# Pathogen Detection – Salmonella

Advances in DNA sequencing technology have made it possible for scientists all over the world to sequence complete microbial genomes rapidly and efficiently. Access to the DNA sequences of entire microbial genomes offers new opportunities to analyse and understand microorganism at the molecular level. Whole-genome sequencing allows us to detect pathogens in biological tissues and study variations in gene expression in response to the pathogenic invasion. These responses help in designing novel approaches for microbial pathogen detection and drug development. Identification of certain microbial pathogens as etiologic agents responsible for chronic diseases is leading to new treatments and prevention strategies for these diseases.  
Pathogen Detection Mechanism

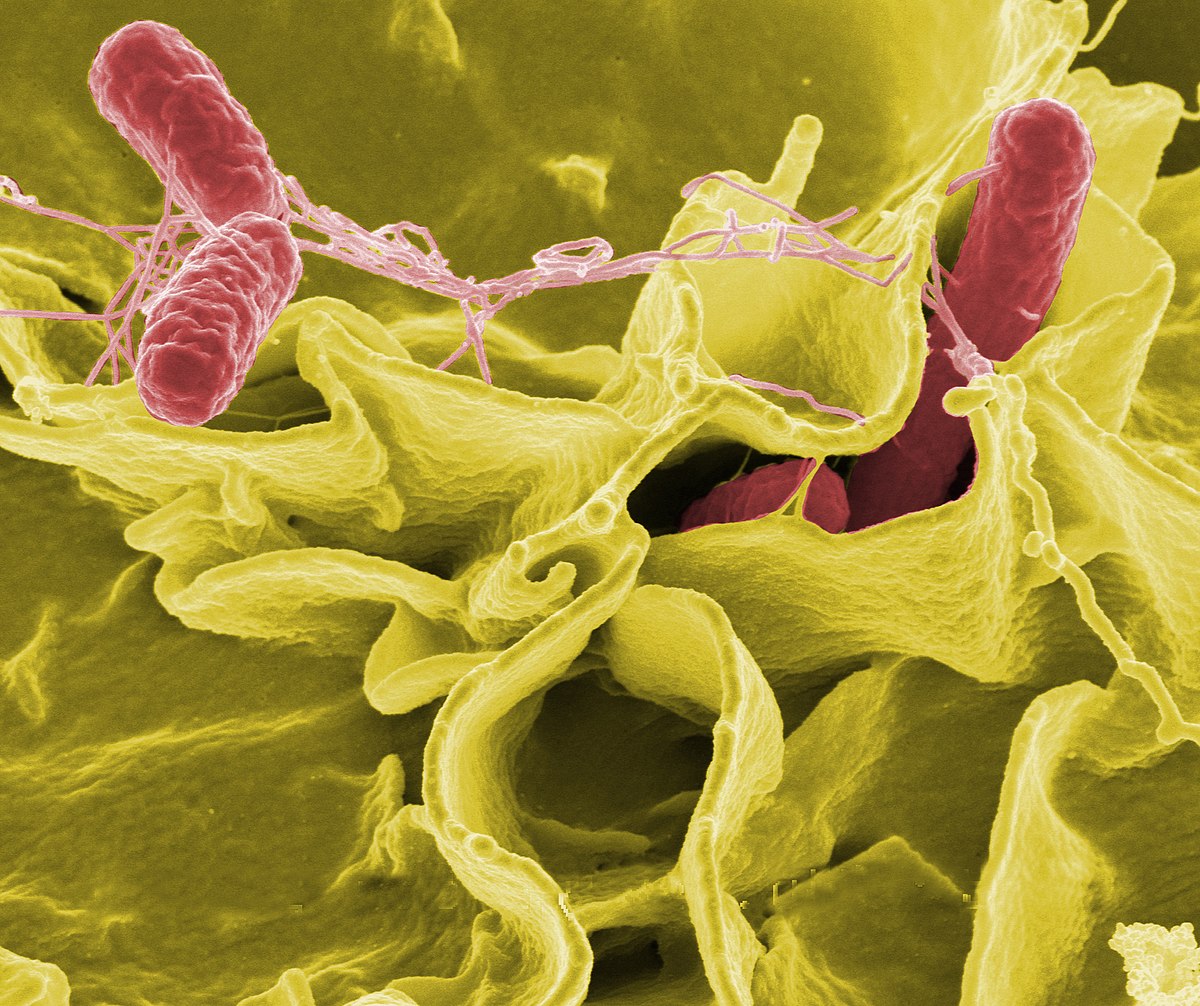


Figure 1: Colour-enhanced scanning electron micrograph showing Salmonella enterica (red) invading human cells.

*Salmonella enterica* (formerly *Salmonella choleraesuis)* is a rod-headed, flagellate, facultative anaerobic, Gram-negative bacterium and a species of the genus Salmonella. A number of its serovars are serious human pathogens; many of them are (more specifically) serovars of *Salmonella enterica*.

Most cases of salmonellosis are caused by food infected with S. enterica, which often infects cattle and poultry, though other animals such as domestic cats and hamsters have also been shown to be sources of infection in humans.

**Dataset Description**

The dataset pathogen\_detection\_salmonella.csv describes 558,820 observations of *Salmonella enterica* taken from environmental sampling. The data includes information about the date and broad location of the sample, along with the source of the sample. There is also detailed genetic information about the sample, including serotype, Isolate, SNP cluster and AMR Genotypes. Details about the most relevant features are given below.

|  |  |  |
| --- | --- | --- |
| Column | Data Type | Description |
| “Strain” | String/Category | The unique code identifying the strain of the isolated sample. |
| “Serovar” | String/Category | The serovar of the isolated sample. |
| “Create date” | String | The date of sampling, formatted as yyy-mm-dd |
| “Location” | String/Category | The broad location of the sample (e.g. USA:NY) |
| “Isolation source” | String | The source of the sample, e.g. “Chicken Breast”. |
| “SNP cluster” | String/Category | The SNP cluster targeted in the serotyping. |
| “AMR genotypes” | String/Category | Presence of genes associated with antimicrobial resistance (see below). |

\*The data in the “AMR genotypes” column is structured as a list of genes, along with a “COMPLETE” or “MISTRANSLATION” tag. For example, one entry might be “mdsA=COMPLETE,mdsB=MISTRANSLATION” which would indicate the presence of a complete sequence for gene “mdsA” and a partially mistranslated gene “mdsB”.