Joint Graduate Seminar on 11 Dec 2009

Comparative Genomic Analysis for a Clinical GBS Group in the Streptococcus Pangenome Model

M.Phil. Student: Haokui ZHOU
Supervisor: Prof. Guoping ZHAO
Clinical Microbial Genomics Laboratory
Department of Microbiology
Prince of Wales Hospital
The Chinese University of Hong Kong

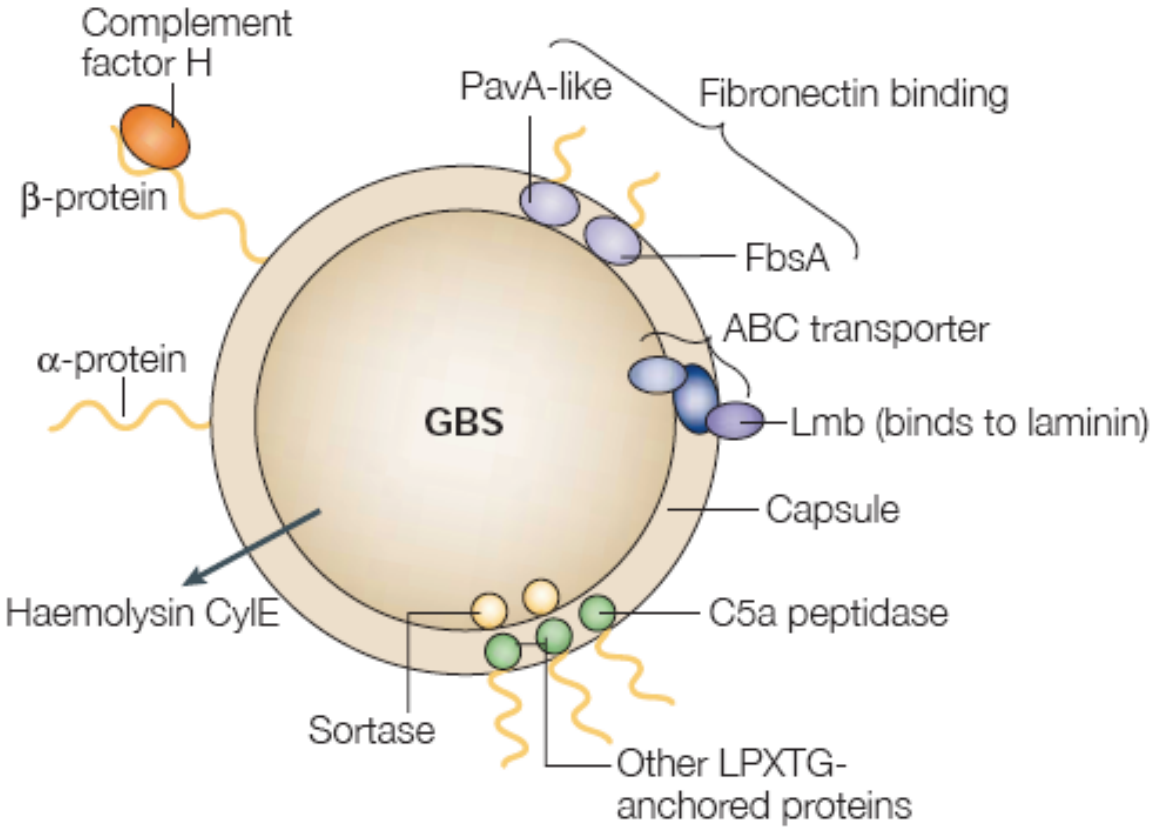
GBS in focus

Pathogenesis & Epidemiology

- Group B streptococci (GBS), or *Streptococcus agalactiae*
  - commensal bacterium colonizing the intestinal tract of a significant proportion of the human population
  - but causes invasive infection in neonate (sepsis, meningitis, pneumonia, et al.)
  - also in some immunocompromised adults
  - also in animals
- Nine serotypes identified, and vary in distribution among human populations

The GBS species defined traditionally

- Mostly based on virulence factors
- Immunological antigens
  - Lancefield Group B antigen
  - Capsular polysaccharide
- Surface proteins
  - alpha protein
  - beta protein
  - Rib protein
  - …



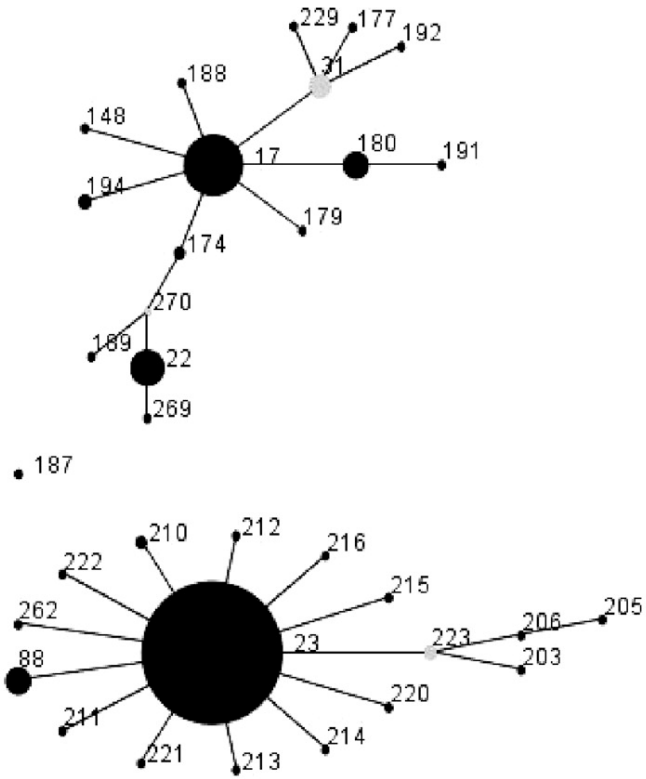
Mitchell, T.J., Nature Reviews Microbiology 1 (3), 219-230 (2003)

MLST & Clonal complexes

aim for the population structures

- Multi locus sequence typing (MLST) profiles
  - sequences of internal fragments (400-500 bp) of seven house-keeping genes
  - sequence type defined from allelic profile
  - sequence types are grouped into clonal complexes by their similarity to a central allelic profile

- Only take into account the conserve fragments of the genome
- Not necessary related to virulence



USA GBS population
1995-1999

adapted from:
Bohnsack, J.F. *et al.*, *J Clin Microbiol* 46 (4), 1285-1291 (2008)

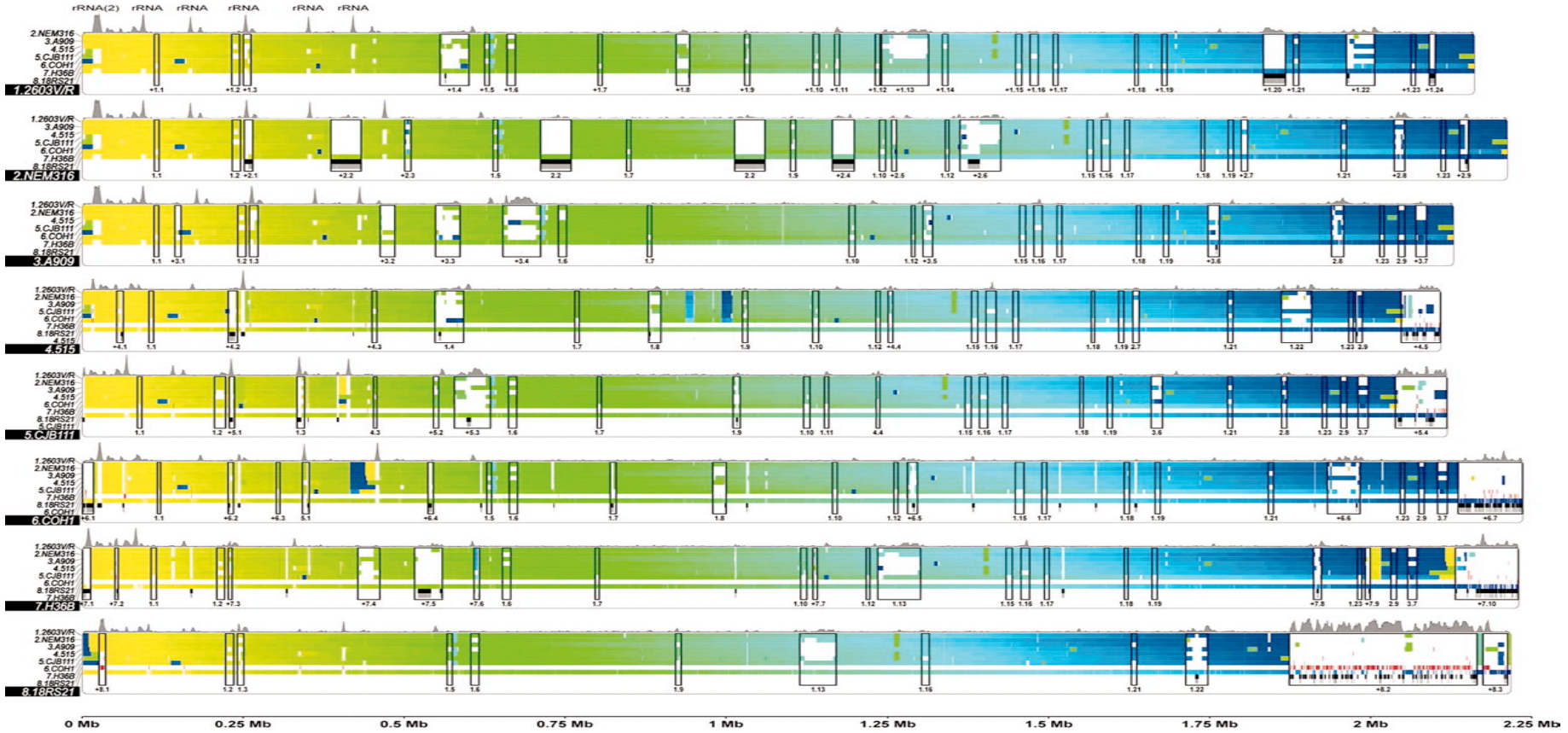
Typing with more and more markers…

- Capsular polysaccharide
- Surface proteins
  - C alpha protein (bca), C alpha-like protein 1, 2 and 3 (alp1, alp2, alp3), Rib protein (rib)
- mobile genetic elements
  - IS861, IS1548, IS1381, ISSa4, ISSag1, ISSag2, GBSi1
- Antibiotic resistance-related genes
  - aad, aph, erm^{B}, erm^{TR}, int-Tn, mef, mre, tet^{M}, tet^{O}

- BUT, when "examined 912 human GBS isolates in which 18 distinct molecular markers"
  - "While some molecular epidemiological markers are important in defining GBS clusters, a definitive predictive relationship between the molecular markers and clinical outcomes may be lacking" (Lin, F. et al., *Pathology* 41 (6), 576-581 (2009))

When there are multiple genomes available …

The proposing of the “pangenome” concept



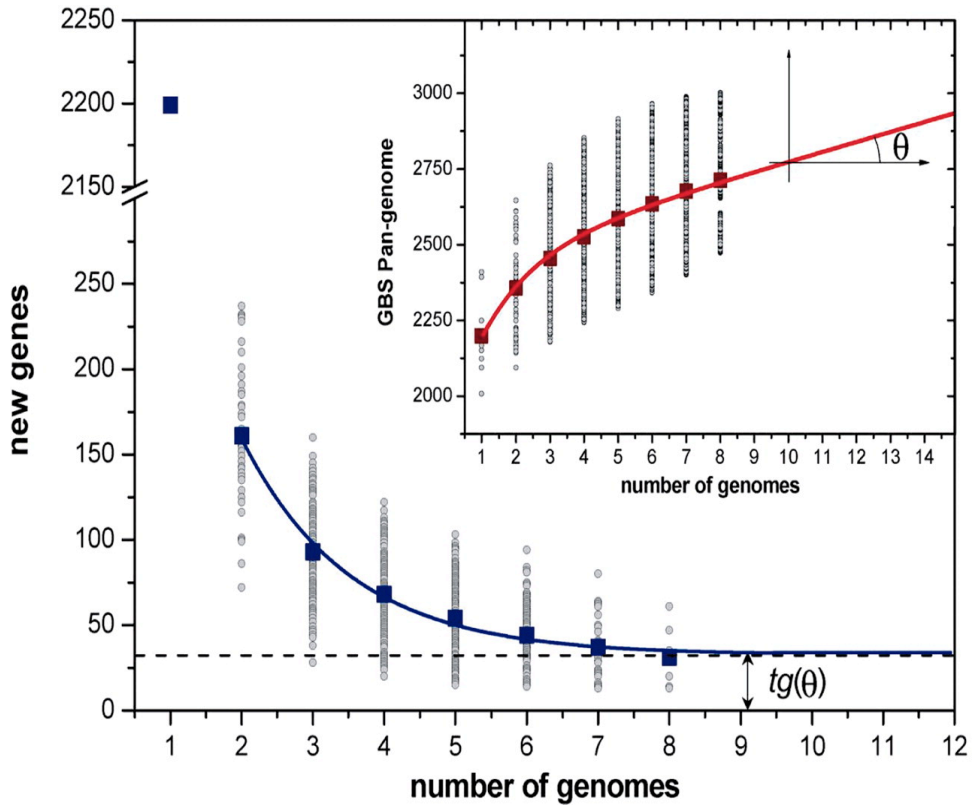
S. agalactiae species can be described by a pan-genome consisting of a core genome shared by all isolates, accounting for ≈80% of any single genome, plus a dispensable genome consisting of partially shared and strain-specific genes.

Tettelin, H. *et al.*, *PNAS* 102 (39), 13950-13955 (2005)

The "opened" pangenome, but how can it be endless?

- "The model predicts that for every new GBS genome sequenced, an average of 33 new strain-specific genes will be identified and added to the pan-genome."

- "This finding suggests that the GBS pan-genome is open and that its size grows with the number of independent strains sequenced."

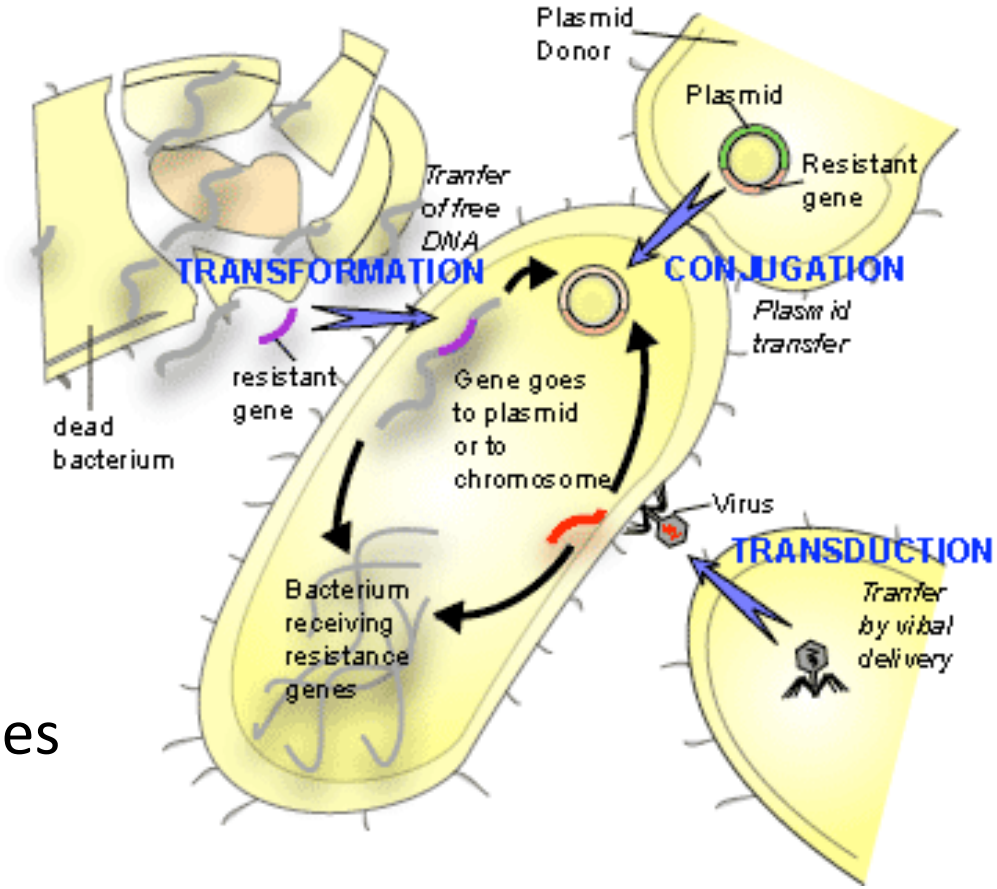


Tettelin, H. *et al.*, *PNAS* 102 (39), 13950-13955 (2005)

The driving force in action

Mobile genetic elements and horizontal gene transfer

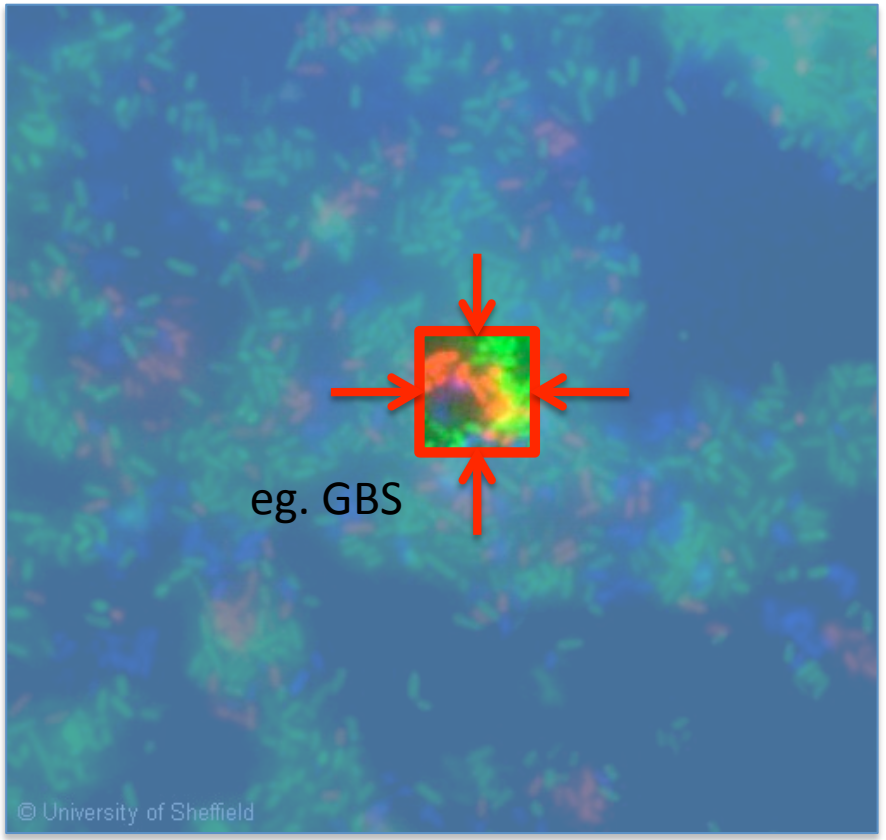
- What is in the dispensable genome?
  - genes associated with mobile and extrachromosomal elements
  - suggesting lateral gene tranfer events
  - most strain specific genes in genomic islands



Grace Kim, Aug 11, 2006. The Science Creative Quarterly 3

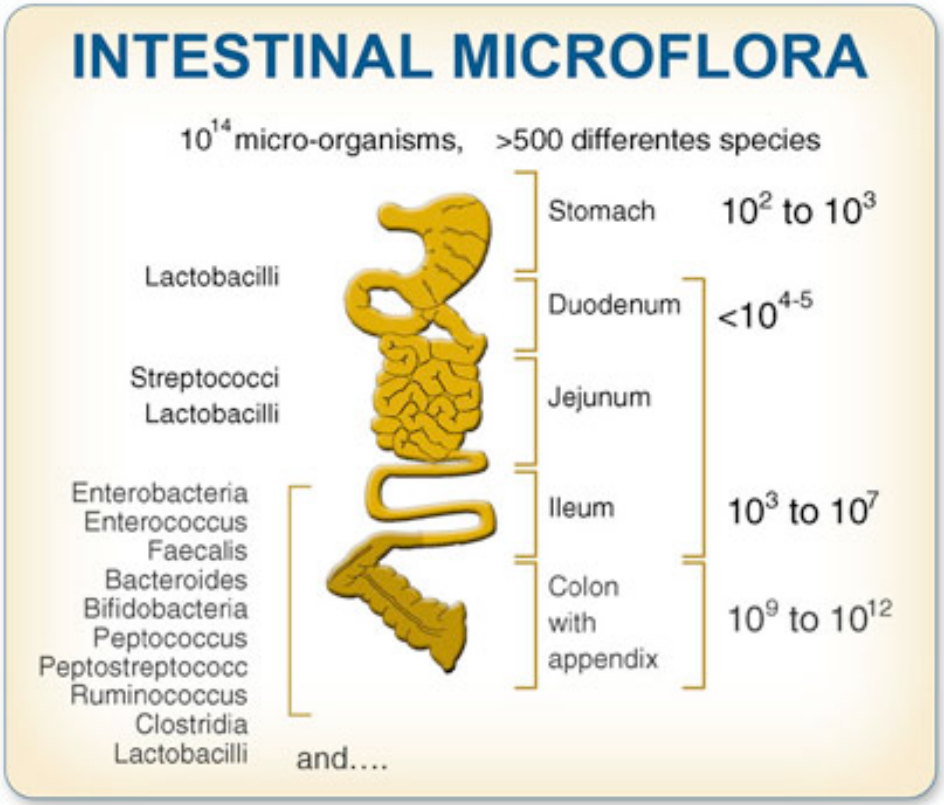
Where is the “open source”?

- Of course, GBS is not isolated, but living in an open environment, a microbial community.
- And a community predominated genetic flow among microbial genomes via HGT, mediated mostly by the associated viruses community.
- So the species concept is uselessness unless incorporating the collective behavior of the community



The ignored side from human microbiome

- There are a great number of microbes inhabiting inside or outside the human body.
- about 100 trillion cells that outnumber human cells 10 to 1 (Savage, 1977)
- The human microbiome project (HMP)
- There is a great diverse viral community associated (Tao Zhang et al, 2006)
- Antibiotic Resistance Reservoir in the Human Microflora (Sommer et al, 2009)



http://www.customprobiotics.com/

Collective emergence in evolution

- *"The emerging picture of microbes as gene-swapping **collectives** demands a revision of such concepts as organism, species and evolution itself."*

  ...

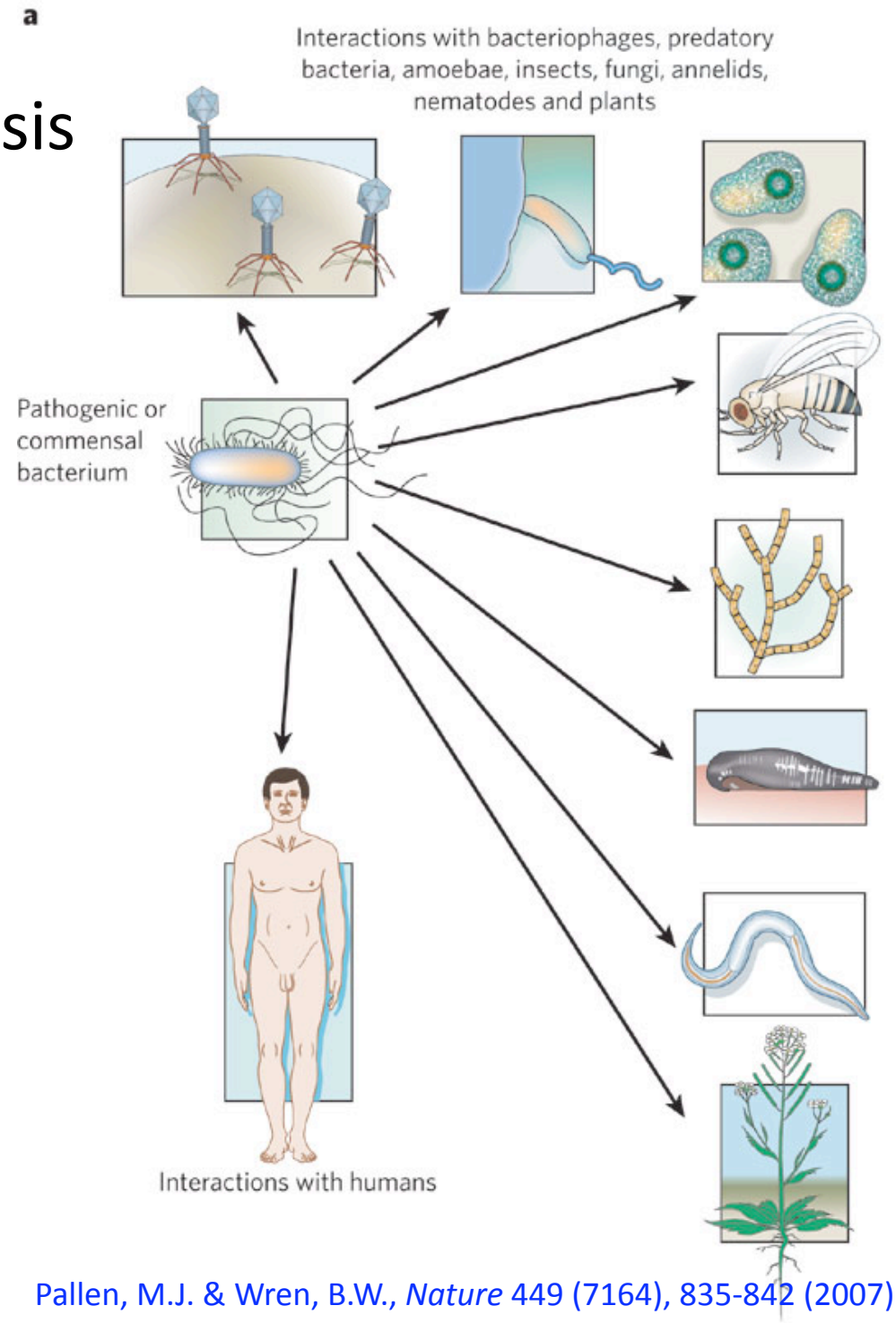
  *"the molecular reductionism that dominated twentieth-century biology will be superseded by an interdisciplinary approach that embraces **collective phenomena**."*

--- Goldenfeld, N. & Woese, C.,
Nature 445 (7126), 369 (2007)

Emergence of pathogenesis
the “eco-evo” view

- A pathogen can emerge from non-pathogen by acquiring virulence gene on plasmids, bacteriophages or pathogenicity islands.

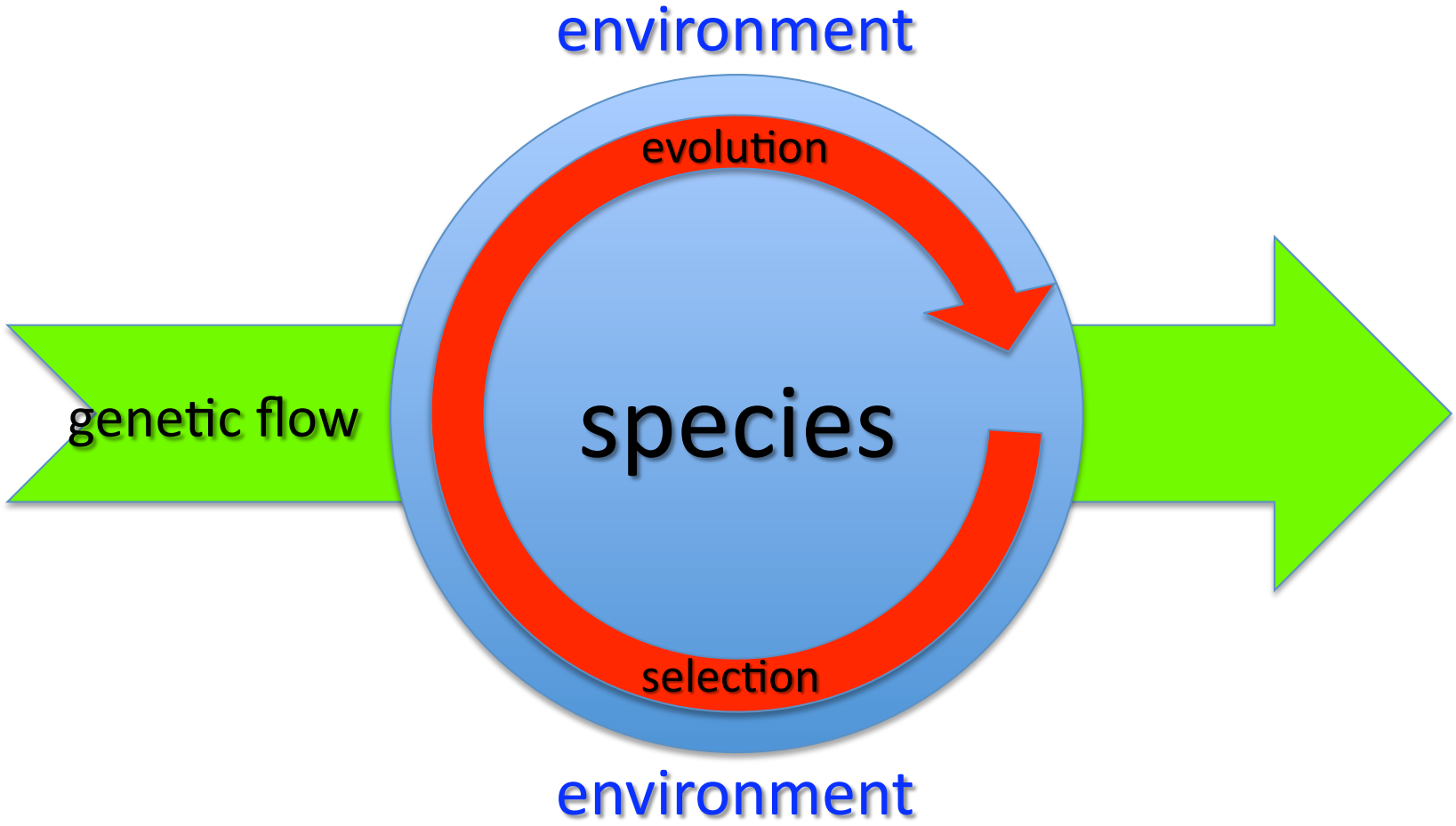
- Ecology analysis of the associated community, with evolutionary genomics, should be important to understand the mechanism of pathogenesis and the epidemiology dynamics.



Pallen, M.J. & Wren, B.W., *Nature* 449 (7164), 835-842 (2007)

Revision of the GBS pangenome

not endless in size, but endless in novelty



A math model learned from the genome database?

Back to GBS pangenome, a new way

- Survey draft genomes from a representative clinical and none-clinical collections.
- Categorize the associated mobile genetic elements.
- Incorporate the available human microbiome data.
- Comparative analysis of the dispensable part.
- Ecological analysis of the diversity and abundance.
- Finding patterns from ecological results, for example, phylogenetic association analysis

Thank you for your attention ☺