Evolutionary pathways and genotype networks of foodborne pathogens

Prof. Mikhail Prokopenko Centre for Complex Systems School of Computer Science, Faculty of Engineering Sydney Institute for Infectious Diseases

Digital Epidemiology and Pathogen Genomics Workshop, Westmead 22 September 2023





Network-based modelling

SCIENTIFIC **Reports**

OPEN

Received: 7 January 2019 Accepted: 3 April 2019 Published online: 16 April 2019

Network properties of salmonella epidemics

Oliver M. Cliff¹, Vitali Sintchenko^{2,3}, Tania C. Sorrell^{2,3}, Kiranmayi Vadlamudi¹, Natalia McLean¹ & Mikhail Prokopenko^{1,3}

We examine non-typhoidal Salmonella (S. Typhimurium or STM) epidemics as complex systems, driven by evolution and interactions of diverse microbial strains, and focus on emergence of successful strains. Our findings challenge the established view that seasonal epidemics are associated with random sets of co-circulating STM genotypes. We use high-resolution molecular genotyping data comprising 17,107 STM isolates representing nine consecutive seasonal epidemics in Australia, genotyped by multiple-locus variable-number tandem-repeats analysis (MLVA). From these data, we infer weighted undirected networks based on distances between the MLVA profiles, depicting epidemics as networks of individual bacterial strains. The network analysis demonstrated dichotomy in STM populations which split into two distinct genetic branches, with markedly different prevalences. This distinction revealed the emergence of dominant STM strains defined by their local network topological properties, such as centrality, while correlating the development of new epidemics with global network features, such as small-world propensity.

Evolutionary pathways



PLOS COMPUTATIONAL BIOLOGY

G OPEN ACCESS

Citation: Cliff OM, McLean N, Sintchenko V, Fair KM, Sorrell TC, Kauffman S, et al. (2020) Inferring evolutionary pathways and directed genotype networks of foodborne pathogens. PLoS Comput Biol 16(10): e1008401. <u>https://doi.org/10.1371/</u> journal.pcbi.1008401

Editor: James Lloyd-Smith, University of California, Los Angeles, UNITED STATES

Received: April 30, 2020

Accepted: September 25, 2020

Published: October 30, 2020

RESEARCH ARTICLE

Inferring evolutionary pathways and directed genotype networks of foodborne pathogens

Oliver M. Cliff¹, Natalia McLean¹, Vitali Sintchenko^{2,3}, Kristopher M. Fair¹, Tania C. Sorrell^{2,3}, Stuart Kauffman⁴, Mikhail Prokopenko^{1,2*}

1 Centre for Complex Systems, Faculty of Engineering, University of Sydney, Sydney, New South Wales, Australia, 2 Marie Bashir Institute for Infectious Diseases and Biosecurity, University of Sydney, Westmead, New South Wales, Australia, 3 Centre for Infectious Diseases and Microbiology – Public Health, Institute of Clinical Pathology and Medical Research, NSW Health Pathology, Westmead Hospital, Sydney, New South Wales, Australia, 4 Institute for Systems Biology, Seattle, Washington, USA



Genome-wide networks

International Journal of Infectious Diseases 117 (2022) 65-73



Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

GENOME-WIDE NETWORKS REVEAL EMERGENCE OF EPIDEMIC STRAINS OF SALMONELLA ENTERITIDIS

Adam J. Svahn^{a,b,*}, Sheryl L. Chang^a, Rebecca J. Rockett^{b,c}, Oliver M. Cliff^{a,d}, Qinning Wang^{c,e}, Alicia Arnott^c, Marc Ramsperger^c, Tania C. Sorrell^{b,c}, Vitali Sintchenko^{b,c,e}, Mikhail Prokopenko^{a,b}

^a Complex Systems Research Group, Faculty of Engineering, The University of Sydney, Sydney, NSW, Australia

^b The University of Sydney Institute of Infectious Diseases, The University of Sydney, Westmead, NSW, Australia

^c Centre for Infectious Diseases and Microbiology–Public Health, Westmead Hospital, Westmead, NSW, Australia

^d School of Physics, Faculty of Science, The University of Sydney, Sydney, NSW, Australia

^e NSW Enteric Reference Laboratory, Institute of Clinical Pathology and Medical Research NSW Health Pathology, Westmead, NSW, Australia







Nine consecutive seasons of instances in NSW

1st January 2008 to 31st December 2016

17,107 isolates of STM

99.3% of all STM found from human cases in NSW over 3,287 days

Genotyped through MLVA

- Multiple-Locus Variable-number tandem repeat (VNTR) Analysis
- 1675 unique MLVAs identified



What is MLVA?

Tandem repeats

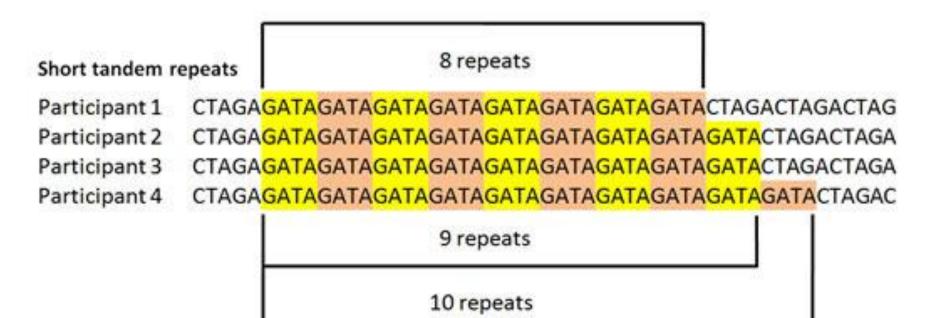
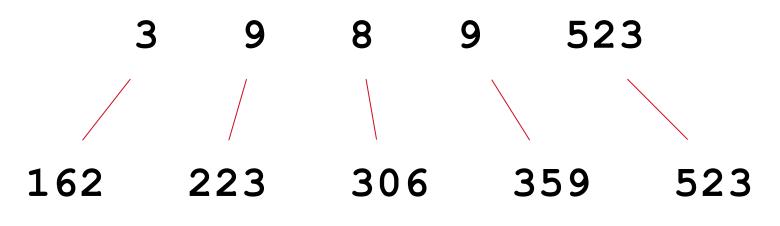


Image source: https://www.stewartsociety.org/images/bannockburn-tandem-repeats.jpg



What is MLVA?

Number of tandem repeats



Fragment sizes



Top 10 Salmonella Typhimurium MLVA profiles and number of isolates from 2011-2015, NSW*

	2011	2012	2013	2014	2015
1	3-9-7-13-523 (259)	3-17-9-12-523 (150)	3-17-9-12-523 (159)	3-17-9-11-523 (210)	3-12-11-14-523 (100)
2	3-10-8-9-523 (149)	3-9-8-13-523 (124)	3-9-8-13-523 (83)	3-12-11-14-523 (149)	3-17-9-11-523 (91)
3	3-9-8-13-523 (113)	3-9-7-13-523 (100)	3-9-7-13-523 (74)	3-12-12-9-523 (141)	3-12-12-9-523 (82)
4	3-9-7-14-523 (92)	3-16-9-12-523 (66)	3-10-14-12-496 (61)	3-10-7-12-523 (99)	3-12-13-9-523 (54)
5	3-12-9-10-550 (76)	3-10-8-9-523 (50)	3-10-7-14-523 (55)	3-9-7-12-523 (98)	3-24-13-10-523 (53)
6	3-9-7-15-523 (59)	3-9-8-12-523 (38)	3-13-11-9-523 (48)	3-9-8-12-523 (97)	3-10-8-12-523 (42)
7	3-14-11-12-523 (50)	3-9-8-14-523 (38)	3-9-7-14-523 (45)	3-16-9-11-523 (94)	3-9-7-12-523 (40)
8	3-10-14-12-496 (48)	3-9-9-13-523 (37)	3-23-23-11-523 (43)	3-17-10-11-523 (94)	3-17-8-11-523 (39)
9	3-12-15-13-523 46)	3-9-9-12-523 (34)	3-10-8-9-523 (39)	3-10-13-11-496 (52)	3-16-9-11-523 (37)
10	3-13-11-9-523 (30)	3-12-11-13-523 (29)	3-17-9-11-523 (39)	3-16-9-12-523 (50)	3-24-14-10-523 (33)

*Colour code indicates closely related MLVA patterns.

OzFoodNet. Enhancing Foodborne Disease Surveillance Across Australia. NSW Annual report, 2015.



MLVA isolates as a complex network

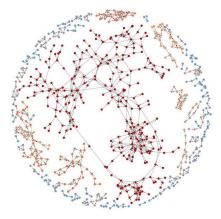
Construct a complete graph (network)

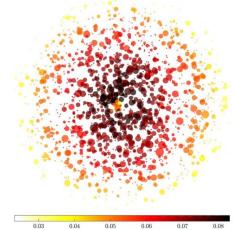
- 1675 nodes (unique MLVA profiles)
- edge weights are inverse of pairwise MLVA distance
- Compute centrality of MLVA profiles within network

Cluster nodes (MLVA profiles)

Trace changes in MLVA centrality over time

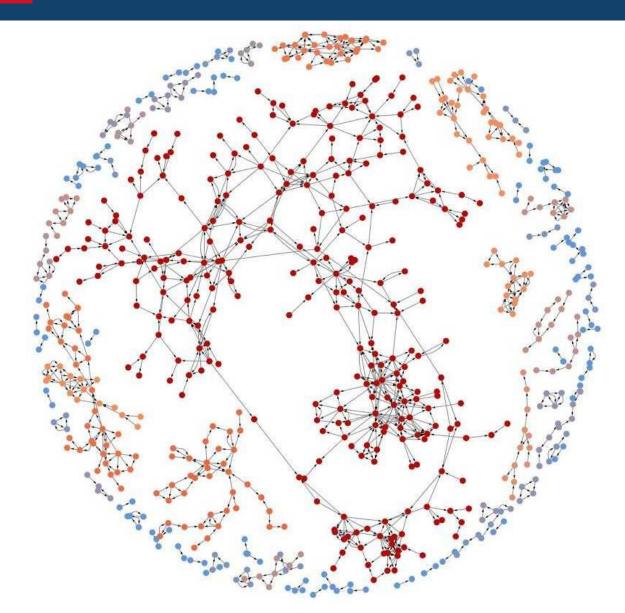






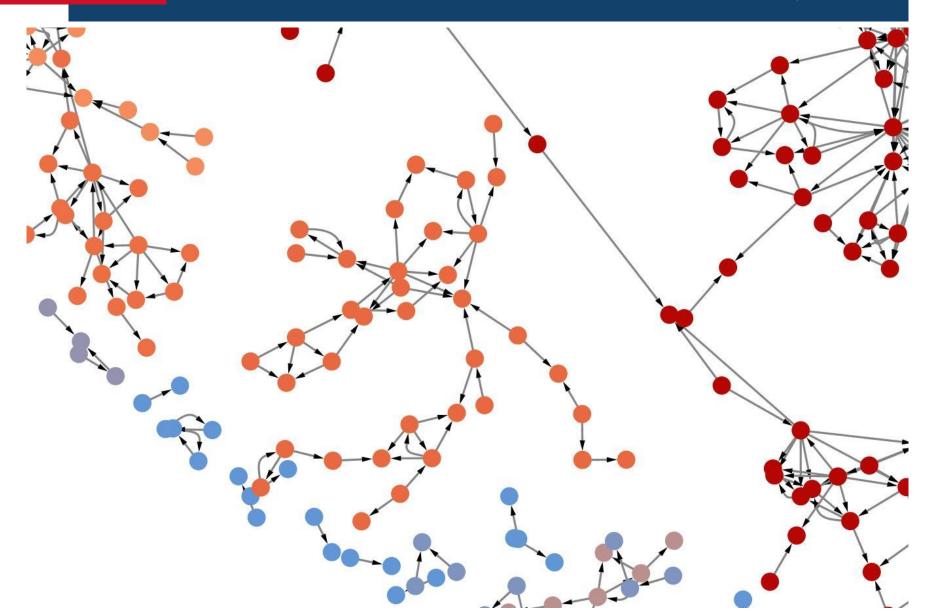
Network of MLVA profiles





Network of MLVA profiles







Edge weights

Inverse of Manhattan distance (L1-norm)

X	3	9	8	9	2	12
У	3	10	8	8	2	12
d(x,y)	0	+1	0	-1	0	0



Edge weights

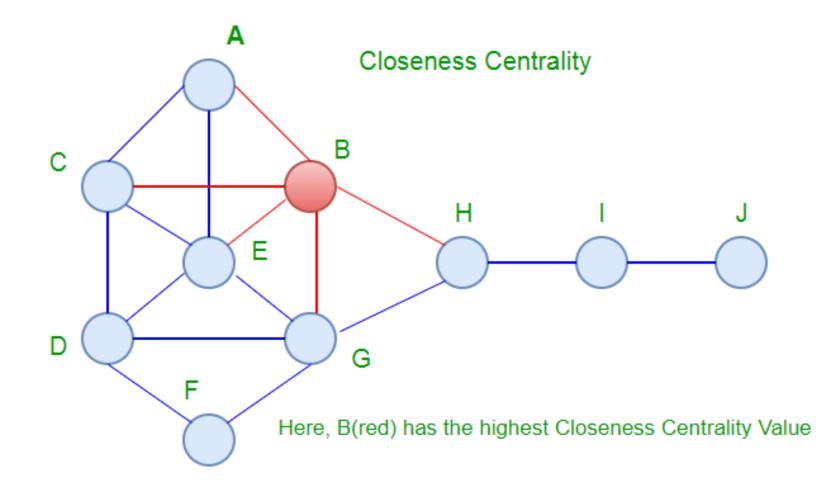
Inverse of Manhattan distance (L1-norm)

Closeness centrality
$$C(x) = \frac{1}{\sum_{y} d(x, y)}$$

Path length: average distance to all other nodes



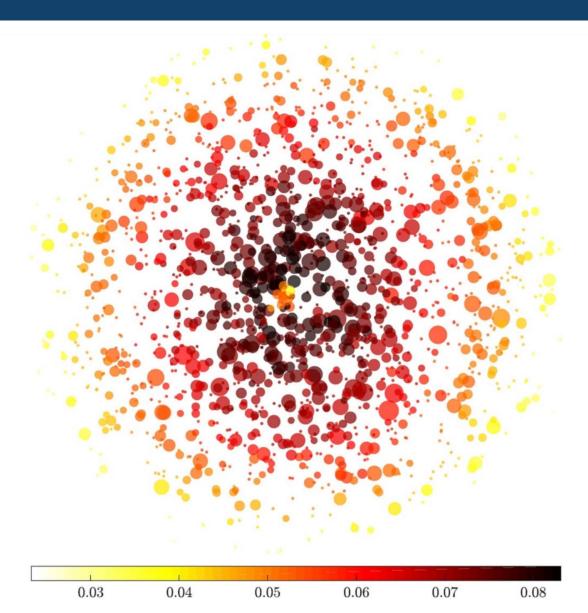
Closeness centrality (example)



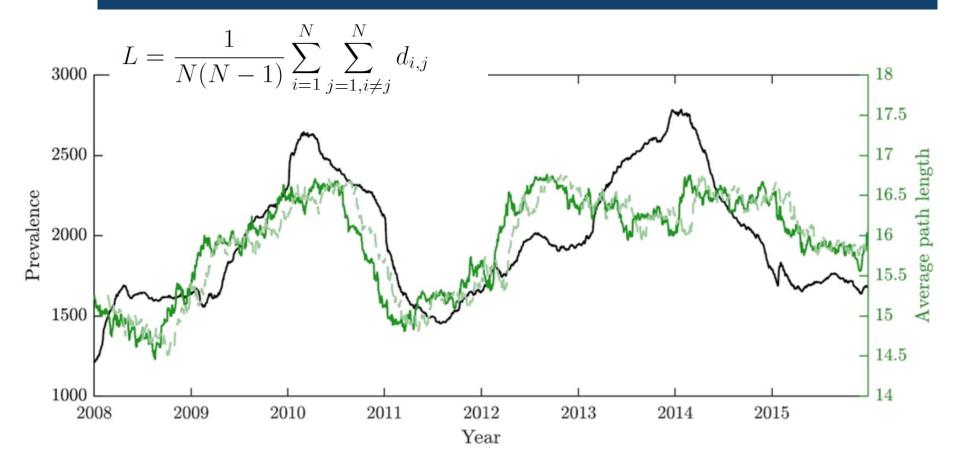
https://www.geeksforgeeks.org/closeness-centrality-centrality-measure/

Closeness centrality of MLVA profiles





Global network properties: average path length



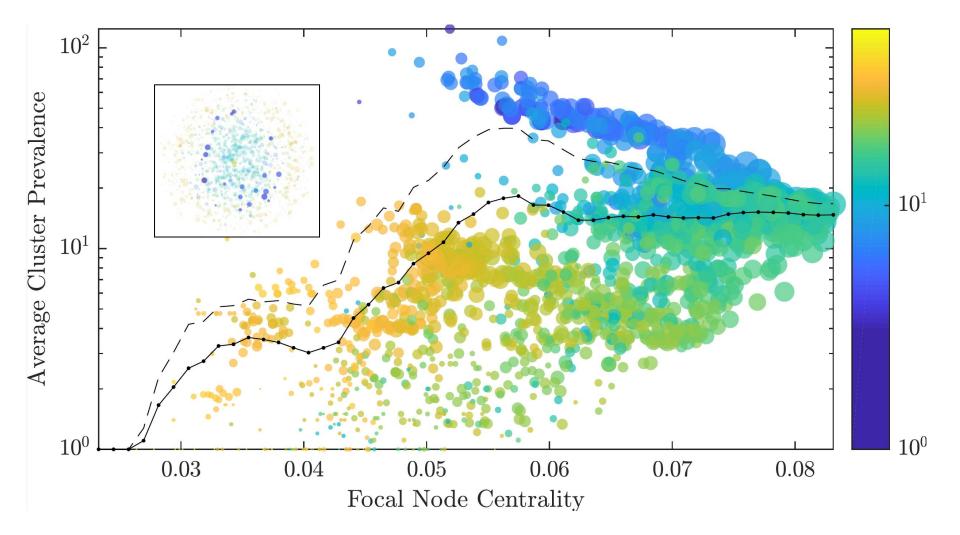
THE UNIVERSITY OF

Average path length (average distance to all other nodes)

correlates with prevalence: $\rho \approx 0.7$ at ~100 days



Centrality vs cluster prevalence



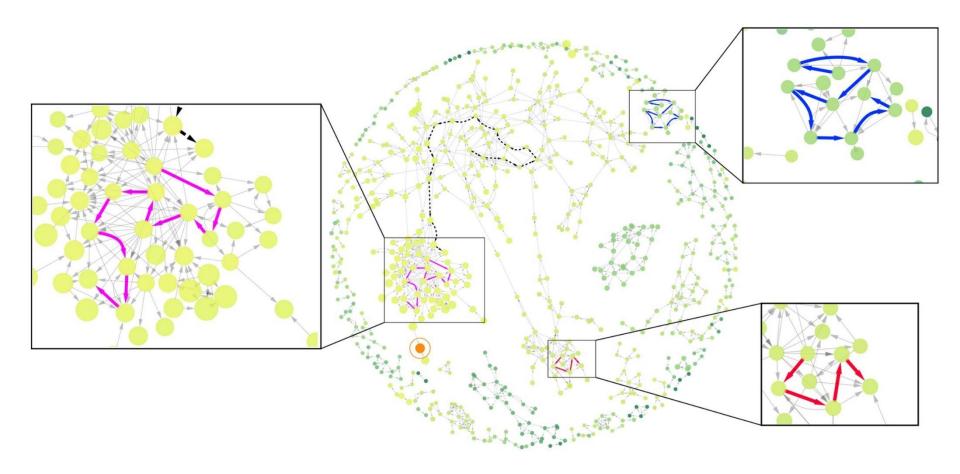
Colours denote distance to the node with the highest average cluster prevalence



Directed genotype (MLVA) networks

PLOS COMPUTATIONAL BIOLOGY

Inferring evolutionary pathways and directed genotype networks of foodborne pathogens

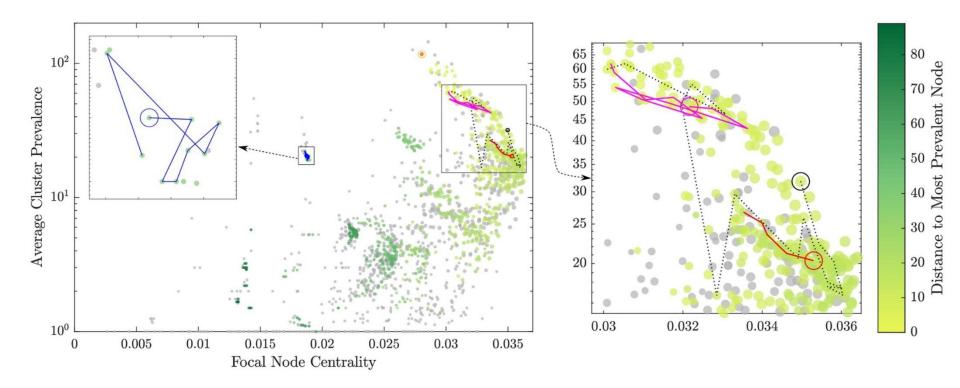




Directed genotype (MLVA) networks: evolutionary drive

PLOS COMPUTATIONAL BIOLOGY

Inferring evolutionary pathways and directed genotype networks of foodborne pathogens

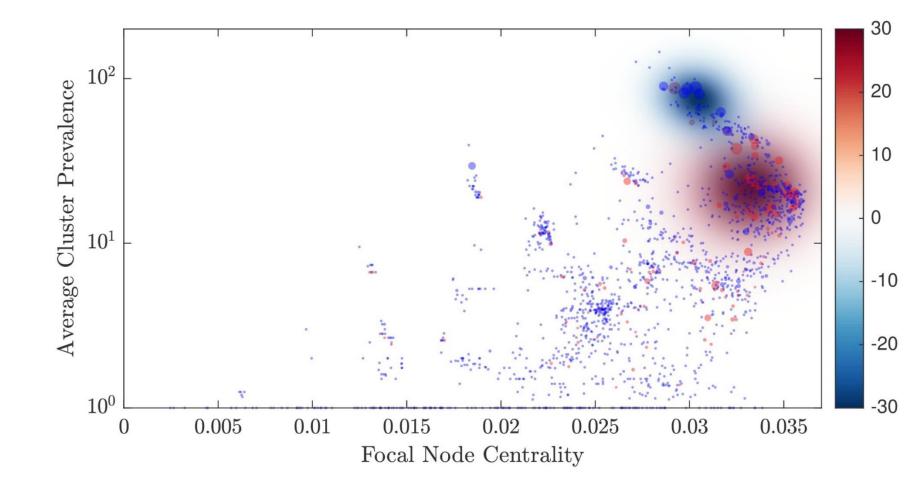




Directed genotype (MLVA) networks: transitions and bottlenecks

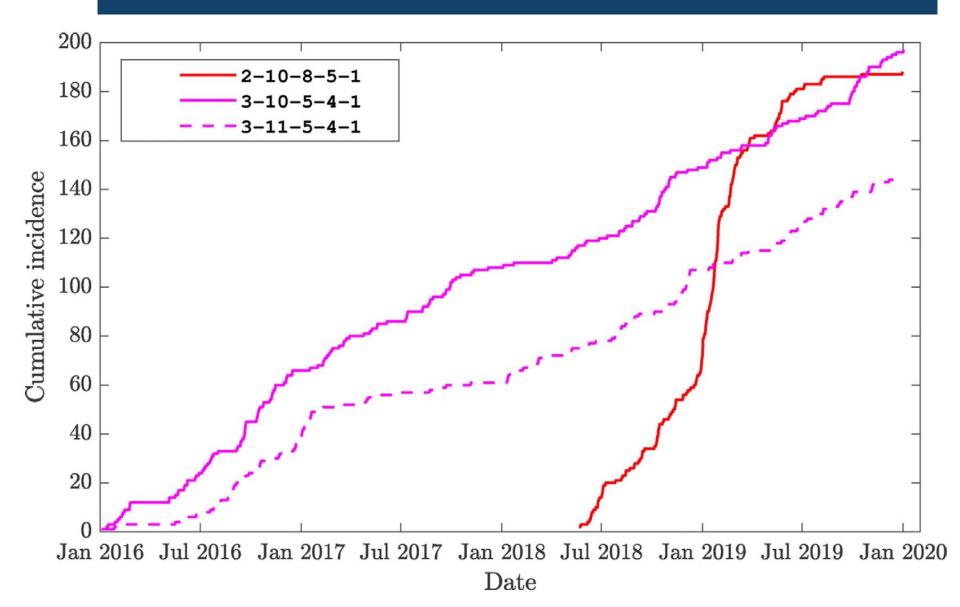
PLOS COMPUTATIONAL BIOLOGY

Inferring evolutionary pathways and directed genotype networks of foodborne pathogens



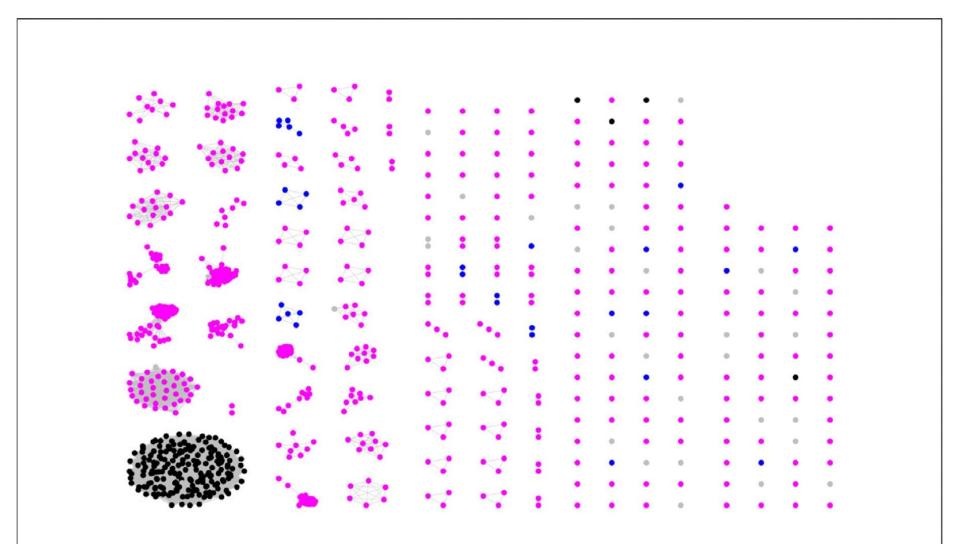
SE outbreak emergence



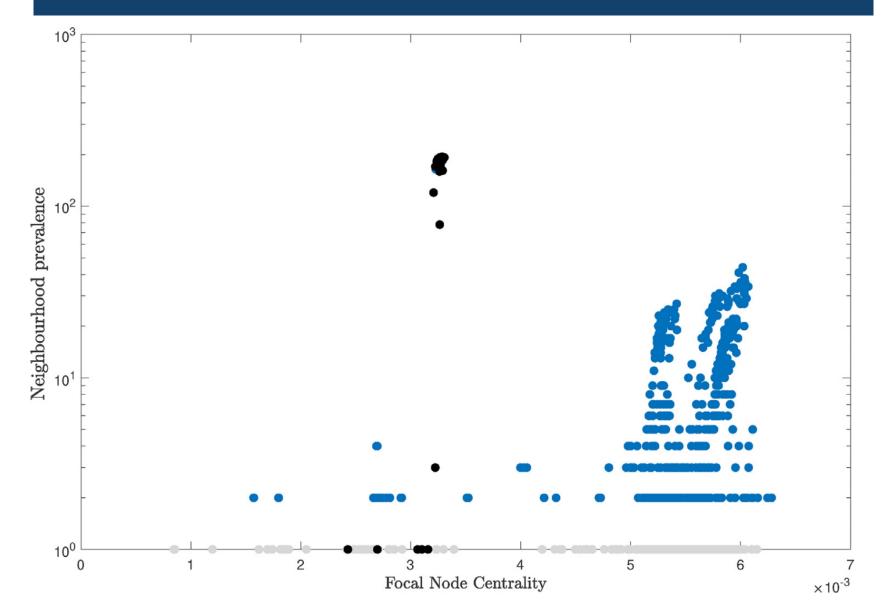




Genome-wide networks: undirected SNP sub-network



Centrality-prevalence plot: undirected SNP network



THE UNIVERSITY OF



- inferred undirected and directed networks from surveillance and molecular genotyping data for STM and SE
- quantified diversity and variability of evolving STM and SE networks
- correlated network properties with the epidemic severity
- identified two distinct evolutionary branches in terms of centrality
- suggested to monitor ongoing STM and SE population diversity and focus on new genotypes as reservoirs of future epidemics



- O. M. Cliff, V. Sintchenko, T. C. Sorrell, K. Vadlamudi, N. McLean, M. Prokopenko, Network properties of Salmonella epidemics, *Scientific Reports*, 9, 6159, 2019.
- O. M. Cliff, N. McLean, V. Sintchenko, K. M. Fair, T. C. Sorrell, S. Kauffman, M. Prokopenko, Inferring evolutionary pathways and directed genotype networks of foodborne pathogens, *PLoS Computational Biology*, 16(10): e1008401, 2020.
- A. J. Svahn, S. L. Chang, R. J. Rockett, O. M. Cliff, Q. Wang, A. Arnott, M. Ramsperger, T. C. Sorrell, V. Sintchenko, M. Prokopenko, Genome-wide networks reveal emergence of epidemic strains of Salmonella Enteritidis, *International Journal of Infectious Diseases*, 117: P65-73, 2022.
- A. J. Svahn, C. J. E. Suster, S. L. Chang, R. J. Rockett, E. M. Sim, O. M. Cliff, Q. Wang, A. Arnott, M. Ramsperger, T. C. Sorrell, V. Sintchenko, M. Prokopenko, Pangenome Analysis of a Salmonella Enteritidis Population Links a Major Outbreak to a Gifsy-1-Like Prophage Containing Anti-Inflammatory Gene gogB, *Microbiology Spectrum*, 11(2): e02791-22, 2023.
- S. L. Chang, C. J. E. Suster, R. J. Rockett, A. J. Svahn, O. M. Cliff, Alicia Arnott, Q. Wang, R. Kim, B. Suliman, M. Gall, T. C. Sorrell, V. Sintchenko, M. Prokopenko, Genome entropy and network centrality contrast exploration and exploitation in evolution of foodborne pathogens, *Physical Biology*, 20, 046006, 2023.