal probabilities for the true number i of strains ($i = 1 \dots k$) of *S. uberis* in a random sample:
 - $(\theta_1 \dots \theta_k)' \sim \text{Dirichlet}(1/k \dots 1/k)$.
 - Assumption: all isolates are equally likely to be found
 - Altekruse et al. 2003, Singer et al. 2000

Case Study

- ribotyping of *Streptococcus uberis* from environmental samples

Sample	Description	# Isolates tested	# ribotypes detected
1-3	Mud, doorway	5	4
2-6	Lying area, pasture	5	3
3-3	Mud, trough 1	5	4
3-4	Mud, trough 2	7	2
3-5	Lying area, pasture	8	5
3-6	Lying area, pasture	7	4
6-5	Mud, trough 3	6	5
8-9	Mud, trough 4	5	3
9-5	Lying area, pasture	5	5

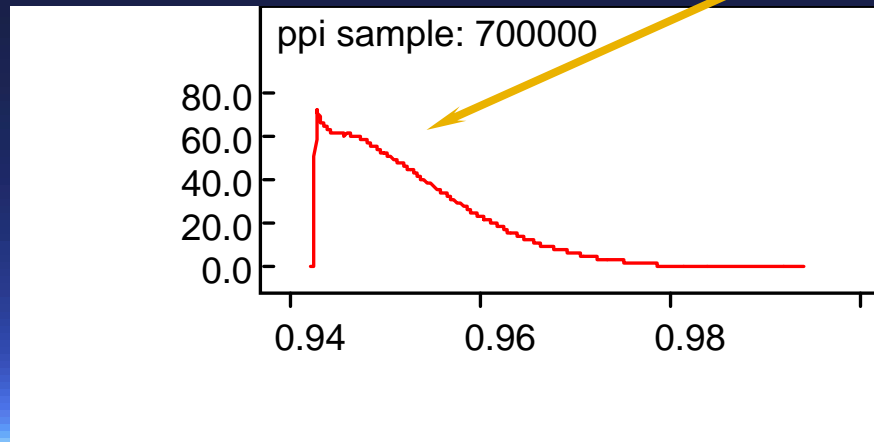
■ assemble.....

- expert opinion as prior knowledge,
- data and
- probabilities

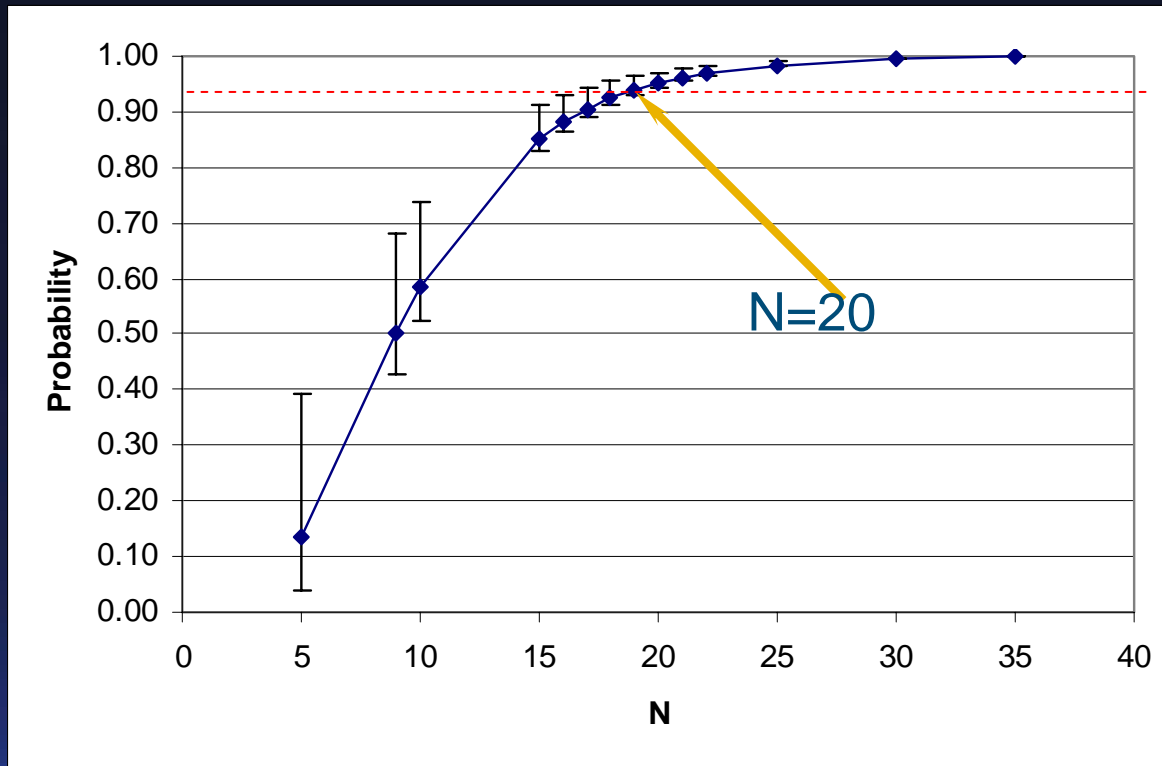
- to calculate the **posterior likelihood distribution** for finding all ribotypes in a given sample while analyzing a fraction N

$$p(N) = \sum_{i=1}^k \theta_i P_N(i | i)$$

- using a Gibbs sampler (Spiegelhalter et al. 2000)



- the results:
- for calculating sample sizes for multiple isolates of one environmental sample



Probability (with 95% CI) of finding all strains of *S. uberis* when ribotyping N isolates per sample

Implications

- 100 environmental samples

Technique	Cost/test	1 isolate	20 isolates
RAPD	\$10	\$1,000	\$20,000
Ribotyping	\$70-135	\$7,000-\$13,500	\$140,000-\$270,000
PFGE	\$20-80	\$2,000-\$8,000	\$40,000-\$160,000
MLST	\$30-70	\$3,000-\$7,000	\$60,000-\$140,000

Fecal Samples vs. Soil Samples

■ Environmental pathogens from Dairy Farms

		<i>S. uberis</i>	<i>L. monocytogenes</i>
Soil	n (samples)	44	10
	> 1 strain	86%	50%
	# of strains/sample	2.5	2.0
Fecal	n (samples)	17	10
	> 1 strain	59%	90%
	# of strains/sample	1.7	2.2

Presence vs. Absence

- How many isolates/tests does it take to prove presence?
- How many isolates/tests does it take to prove absence?

- E. coli O157:H7 – well established tests, PCR
- “Gas-forming spoilage organism”
 - gas-formation in cottage cheese
 - specific farm blamed
 - “Prove that the farm was not the source”
- Product testing
- Raw milk testing
 - how many farms/tanks/trucks/batches
 - how many isolates per sample
 - budget
 - legal implications

Considerations

- Cost vs. quality of typing methods
 - *E. coli* in bulk tank milk
 - world wide reference collection of *Listeria monocytogenes*
- Budget constraints
 - \$ for one large project or multiple smaller projects
 - few samples in great detail or many samples superficially
- How much detail is relevant? General picture vs. finding last reservoir
 - *S. uberis* in farm samples
 - *L. monocytogenes* in food processing plant

Conclusions

- **general problem** in microbiology
- applicable to surveys with the purpose of genotyping, etc.
- **use stepwise iterations to update sample sizes in surveys**
- **optimise the allocation of resources** using expert opinion, real-world data and probability methods
- “How many \$\$\$ should I budget in my grant proposal, control program, HACCP system, etc?”