Module 3 Microbial Species Concepts

Evidence worksheet\_05 “Extensive mosaic structure”

Part 1: **Learning objectives:**

* Evaluate the concept of microbial species based on environmental surveys and cultivation studies.
* Explain the relationship between microdiversity, genomic diversity and metabolic potential
* Comment on the forces mediating divergence and cohesion in natural microbial communities

**General Questions:**

• *What were the main questions being asked?*

* Is the difference in the genome sequence the cause of pathogenicity in difference strain of *E. coli*?
* Three-way genome comparison of the CFT073, enterohemorrphagic *E. cli* EDL933, and laboratory strain MG1655 revealed that pathogen genomes are as different from each other as each pathogen is from the benign strain.
* Comparison of 3 different strains of E. coli (2 uropathogenic, 1 non-pathogenic)
* Compare genetic similarity/OUT similarity between pathogenic and non-pathogenic strains
* Determine conserved pathogenicity associated/PAIs
* How later gene transfer contributes to the emergence of new uropathogenic E.coli strains

• *What were the primary methodological approaches used?*

* Different strains are collected from …
* Data are collected on Applied Biosystems ABI377 and 3700 automated sequencers.
* The genome sequence of difference strains are assembled by SEQMANII, and they used sequencing of opposite ends of linking clones, and several PCR-Based techniques and primer walking.
* The genome sequence of different strains are annotated in the multiuser, web-based annotation environment called MAGPIE, which defined and assigned automatic annotation for all ORFs, which then were checked individually and corrected.
* Predicted proteins were search against the nonredundant database by using BLAST.
* The assembly of DNA sequences from a shotgun library of CFT073 DNA fragments, combined with PCR strategies and primer walking experiments for finishing, resulted in a circular, 5 million-bp chromosomal sequence with seven times coverage.
* Distinctive codon usage is considered to be a hallmark of lateral gene transfer.

• *Summarize the main results or findings.*

* The CFT073 genome is circular.
* Although virulence plasmids are common to many *E. coli* isolates, they are not usually associated with uropathogenic strains, and non were found in CFT073.
* There are five cryptic prophage genomes in the CFT073 chromosome, none with sufficient genetic information to produce viable phage.
* 70% of the ORF previously identified as unique to either MG1655 or EDL933 are replaced with new genes specific to the uropathogenic isolate. The distrupted ORF resulted in detection of only 62 pesudogenes.
* Distinctive codon usage test reveals the CFT073 has 52 of 61 codons in island ORFs had frequency distributions significantly different from those in backbones with 95% confidence values.
* The CFT073 specific islands contain 2004 genes, of each only 204 also occur among the EDL9330 specific genes. Two-third of these island genes shared by EDL933 and CFT073 have unknown functions or associated with phage or insertion sequence elements. The remaining encode putative iron-uptake systems, a complex set of potential fatty acid biosynthetic enzymes, several adhesins, phosphotransferase system and ATP-binding cassette-type transport systems.
* Many island locations are at the same relative backbone position in the two pathogens although the island contents are unrelated
* 13 CFT073 and 10 EDL933 islands are closely associated with known tRNA genes.
* The differences in disease potential between enterophemorphaggic and ropathogenic *E. coli* are reflected specifically in the absence of genes in CFT073 for type III secretion system and phage- and plasmid- encoded virulence genes common to *E. coli* O157:H7 isolates. In CFT073, the strain-specific regions contain genes that encode specific fibrial adhesins, secreted autoransporters and phase switch recombinases.
* The ability to inhabit the different niches during an ascending urinary tract infection and cause particular pathologies at each site resides largely in the island gens specific to uropathogenic *E. coli*. The CFT073 genome sequence reveal many possible factors that may contribute to colonization of the urinary tract tissues and the disease:
  + Surface structure known as fimbriae or pili mediate specificity for and attachment to host cells, an essential event for host colonization.
  + They found that in CFT073, these proteins are highly divergent from those in MG1655 and EDL933, with amino acid sequence identities ranging from 53% to 81%, suggesting that the selective pressure on the expression of this pilus has varied among *E. coli* lineages.
  + Presumably, the variable sequences of the shared operons allows for the specificity of each adhesion to its individual target tissue.

Summary:

* Through long history of evolution, the ancestral backbone genes that define *E. coli* have undergone slow accumulation of vertically acquired sequence changes, but there are numerous newly introduced genes via independent horizontal gene-transfer events at many discrete site.
* The condon usage analysis supports that there are a set of backbone E. coli genes that have a shared codon bias that is not seen in the genes unique in each of the three genomes. This result in a mosaic genome structure in which newly acquired genes in each of the E. coli types are placed into a framework made of genes that distinguished E. coli from its closer relatives such S. enterica.
* Each type of *E. coli* possess combinations of island genes that confer its characteristic lifestyle or disease-causing traits, for example, the uropathogenic strains of E. coli, the island acquisition resulted in the ability to infect the urinary tract and bloodstream and evade host defense without compromising the ability to harmlessly colonize the intestine.
* Extraintestinal *E. coli* may be oligoclonal despite the apparent linkage relationships of a handful or virulent genes, and suggests that the ropathogenic *E. coli* may be as diverse as the intestinal strains.
* The sheer amount of unique DNA in each *E. coli* strain that can be explained by the frequent gain and loss of accessory genes suggests that careful reconsideration is due for defining species by a few phenotypic traits and low-resolution mapping.

• *Do new questions arise from the results?*

* The CFT073 and EDL933 genome sequences enable us to design far more discriminating tools for diagnosis of particular *E. coli* pathotypes that cause such a wide range of intestinal and extraintestinal disease.
* Do the pathogenity of E. coli can be inferred from the island gene?
* Comparing three genome, what is the similarity of backbone gene between three genome? How similar is the backbone?
* How can species be defined to account for frequent gain and loss of accessory gene -> cannot simply define species by phenotypic analysis and low resolution mapping gene?

• *Were there any specific challenges or advantages in understanding the paper (e.g. did the authors provide sufficient background information to understand experimental logic, were methods explained adequately, were any specific assumptions made, were conclusions justified based on the evidence, were the figures or tables useful and easy to understand)?*

* The author assumed the reader to have detail understanding of strains of *E. coli*
* The figures and finding requires the reader to have prior understanding of certain genes

Part 2: **Learning objectives:**

* Comment on the creative tension between gene loss, duplication and acquisition as it relates to microbial genome evolution
* Identify common molecular signatures used to infer genomic identity and cohesion
* Differentiate between mobile elements and different modes of gene transfer

Based on your reading and discussion notes, explain the meaning and content of the following figure derived from the comparative genomic analysis of three *E. coli* genomes by Welch et al. Remember that CFT073 is a uropathogenic strain and that EDL933 is an enterohemorrhagic strain. Explain how this study relates to your understanding of ecotype diversity. Provide a definition of ecotype in the context of the human body. Explain why certain subsets of genes in CFT073 provide adaptive traits under your ecological model and speculate on their mode of vertical descent or gene transfer.

