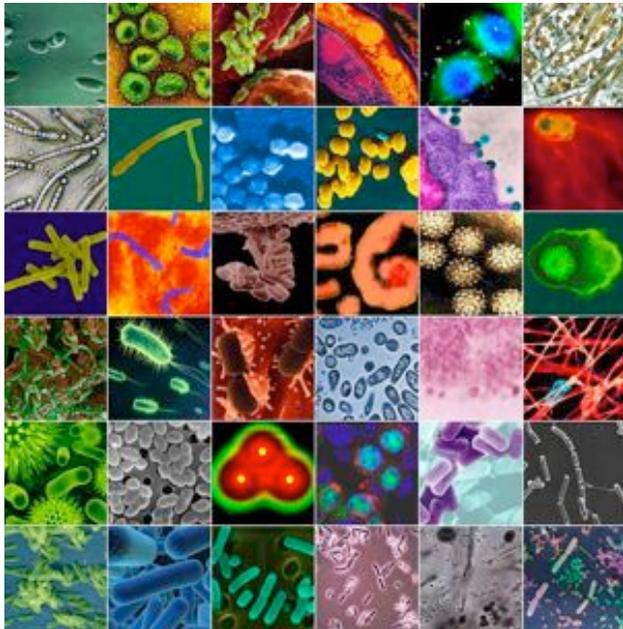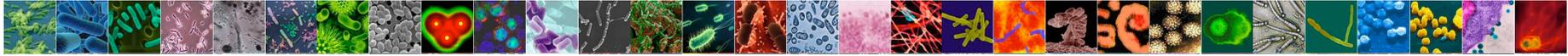# Evolutionary Genetics: What Use to Anyone?

Daniel Wilson

Nuffield Department of Clinical Medicine

17th June 2010

daniel.wilson@ndm.ox.ac.uk
www.danielwilson.me.uk

Gil McVean
Wellcome Trust Centre for Human Genetics
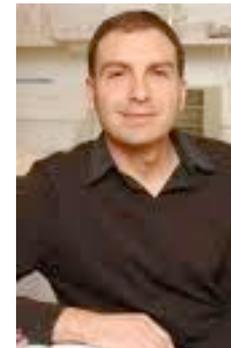
Martin Maiden
Department of Zoology

Paul Fearnhead
Lancaster University

Andrew Fox
Health Protection Agency

Molly Przeworski
University of Chicago

Lluis Quintana-Murci
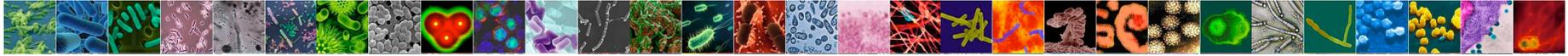Institut Pasteur de Paris
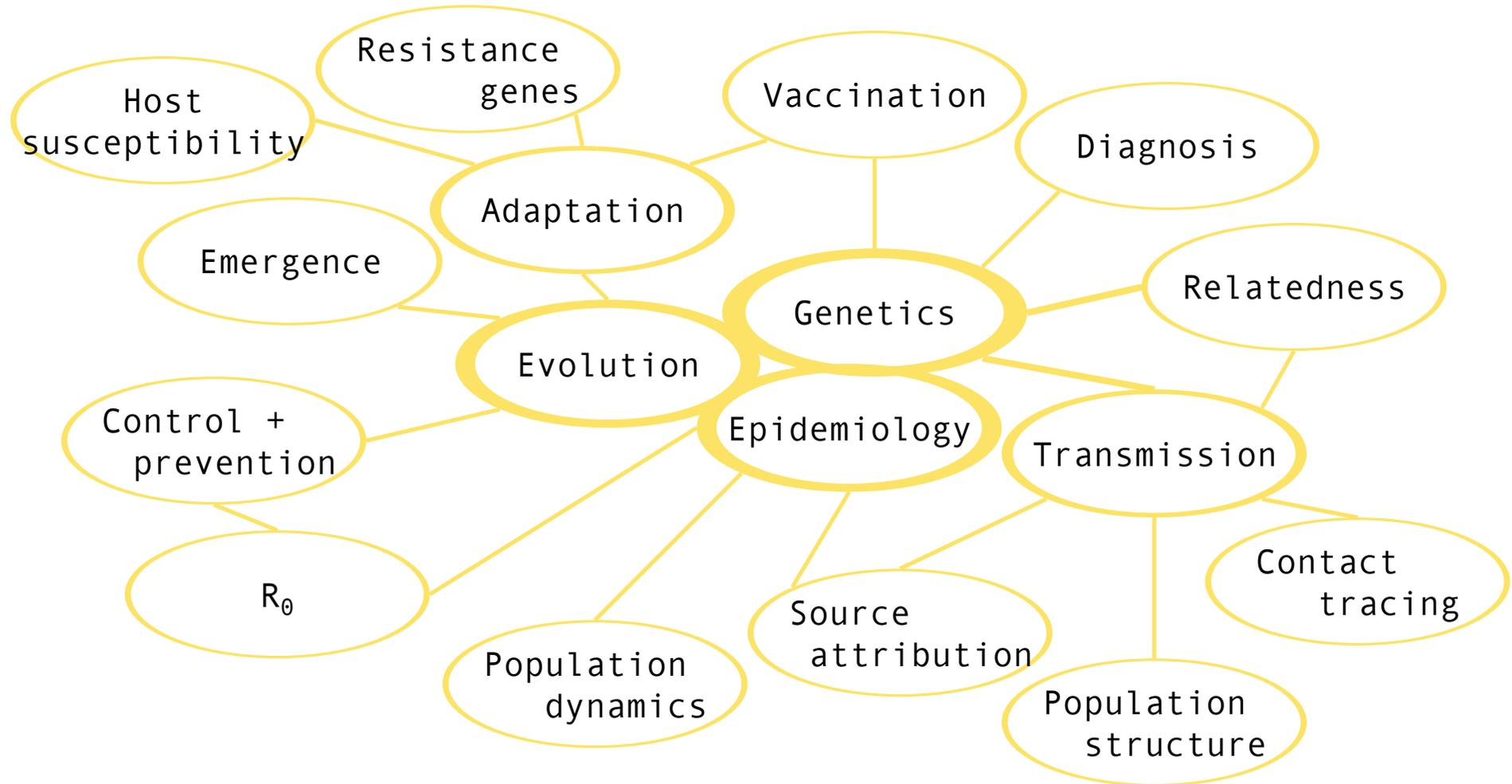
Richard Everitt
University of Bristol

Bethany Dearlove
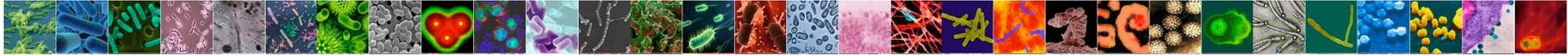University of Reading

Evolutionary Genetics: What Use to Anyone?

Evolutionary Genetics:
What Use to Anyone?

- Visualizing Relatedness

- Understanding Population Structure

- Attributing the Source of Infection

- Reconstructing Demographic History

- Unravelling the Genetic Basis of Disease

- Learning about Evolution

- Ad hoc vs. Model-Based Approaches
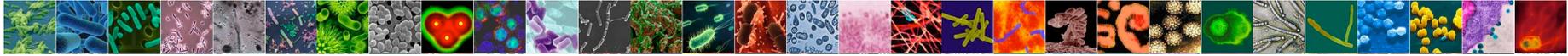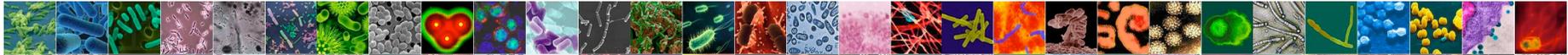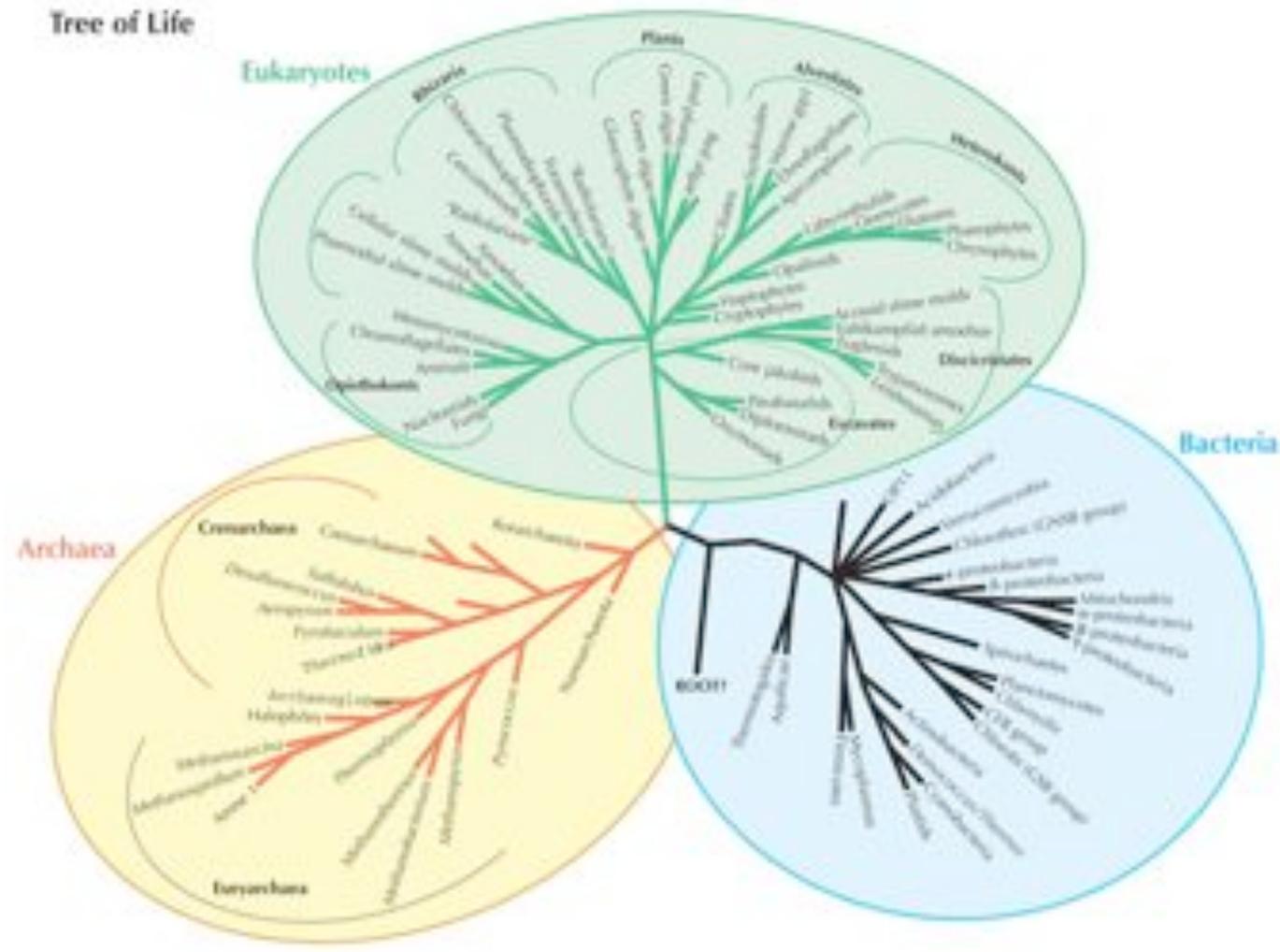
- Technical Challenges

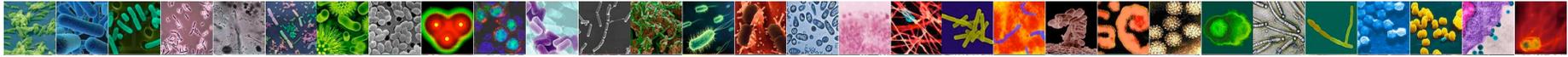# Evolutionary Genetics: What Use to Anyone?

- Visualizing Relatedness
- Understanding Population Structure
- Attributing the Source of Infection
- Reconstructing Demographic History
- Unravelling the Genetic Basis of Disease
- Learning about Evolution

- Ad hoc vs. Model-Based Approaches
- Technical Challenges

Visualizing Relatedness

Transmission Pathways of Foot-and-Mouth Disease Virus in the United Kingdom in 2007

Eleanor M. Cottam[1,2], Jemma Wadsworth[1], Andrew E. Shaw[1], Rebecca J. Rowlands[1], Lynnette Goatley[1], Sushila Maan[1], Narender S. Maan[1], Peter P. C. Mertens[1], Katja Ebert[1], Yanmin Li[1], Eoin D. Ryan[1], Nicholas Juleff[1], Nigel P. Ferris[1], John W. Wilesmith[3], Daniel T. Haydon[2], Donald P. King[1], David J. Paton[1], Nick J. Knowles[1]*
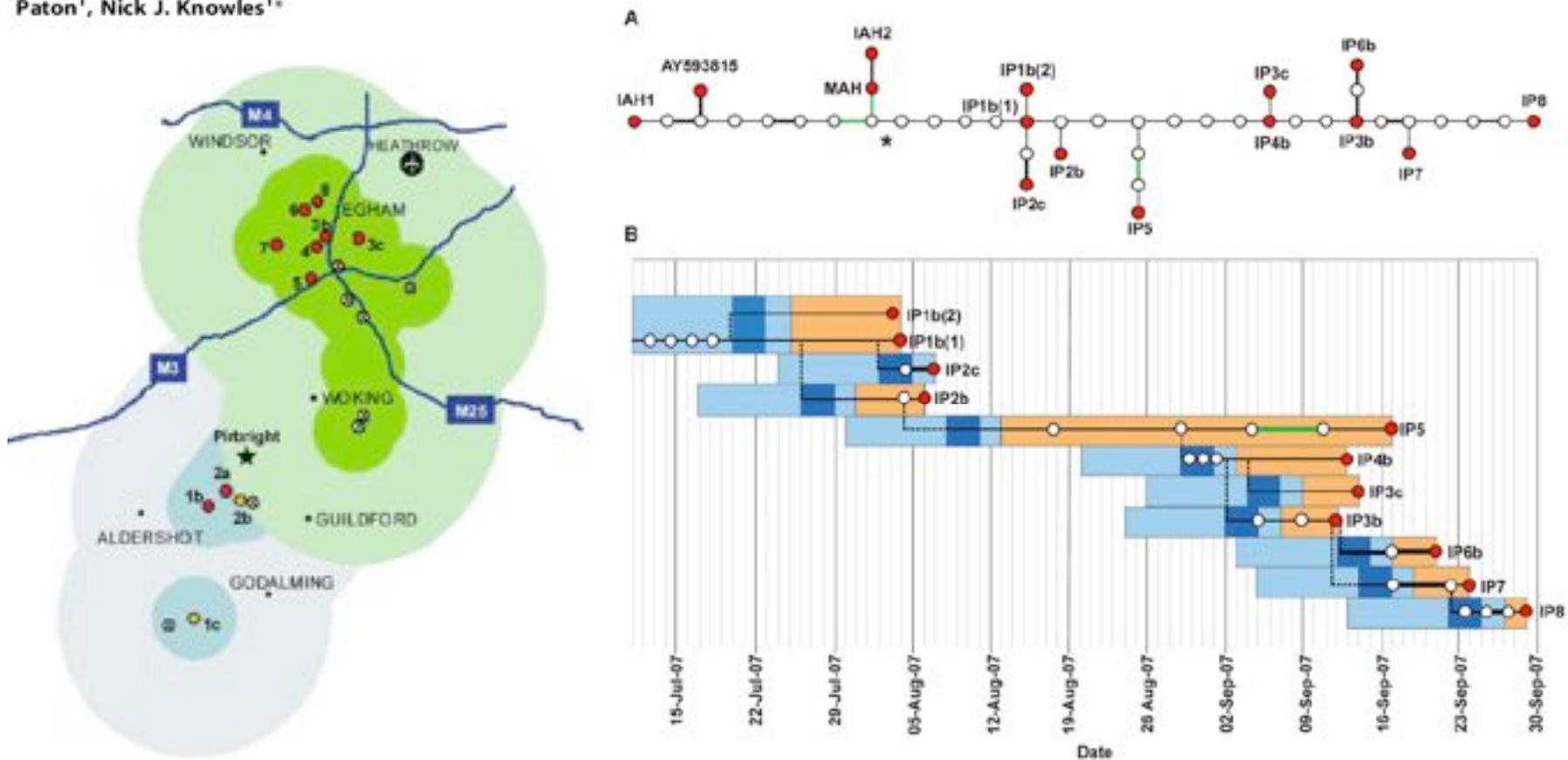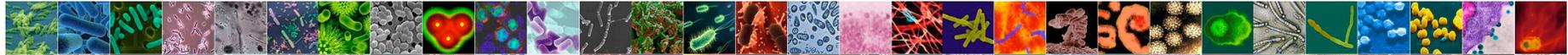

Figure 2. Analysis of sequence data. A) Statistical parsimony analysis by TCS [19] of complete genome sequences of 14 FMDVs;

Molecular evidence of HIV-1 transmission in a criminal case

Michael L. Metzker*[†], David P. Mindell[‡], Xiao-Mei Liu*[§], Roger G. Ptak[¶], Richard A. Gibbs*, and David M. Hillis[||]

*Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; [‡]Department of Ecology and Evolutionary Biol[ogy] Museum of Zoology, University of Michigan, Ann Arbor, MI 48109-1079; [¶]School of Dentistry, Biologic and Materials Sciences, University of Mich[igan], Ann Arbor, MI 48109; and **Section of Integrative Biology and Center for Computational Biology and Bioinformatics, University of Texas, Austin, TX 78712
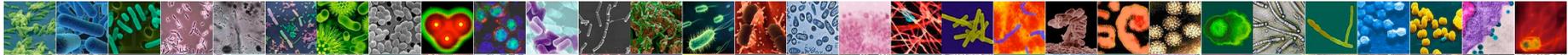
A gastroenterologist was convicted of attempted second-degree murder by injecting his former girlfriend with blood or blood-products obtained from an HIV type 1 (HIV-1)-infected patient under his care. Phylogenetic analyses of HIV-1 sequences were admitted and used as evidence in this case, representing the first use of phylogenetic analyses in a criminal court case in the United States. Phyloge- (13). This case was the first time that phylogenetic analys[es] used as evidence in a United States criminal proceedin[g] present the phylogenetic evidence that constituted p[art of the] prosecution's case that resulted in the conviction of the gastroenterologist on the charge of attempted seco[nd-degree] murder.

a

100% of bootstrap replicates place victim sequences within patient sequences

Visualizing Relatedness
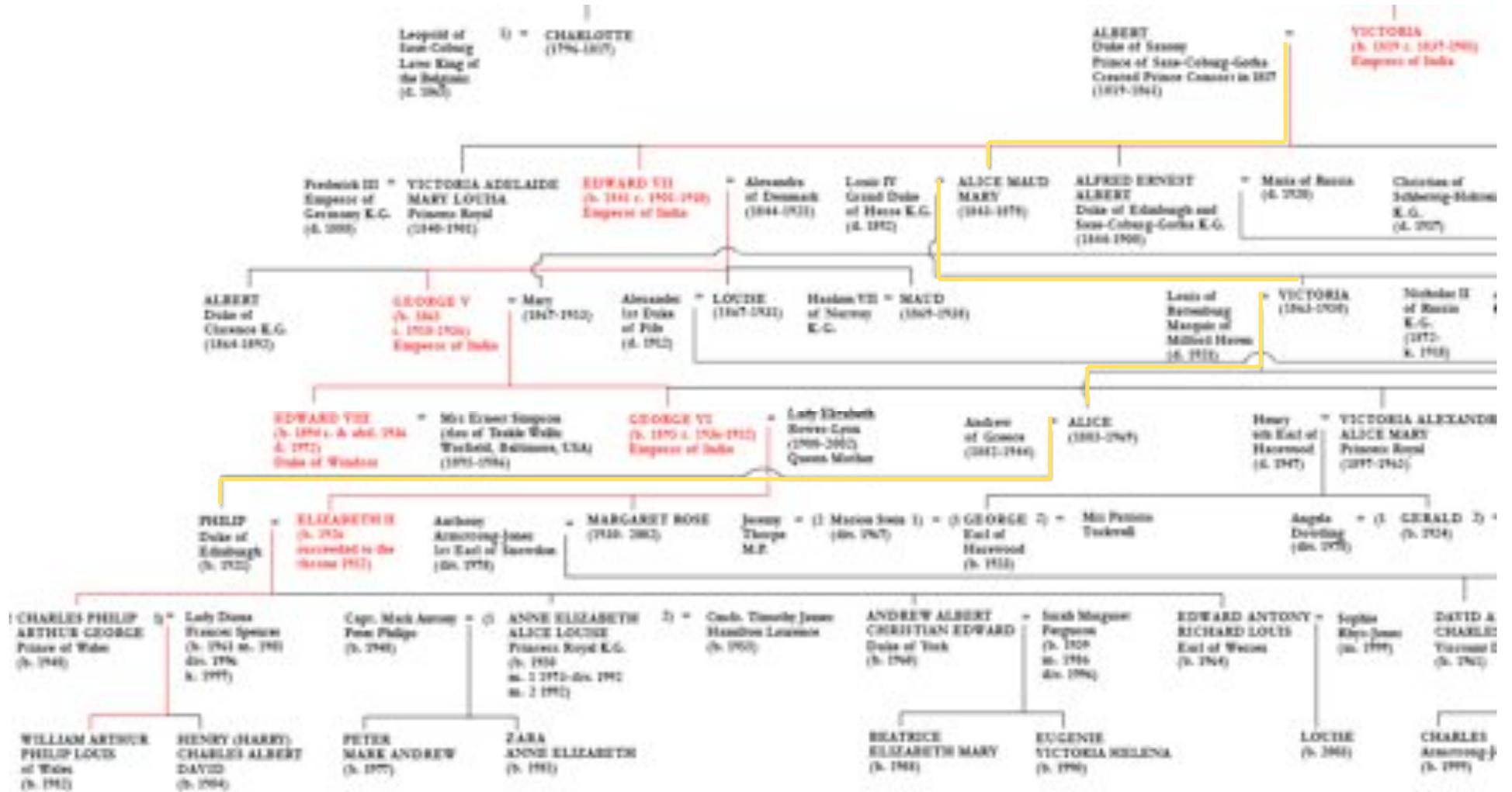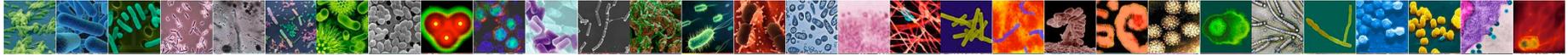
Visualizing Relatedness
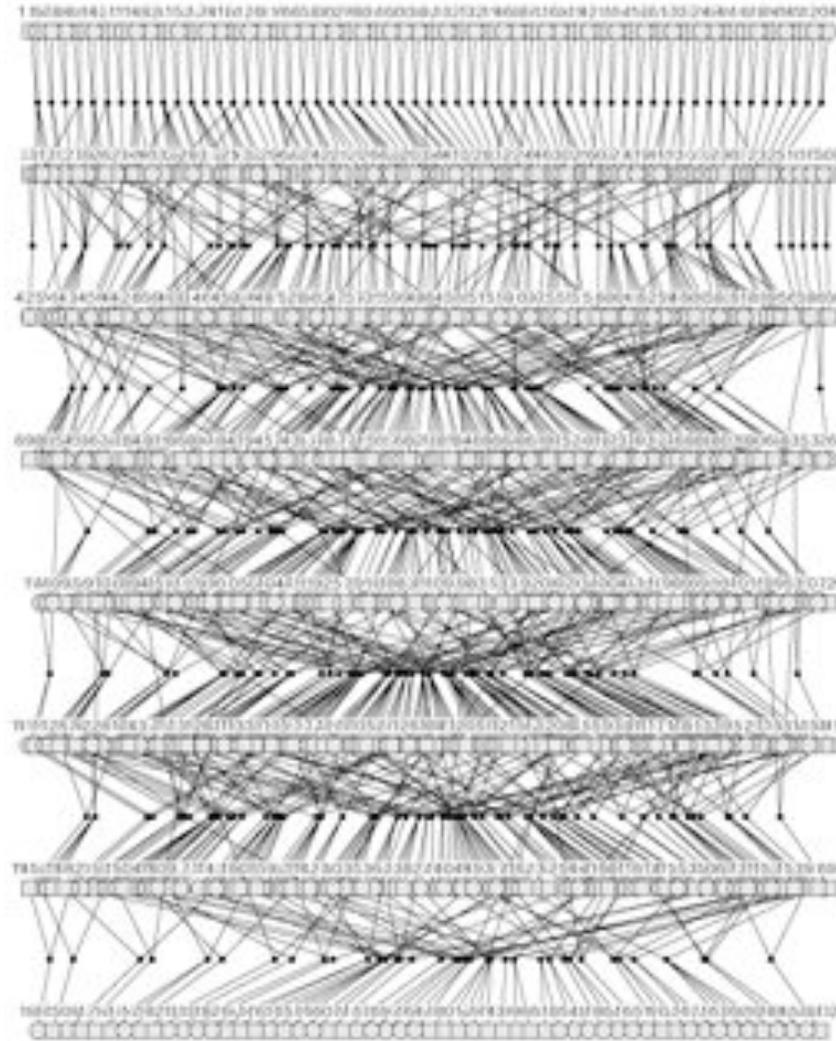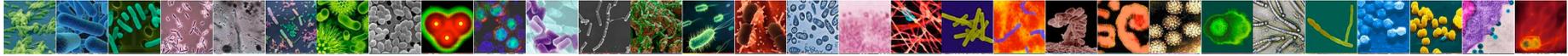
Evolutionary Genetics: What Use to Anyone?

Host susceptibility

Resistance genes

Vaccination

Diagnosis

Adaptation

Emergence

Genetics

Relatedness

Evolution

Epidemiology

Transmission

Control + prevention

R_0

Source attribution

Contact tracing

Population dynamics

Population structure

daniel.wilson@ndm.ox.ac.uk

Evolutionary Genetics:
What Use to Anyone?

- Visualizing Relatedness
- Understanding Population Structure
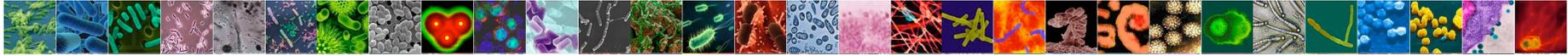- Attributing the Source of Infection
- Reconstructing Demographic History
- Unravelling the Genetic Basis of Disease
- Learning about Evolution

- Ad hoc vs. Model-Based Approaches
- Technical Challenges

Understanding Population Structure

Genes mirror geography within Europe

See associated Correspondence: Detours, Nature **455**, 861 (October 2008)

John Novembre[1,2], Toby Johnson[4,5,6], Katarzyna Bryc[7], Zoltán Kutalik[4,6],
Adam R. Boyko[7], Adam Auton[7], Amit Indap[7], Karen S. King[8], Sven
Bergmann[4,6], Matthew R. Nelson[8], Matthew Stephens[2,3] & Carlos D.
Bustamante[7]
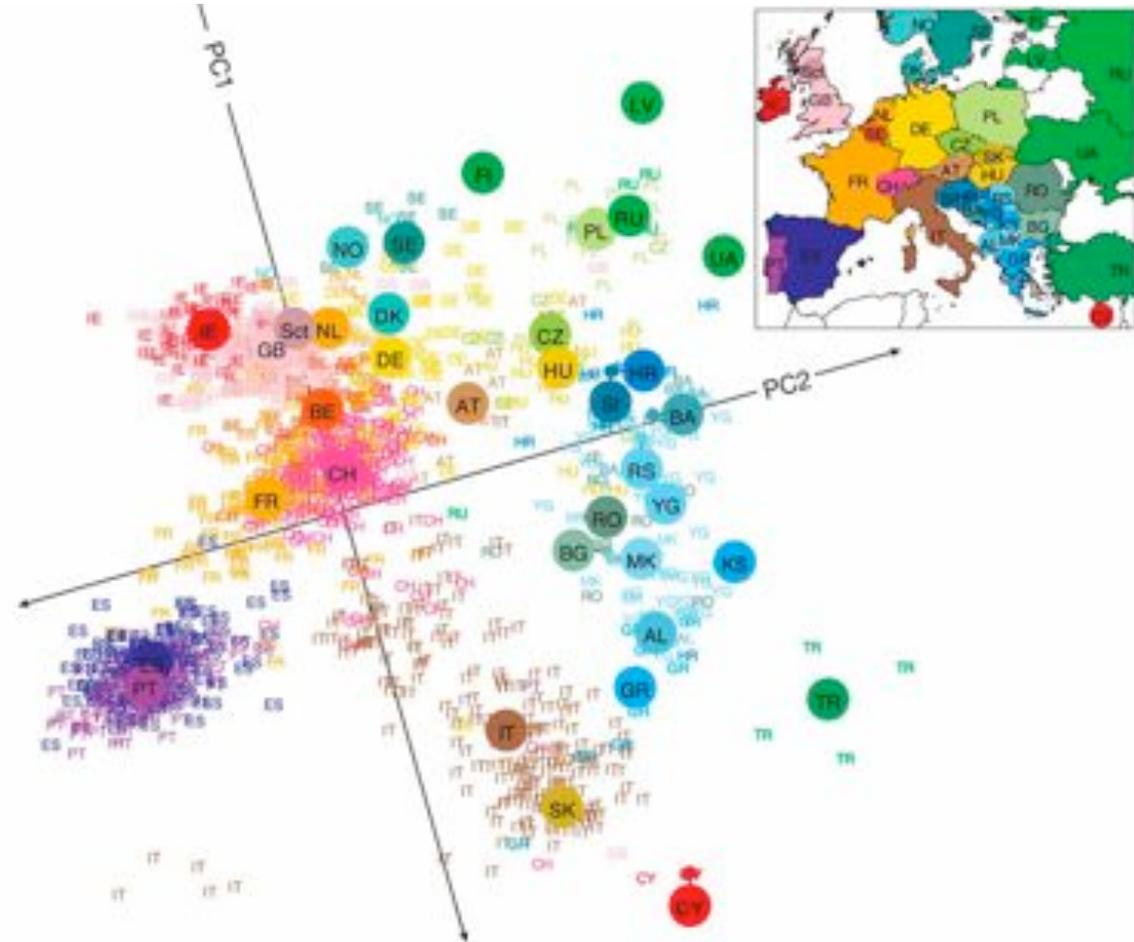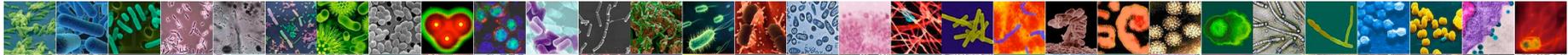
Understanding Population Structure

Traces of Human Migrations in *Helicobacter pylori* Populations

Daniel Falush,[1] Thierry Wirth,[1] Bodo Linz,[1]
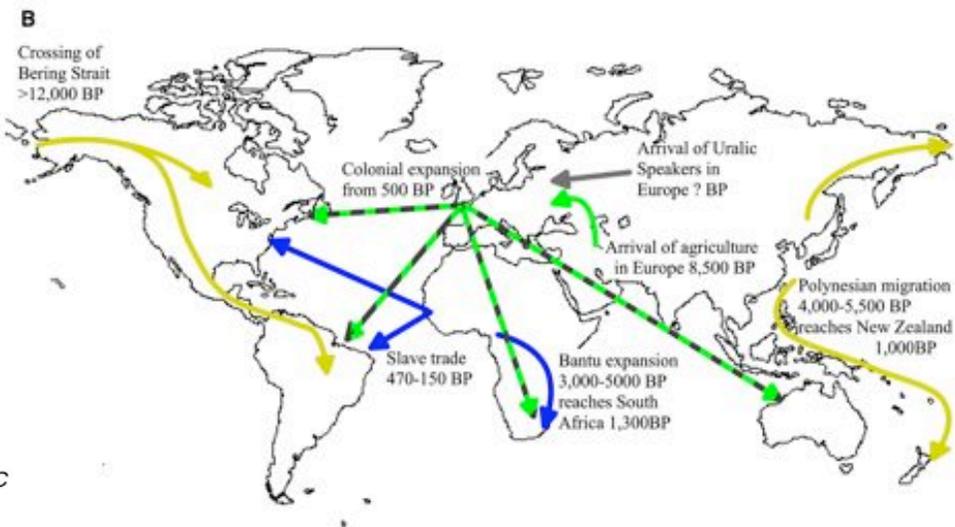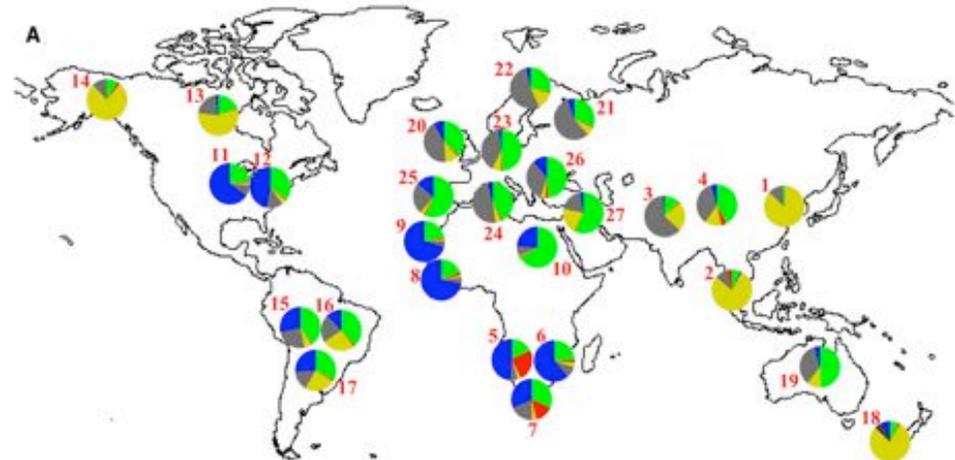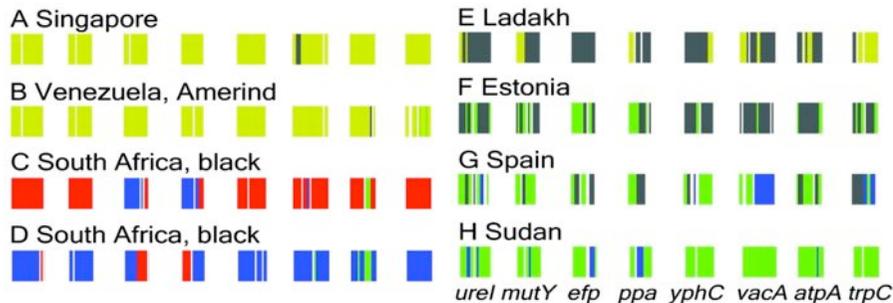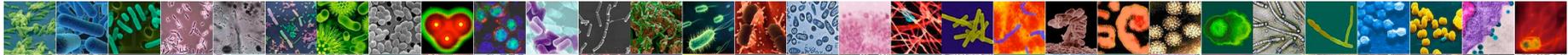Jonathan K. Pritchard,[2] Matthew Stephens,[3] Mark Kidd,[4]
Martin J. Blaser,[5] David Y. Graham,[6] Sylvie Vacher,[7]
Guillermo I. Perez-Perez,[5] Yoshio Yamaoka,[6] Francis Mégraud,[7]
Kristina Otto,[8] Ulrike Reichard,[7] Elena Katzowitsch,[8]
Xiaoyan Wang,[1] Mark Achtman,[1*] Sebastian Suerbaum[8]

Helicobacter pylori, a chronic gastric pathogen of human beings, can be divided into seven populations and subpopulations with distinct geographical distributions. These modern populations derive their gene pools from ancestral populations that arose in Africa, Central Asia, and East Asia. Subsequent spread can be attributed to human migratory fluxes such as the prehistoric colonization of Polynesia and the Americas, the neolithic introduction of farming to Europe, the Bantu expansion within Africa, and the slave trade.
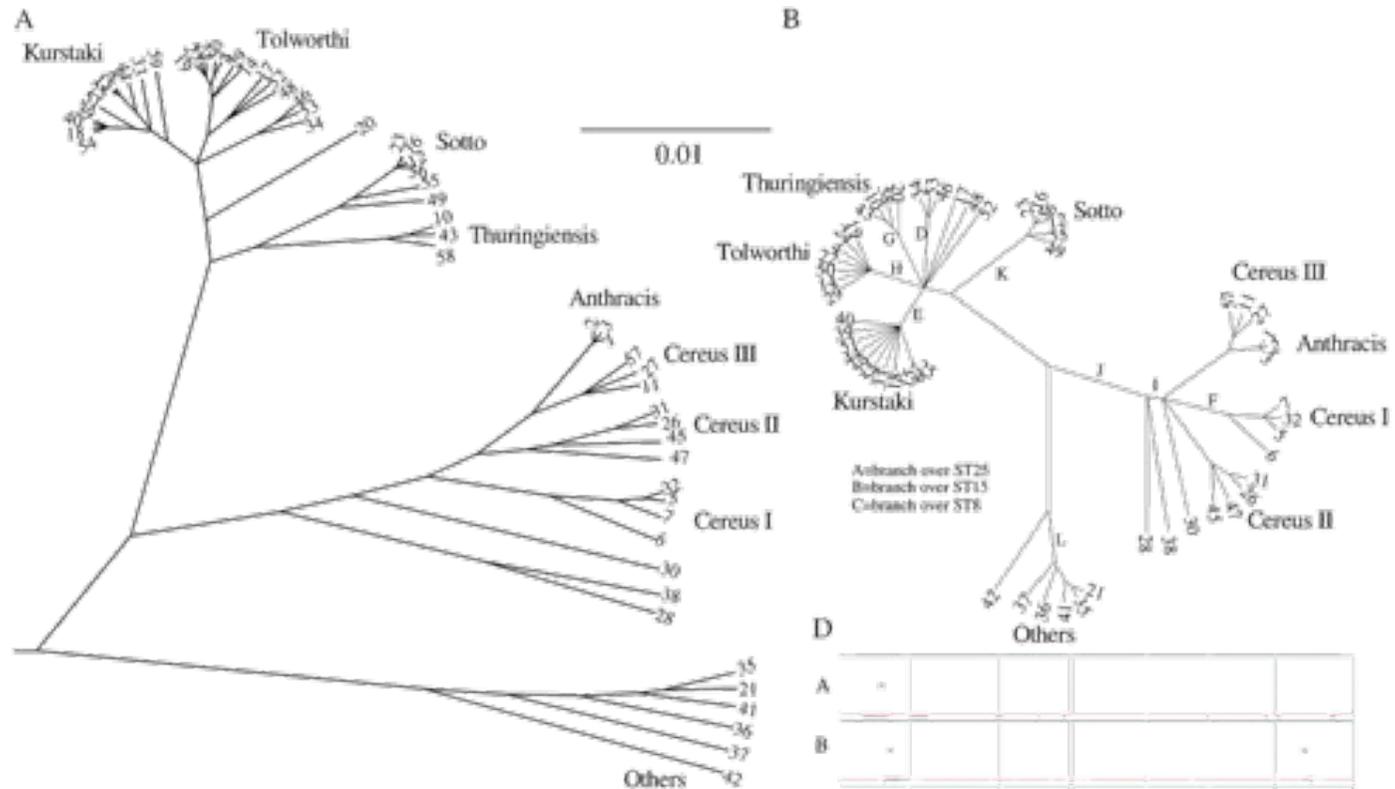
A Singapore

B Venezuela, Amerind

C South Africa, black

D South Africa, black

E Ladakh

F Estonia

G Spain

H Sudan

urel mutY efp ppa yphC vacA atpA trpC

Understanding Population Structure

A

Kurstaki Tolworthi

0.01

Sotto

Thuringiensis

Anthracis

Cereus III

Cereus II

Cereus I

Others

B

Thuringiensis

Tolworthi

Sotto

Cereus III

Anthracis

Kurstaki

Cereus I

A-branch over ST25
B-branch over ST15
C-branch over ST8

Cereus II

Others

D

Inference of Bacterial Microevolution Using Multilocus Sequence Data

Xavier Didelot and Daniel Falush[1]

Department of Statistics, University of Oxford, Oxford OX1 3SY, United Kingdom

Evolutionary Genetics:
What Use to Anyone?

Host susceptibility

Resistance genes

Vaccination

Diagnosis

Adaptation

Emergence

Genetics

Relatedness

Evolution

Epidemiology

Transmission

Control + prevention

R_0

Population dynamics

Source attribution

Population structure

Contact tracing
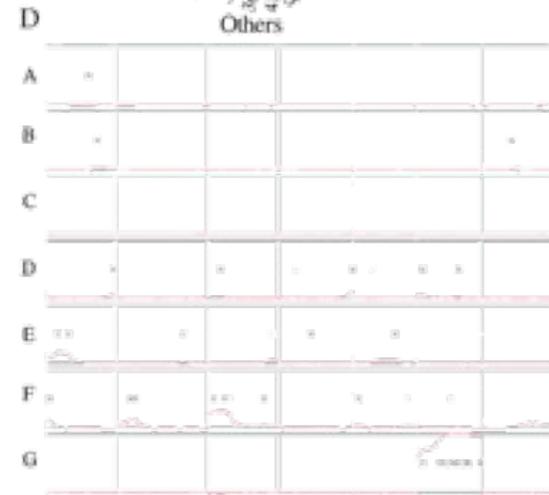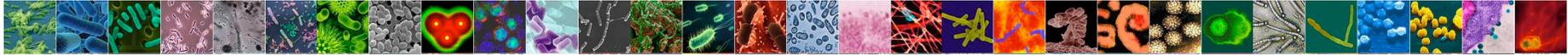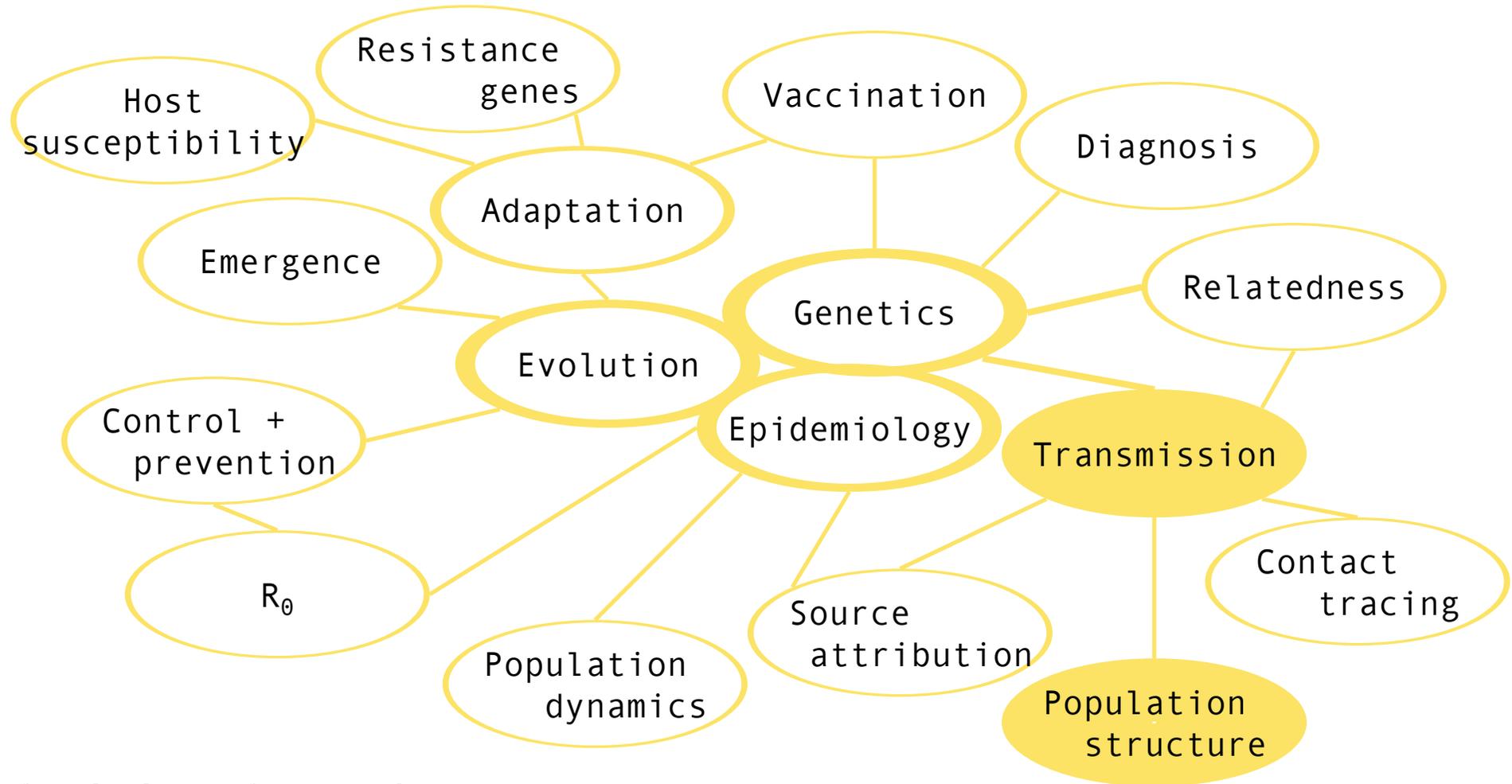
daniel.wilson@ndm.ox.ac.uk

Evolutionary Genetics:
What Use to Anyone?

- Visualizing Relatedness
- Understanding Population Structure
- Attributing the Source of Infection
- Reconstructing Demographic History
- Unravelling the Genetic Basis of Disease
- Learning about Evolution

- Ad hoc vs. Model-Based Approaches
- Technical Challenges

Attributing the Source of Infection

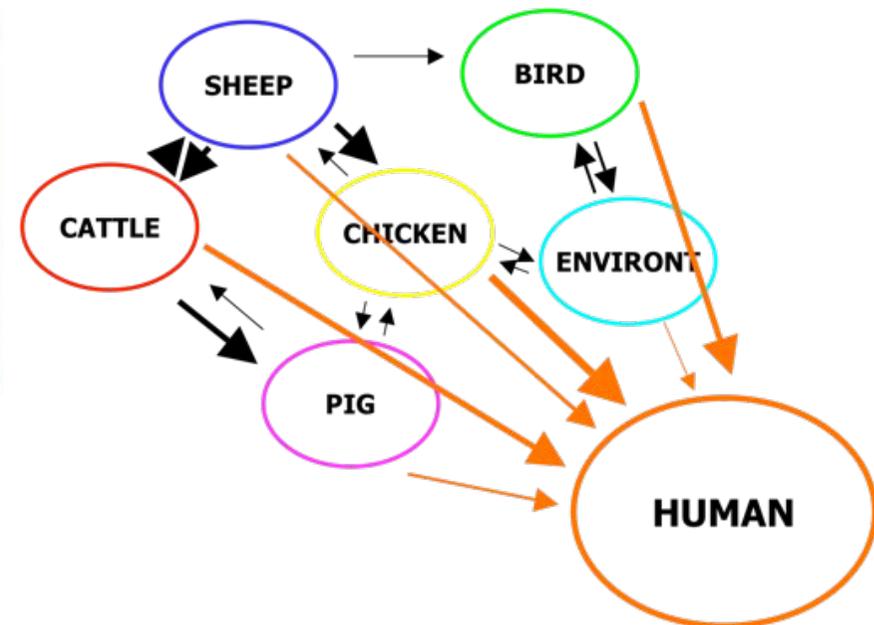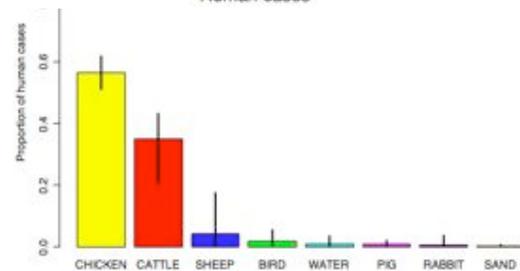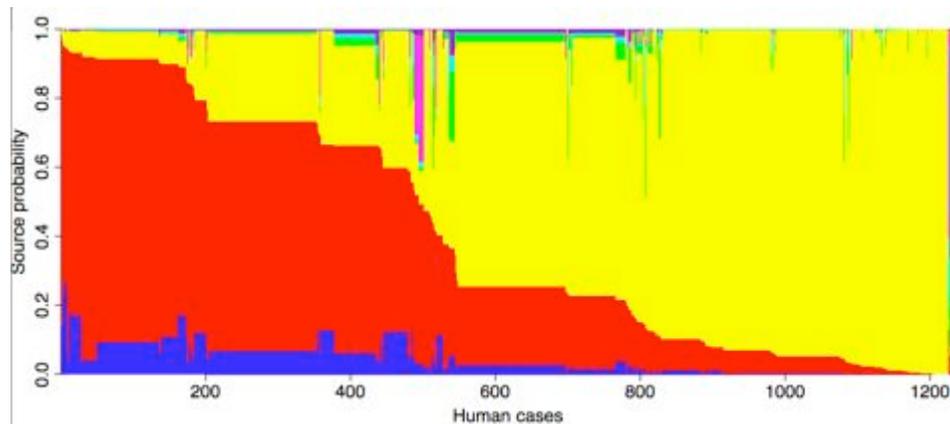## Comparative phylogenomics of the food-borne pathogen *Campylobacter jejuni* reveals genetic markers predictive of infection source

Olivia L. Champion[a], Michael W. Gaunt[a], Ozan Gundogdu[a], Abdi Elmi[a], Adam A. Witney[b], Jason Hinds[b], Nick Dorrell[a], and Brendan W. Wren[a]

[a]Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom; and [b]Bacterial Microarray Group, Medical Microbiology, Department of Cellular and Molecular Medicine, St. George's University of London, Cranmer Terrace, London SW17 0RE, United Kingdom

Campylobacter jejuni is the predominant cause of bacterial gastroenteritis worldwide, but traditional typing methods are unable to discriminate strains from different sources that cause disease in humans. We report the use of genomotyping (whole-genome comparisons of microbes using DNA microarrays) combined with Bayesian-based algorithms to model the phylogeny of this major food-borne pathogen. In this study 111 *C. jejuni* strains were examined by genomotyping isolates from humans with a spectrum of *C. jejuni*-associated disease (79 strains), chickens (17 strains), bovines (11 strains), ovines (5 strains), and the environment (8 strains). From these data, the Bayesian phylogeny of the isolates examines the genetic relatedness of bacterial populations. Examples where whole-genome microarray analysis has been used include the investigation of *Vibrio cholerae* strains associated with the current seventh cholera pandemic (7). A *V. cholerae* microarray based on the El Tor O1 strain N16961 was used to analyze a collection of nine strains of diverse global origin isolated between 1910 and 1992, and it was possible to differentiate classical biotype strains from El Tor biotype strains, and two putative chromosomal islands (VSP-I and VSP-II) with a deviant G+C content were identified in El Tor biotypes (7). Similarly, the pathological outcome of *Yersinia* species infections has been investigated with microarrays (8). DNA

- Study 111 strains of human, chicken, bovine, ovine and environmental strains.

- High posterior probability (*Pr*=1.0) for two clades: the livestock clade contains 89% of livestock isolates and the non-livestock clade contains environmental isolates.

- The majority (56%) of human isolates fall within the non-livestock clade.

- This suggests most human infections may originate from non-livestock (and possibly non-agricultural) sources, which may have implications for control strategies.

Attributing the Source of Infection

Evolutionary Genetics: What Use to Anyone?

Host susceptibility

Resistance genes

Vaccination

Diagnosis

Adaptation

Emergence

Genetics

Relatedness

Evolution

Control + prevention

Epidemiology

Transmission

R_0

Source attribution

Contact tracing

Population dynamics

Population structure

daniel.wilson@ndm.ox.ac.uk

Evolutionary Genetics:
What Use to Anyone?

- Visualizing Relatedness
- Understanding Population Structure
- Attributing the Source of Infection
- Reconstructing Demographic History
- Unravelling the Genetic Basis of Disease
- Learning about Evolution

- Ad hoc vs. Model-Based Approaches
- Technical Challenges

Reconstructing Demographic History
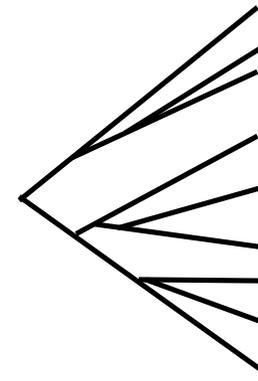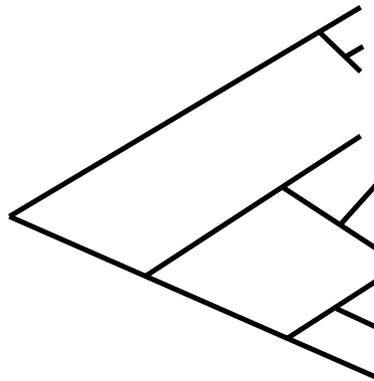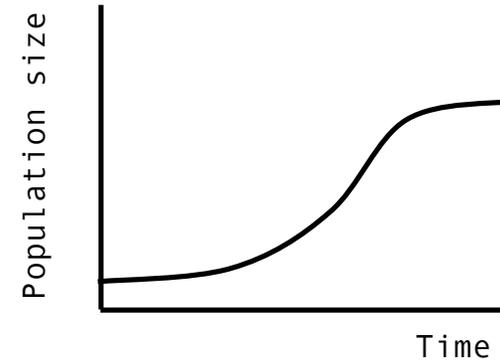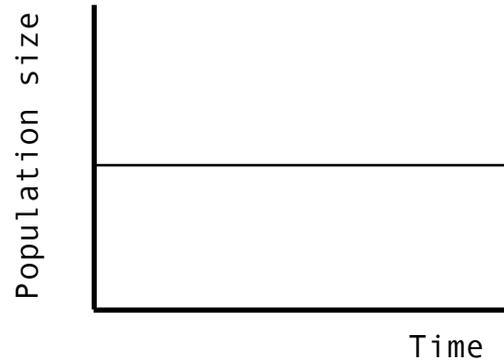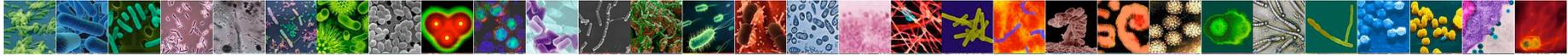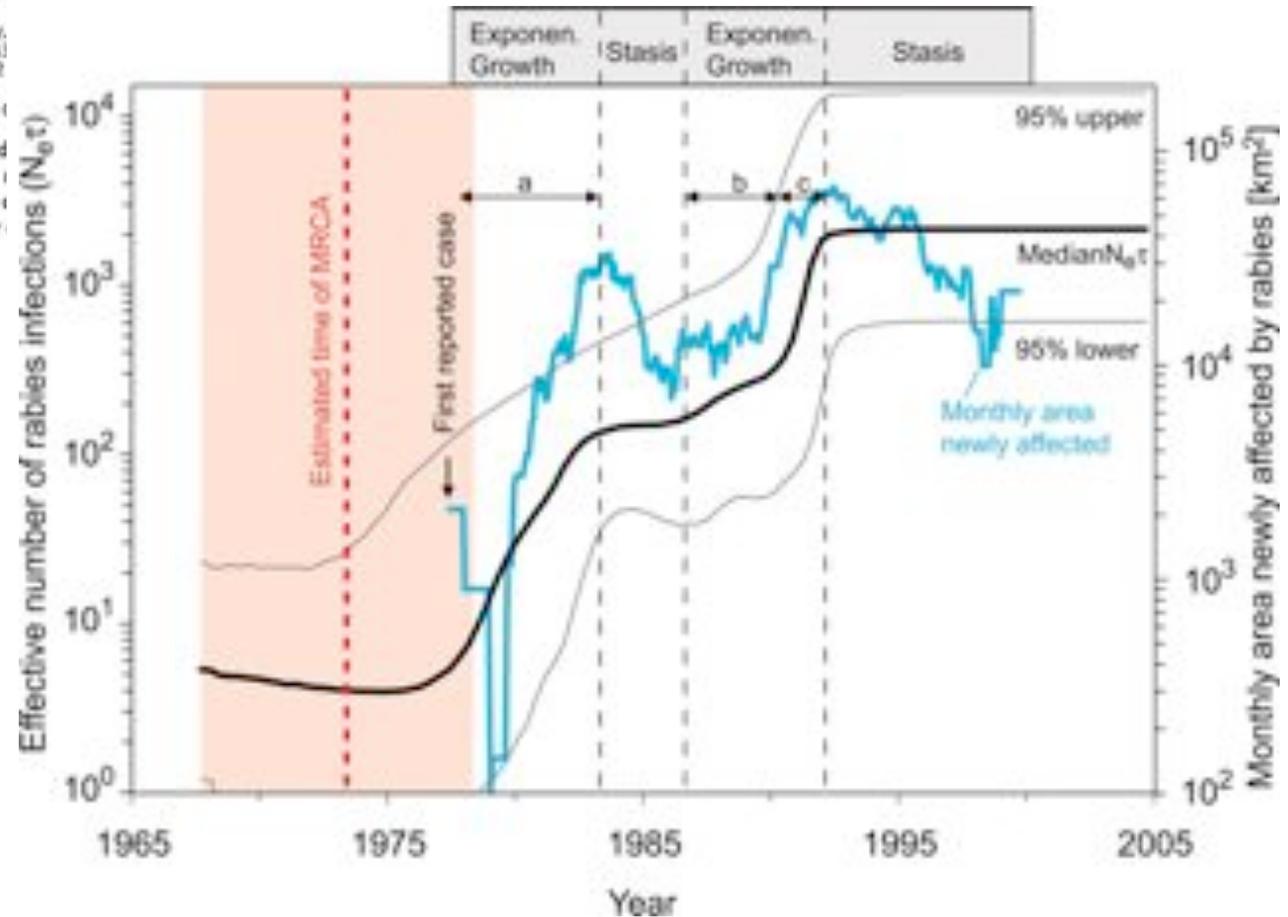
Reconstructing Demographic History

A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus
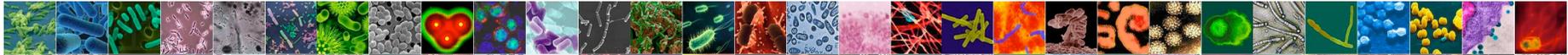
Reconstructing Demographic History

BMC Evolutionary Biology

BioMed Central

Open Access

Population genetic estimation of the loss of genetic diversity during horizontal transmission of HIV-1

Charles TT Edwards*[1], Edward C Holmes[2], Daniel J Wilson[3], Raphael P Viscidi[4], Elaine J Abrams[5], Rodney E Phillips[1] and Alexei J Drummond[6,7]

Abstract

Background: Genetic diversity of the human immunodeficiency virus type 1 (HIV-1) population within an individual is lost during transmission to a new host. The demography of transmission is an important determinant of evolutionary dynamics, particularly the relative impact of natural selection and genetic drift immediately following HIV-1 infection. Despite this, the magnitude of this population bottleneck is unclear.

Results: We use coalescent methods to quantify the bottleneck in a single case of homosexual transmission and find that over 99% of the env and gag diversity present in the donor is lost. This was consistent with the diversity present at seroconversion in nine other horizontally infected individuals. Furthermore, we estimated viral diversity at birth in 27 infants infected through vertical transmission and found there to be no difference between the two modes of transmission.
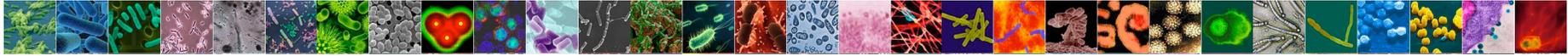
Conclusion: Assuming the bottleneck at transmission is selectively neutral, such a severe reduction in genetic diversity has important implications for adaptation in HIV-1, since beneficial mutations have a reduced chance of transmission.
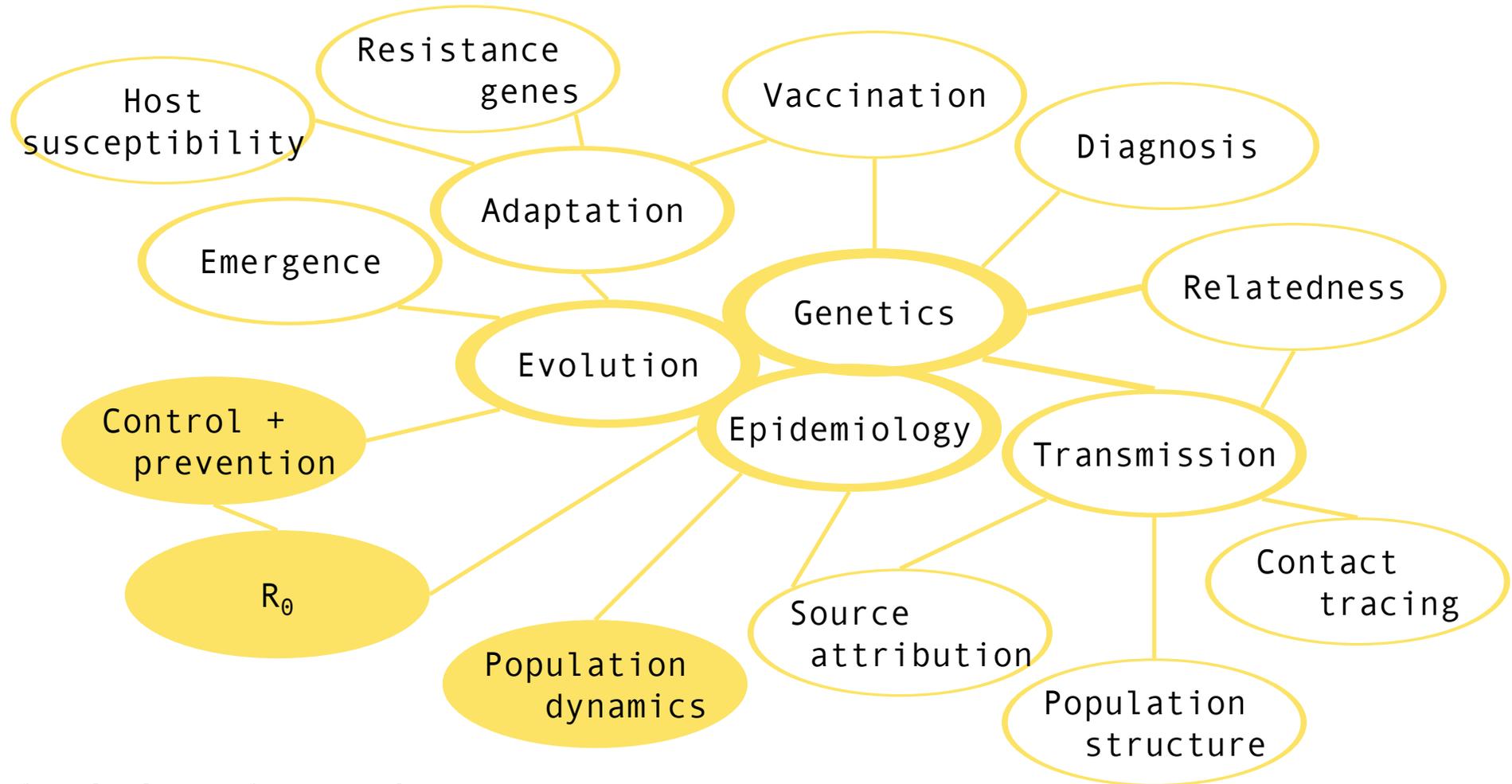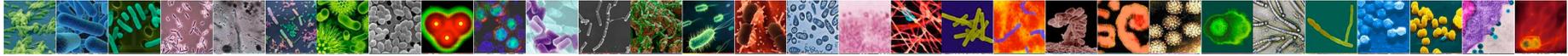
(a) *env V1-V4*
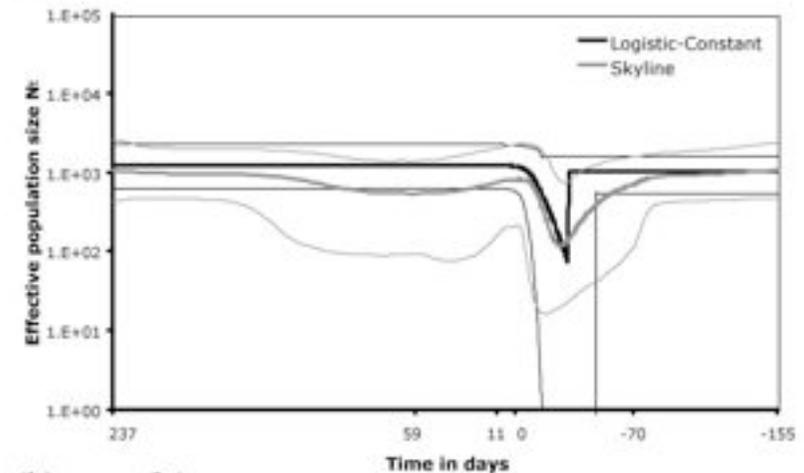
(b) *gag p24*
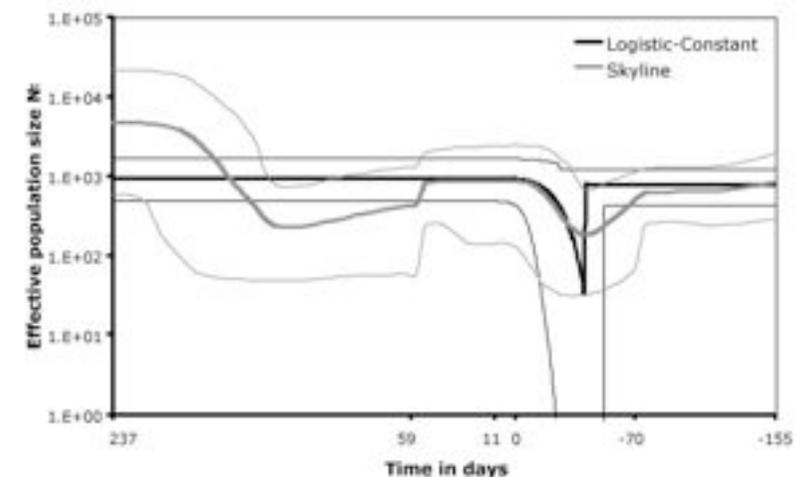
Evolutionary Genetics: What Use to Anyone?
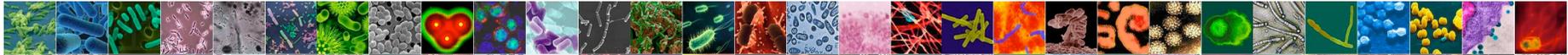
daniel.wilson@ndm.ox.ac.uk

Evolutionary Genetics:
What Use to Anyone?

- Visualizing Relatedness

- Understanding Population Structure

- Attributing the Source of Infection

- Reconstructing Demographic History

- Unravelling the Genetic Basis of Disease

- Learning about Evolution

- Ad hoc vs. Model-Based Approaches

- Technical Challenges
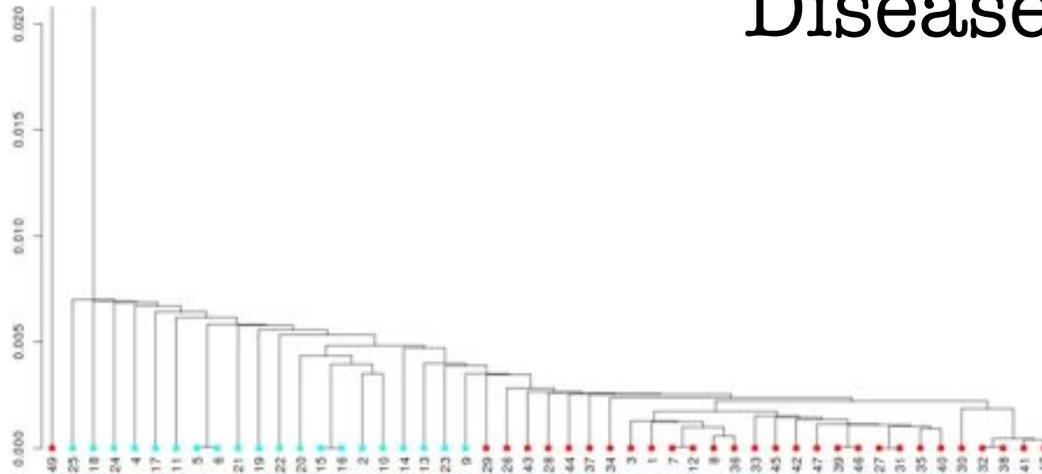
Unravelling the Genetic Basis of Disease

THE PHYLOGENETIC REGRESSION

By A. GRAFEN†

Animal Behaviour Research Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K.
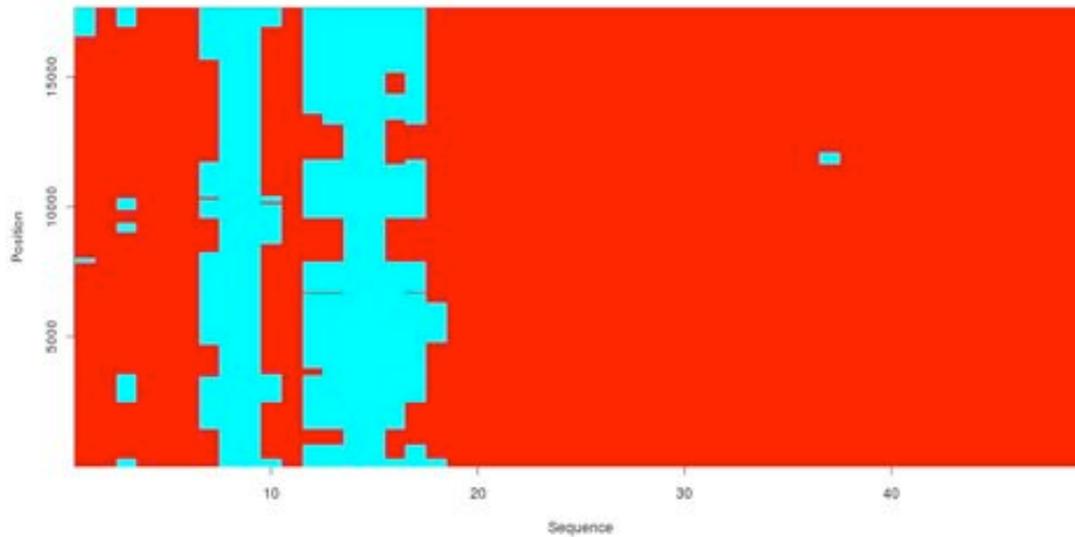
nature genetics

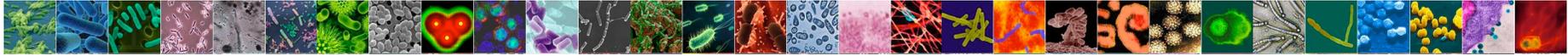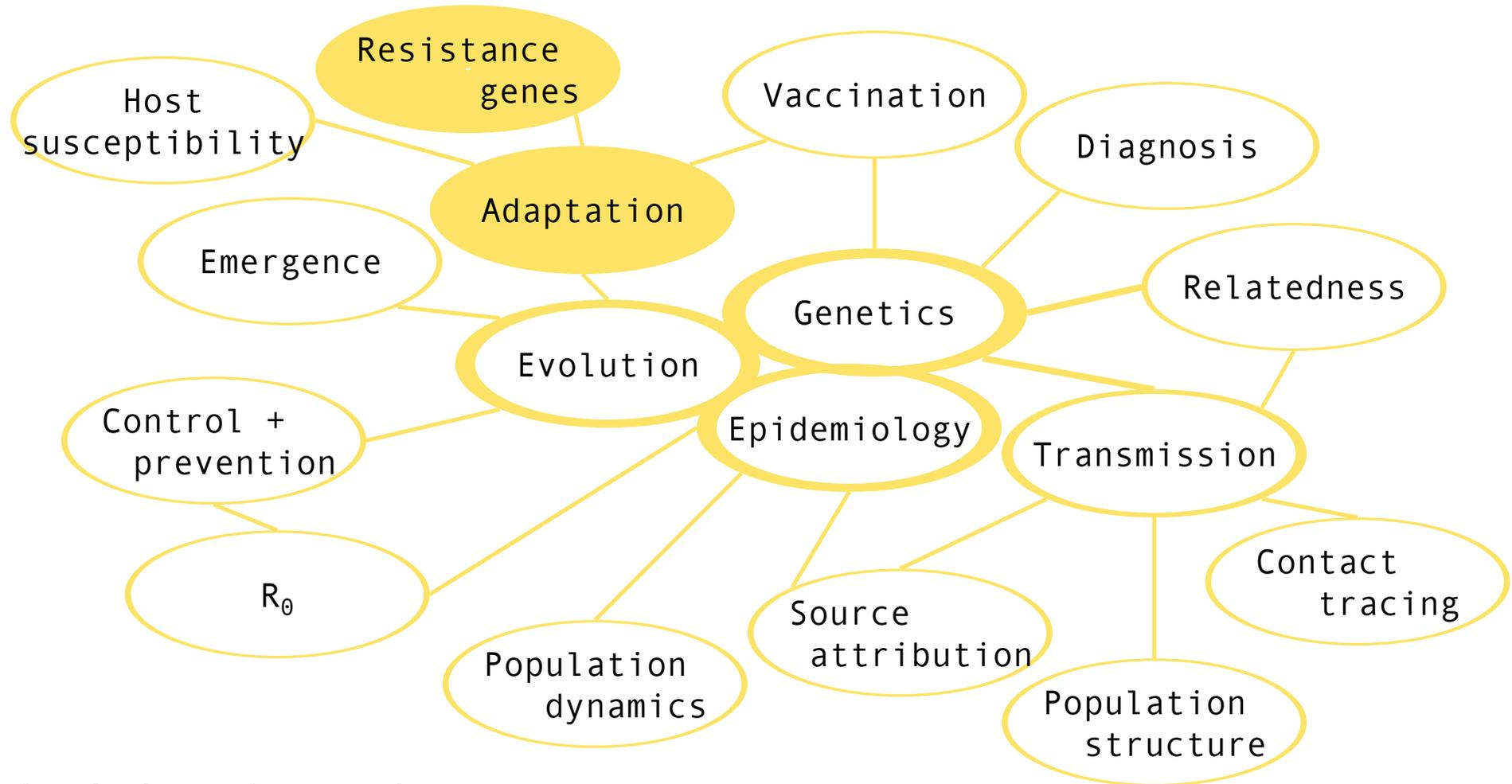## Principal components analysis corrects for stratification in genome-wide association studies

Alkes L Price[1,2], Nick J Patterson[2], Robert M Plenge[2,3], Michael E Weinblatt[3], Nancy A Shadick[3] & David Reich[1,2]

Population stratification—allele frequency differences between cases and controls due to systematic ancestry differences—can cause spurious associations in disease studies. We describe a method that enables explicit detection and correction of population
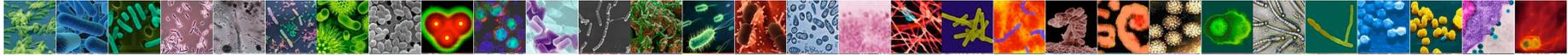
Evolutionary Genetics:
What Use to Anyone?

Resistance genes

Host susceptibility

Vaccination

Diagnosis

Adaptation

Emergence

Genetics

Relatedness

Evolution

Epidemiology

Transmission

Control + prevention

R_0

Population dynamics

Source attribution

Contact tracing

Population structure
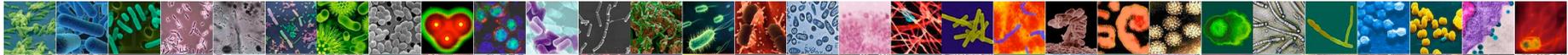
daniel.wilson@ndm.ox.ac.uk

Evolutionary Genetics: What Use to Anyone?

- Visualizing Relatedness
- Understanding Population Structure
- Attributing the Source of Infection
- Reconstructing Demographic History
- Unravelling the Genetic Basis of Disease
- Learning about Evolution

- Ad hoc vs. Model-Based Approaches
- Technical Challenges
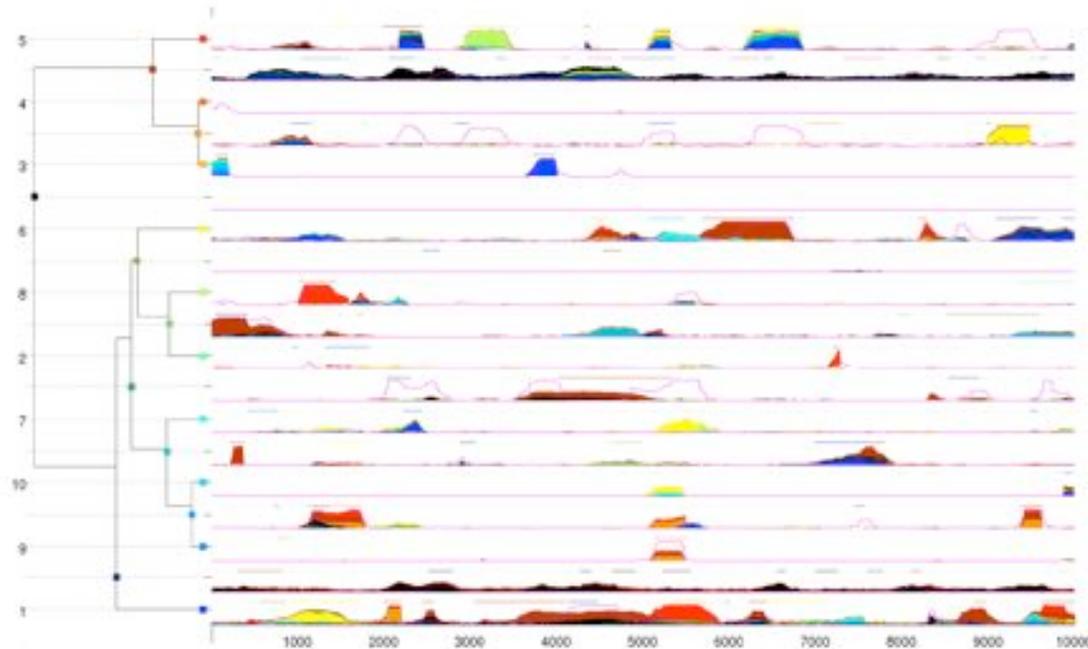
Learning About Evolution

How clonal are bacteria?

(recombination/linkage disequilibrium/population structure/genetic transformation/parasite evolution)

JOHN MAYNARD SMITH, NOEL H. SMITH, MARIA O'ROURKE, AND BRIAN G. SPRATT*

School of Biological Sciences, University of Sussex, Falmer, Brighton BN1 9QG, United Kingdom

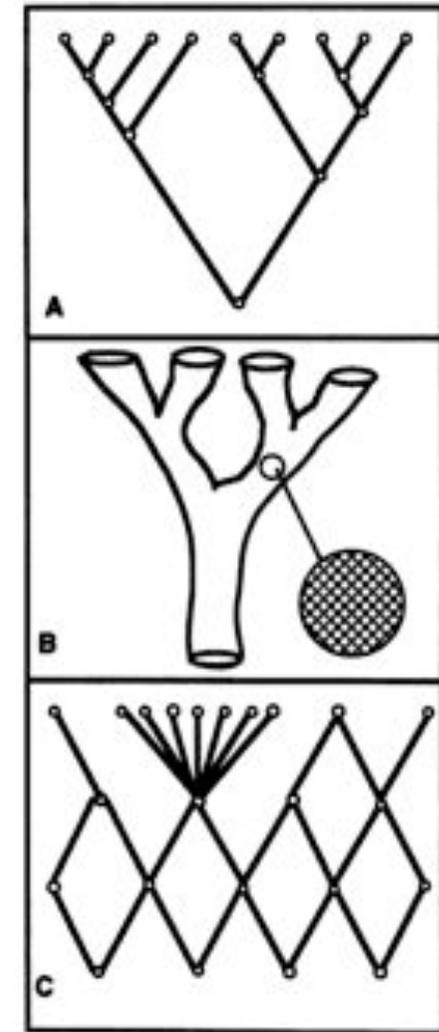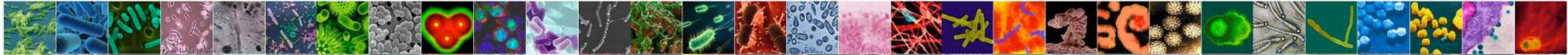**ClonalOrigin** Xavier Didelot et al, in preparation

FIG. 1. Representations of population structures.

Learning About Evolution

Rapid Evolution and the Importance of Recombination to the Gastroenteric Pathogen *Campylobacter jejuni*

Daniel J. Wilson,*[1] Edith Gabriel,†[2] Andrew J.H. Leatherbarrow,‡ John Cheesbrough,§ Steven Gee,§ Eric Bolton,|| Andrew Fox,§|| C. Anthony Hart,¶[3] Peter J. Diggle,† and Paul Fearnhead*

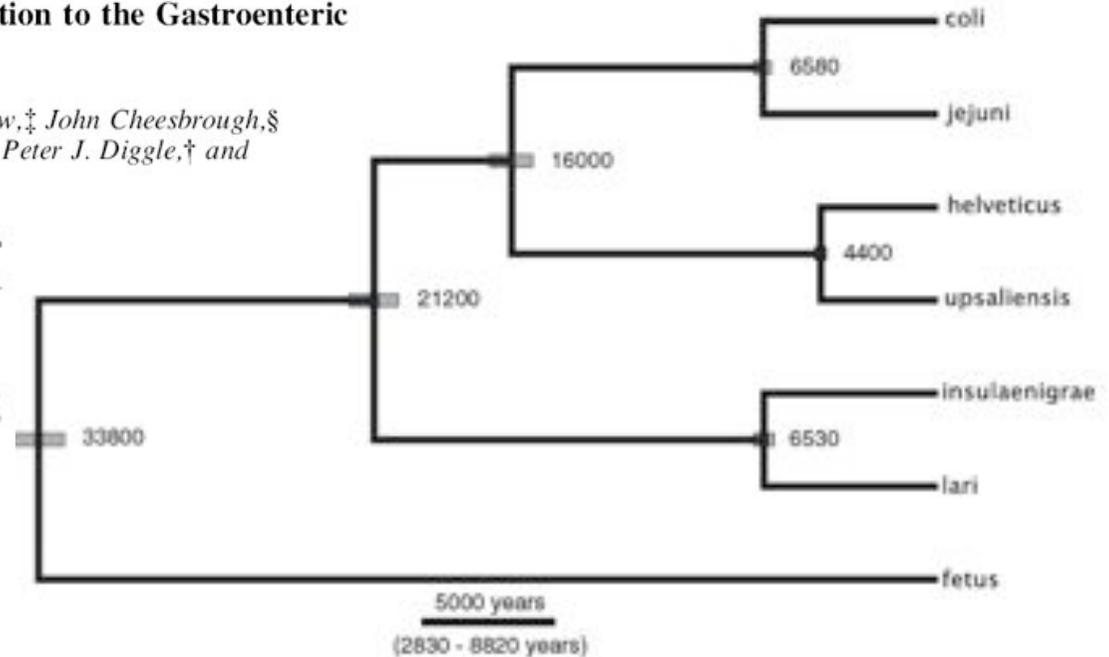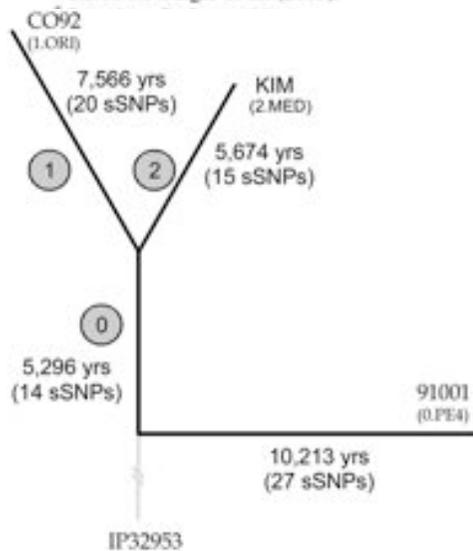**Alternative Scale Bars for *Campylobacter* Phylogeny**

Method	Scale (Years)	95% CI
Intraspecific[a]	5,000	2,830–8,820
Empirical[b]	42,200	2,690–661,000
Ochman–Wilson[c]	7,600,000	Not quantified
Coalescent[d]	$23.6 N_e g$	$14.1\ N_e g$–$39.6\ N_e g$

[a] Based on μ_S from table 3.
[b] Based on μ_S calculated from μ_0 in table 4.
[c] Based on Dingle et al. (2005).

Proceedings of the National Academy of Sciences of the United States of America

Microevolution and history of the plague bacillus, *Yersinia pestis*

Mark Achtman*,‡, Giovanna Morelli*, Peixuan Zhu*,‡, Thierry Wirth*,§, Ines Diehl*, Barica Kusecek*, Amy J. Vogler¶, David M. Wagner¶, Christopher J. Allender¶, W. Ryan Easterday¶, Viviane Chenal-Francisque∅, Patricia Worsham**, Nicholas R. Thomson††, Julian Parkhill††, Luther E. Lindler‡‡,§§, Elisabeth Carniel∅, and Paul Keim¶,¶¶

Evolutionary Genetics: What Use to Anyone?

Host susceptibility

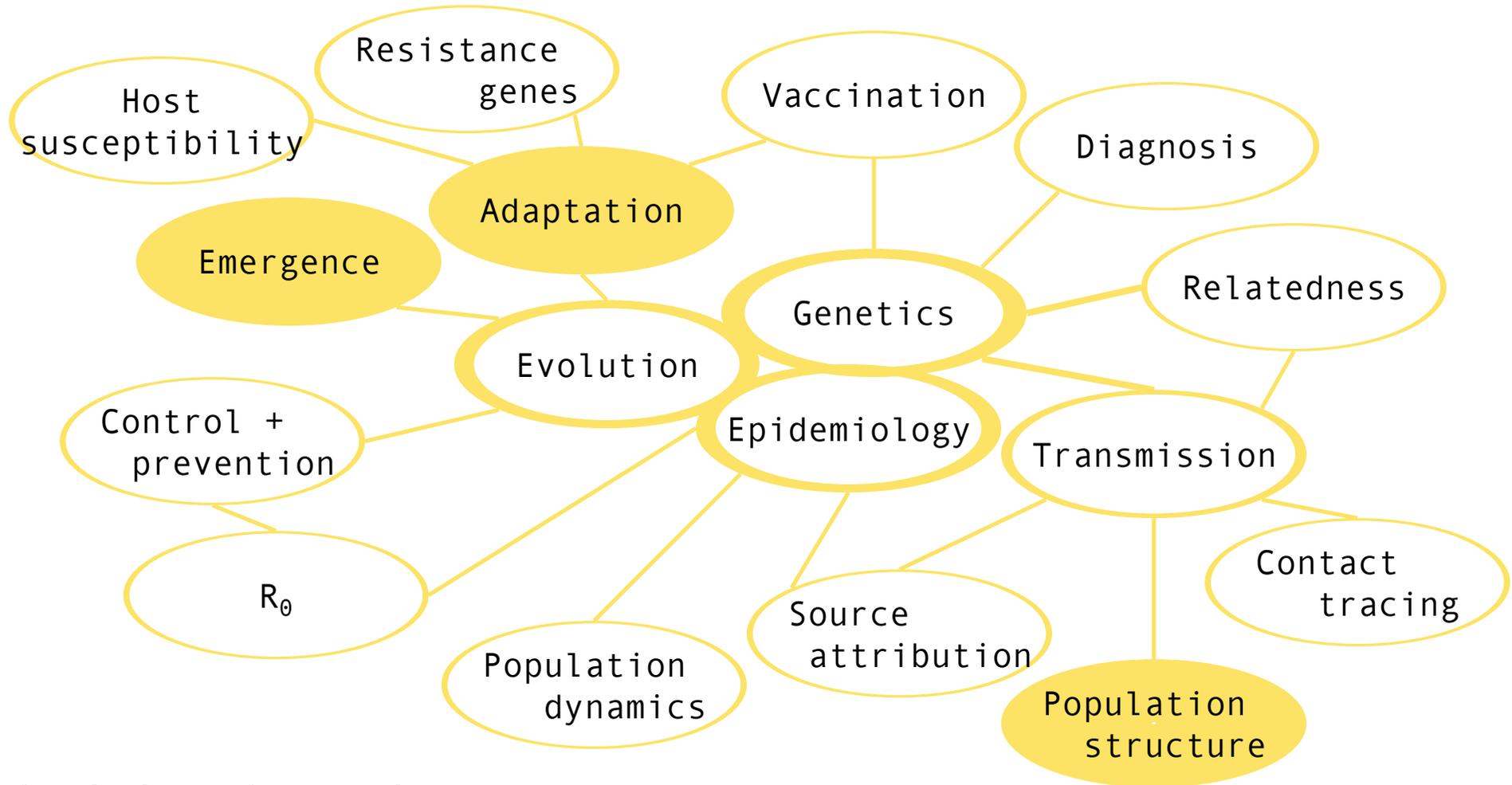Resistance genes

Vaccination

Diagnosis

Adaptation

Emergence

Genetics

Relatedness

Evolution

Control + prevention

Epidemiology

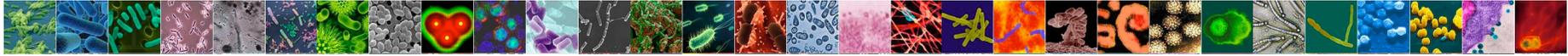Transmission

R_0

Source attribution

Contact tracing

Population dynamics

Population structure
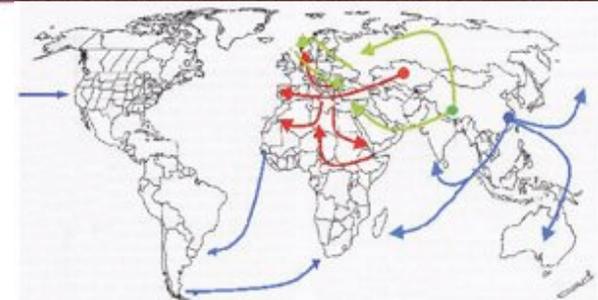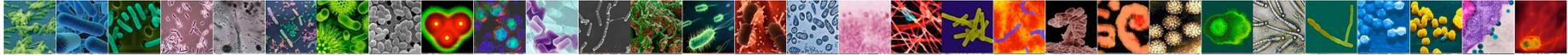
daniel.wilson@ndm.ox.ac.uk

Challenges and Objectives

- Prioritize the most clinically important and scientifically relevant questions.

- Consider appropriate evolutionary models for addressing those questions.

- Plan strategies for coping with
  - Genome instability
  - Recombination

- Develop computationally feasible methods for statistical inference.

- Disseminate findings widely.

- Identify avenues for further research.

Evolutionary Genetics: What Use to Anyone?

Host susceptibility

Resistance genes

Vaccination

Diagnosis

Adaptation

Emergence

Genetics

Relatedness

Evolution

Epidemiology

Transmission

Control + prevention

R_0

Population dynamics

Source attribution

Contact tracing

Population structure

daniel.wilson@ndm.ox.ac.uk