Bioinformatics Tools in Clinical Microbiology and Infectious Disease Prevention Algorithms

Hyunjo Kim*

Department of Bioinformatics, Gachon University, South Korea

***Corresponding author:** Hyunjo Kim, Department of Bioinformatics, Gachon University, Seungnam city, Kyeunggido, South Korea, Tel: 82-10-6563-9755; Email: hyunjo6763@gmail.com

Abstract

Bioinformatics resource is the exploitation of genome sequence data for diagnostic, therapeutic, and prevention purposes. The role of clinical genomics in infectious disease is aimed at describing the structure of outbreak communities, the methodology works equally well to identify pathogens in clinical samples. Furthermore, early detection of infectious disease in outbreaks is one of the significant issues in syndromic surveillance systems. The key concept in using clinical genomics methodology is a process of generating large-scale sequence data sets that adequately sample a specimen for microbial content and then of applying computational methods to resolve the sequences into individual species, genes, pathways, or other features. It helps to provide a rapid epidemiological response and reduce morbidity and mortality. Therefore, bioinformatics model algorithms for virus transmission and prevention in terms of resistome would be taken into consideration. The relevant future study of resistome reveals strategies that can be applied in new antibiotic discoveries.

Keywords: Bioinformatics resource; Clinical microbiology; Infectious diseases surveillance; NGS; Prevention algorithm; Resistome

Abbreviations: NGS: Next-Generation Sequencing; CDS: Coding Sequence; ODEs: Ordinary Differential Equations; MICS: Minimal Inhibitory Concentration

Introduction

Bioinformatics tools and techniques analyzing Next-Generation Sequencing (NGS) data are increasingly used for the diagnosis and monitoring of infectious diseases [1-3]. It is of interest to review the application of bioinformatics tools, commonly used databases and NGS data in clinical microbiology, focusing on molecular identification, genotypic, microbiome research, antimicrobial resistance analysis and detection of unknown disease-associated pathogens in clinical specimens. Furthermore, bioinformatics tools are extensively used in the identification, characterization, and typing of all kinds of pathogens [4,5].

This followed the widespread use of genomic approaches in the diagnosis and management of viral, bacterial, and fungal infections [6,7]. Applications of bioinformatics have been used in pathogen identification, detection of virulence factors, resistance analysis, and strain typing. Next Generation sequencing (NGS) determines the DNA sequence of a complete bacterial genome in a single sequence run, and from these data, information on resistance and virulence, as well as information for typing is obtained, useful for outbreak investigation. The obtained genome data can be further used for the development of an outbreak-specific screening test. In this review, a general introduction to NGS is presented, including the library preparation and the major characteristics of the most common NGS platforms. Although NGS holds enormous promise for clinical infectious disease testing, many challenges remain, including automation, standardizing technical protocols and bioinformatics pipelines, improving reference databases, establishing proficiency testing and quality control measures, and reducing cost and turnaround time, all of which would be necessary for widespread adoption of NGS in clinical microbiology laboratories [8-10].

Furthermore, applications of NGS in the clinical setting are described, such as outbreak management, molecular case finding, characterization and surveillance of pathogens, rapid identification of bacteria using taxonomy, metagenomics approaches on clinical samples, and the determination of the transmission of zoonotic micro-organisms from animals to humans [11-13]. Finally, we share our vision on the use of NGS in personalized microbiology in the near future, pointing out specific requirements. Recently, repeated transmission of animal viruses to humans has prompted investigation of the viral, host, and environmental factors responsible for transmission via aerosols or respiratory droplets [14]. How do we determine out of thousands of virus isolates collected in animal surveillance studies each year which viruses have the potential to become airborne and hence pose a pandemic threat [15].

In this study, using knowledge from pandemic, zoonotic and epidemic viruses, we postulate that the minimal requirements for efficient transmission of an animal virus between humans are efficient virus attachment to upper respiratory tissues, replication to high titers in these tissues, and release and aerosolization of

single virus particles. Investigating airborne transmission of viruses is key to understand and predict virus pandemics [16,17]. In addition to it, antibiotics are essential for the treatment of bacterial infections and are among our most important drugs. Resistance has emerged to all classes of antibiotics in clinical use. Antibiotic resistance has, proven inevitable and very often it emerges rapidly after the introduction of a drug into the clinic. There is, therefore, a great interest in understanding the origins, scope and evolution of antibiotic resistance [18-20].

The review discusses the concept of the antibiotic resistome, which is the collection of all genes that directly or indirectly contribute to antibiotic resistance. The concept of the antibiotic resistome provides a framework for the study and understanding of how resistance emerges and evolves [21-24]. Thus, we seeks to assemble current knowledge of the resistome concept as a means of understanding the totality of resistance and not just resistance in pathogenic bacteria. Furthermore, the study of the resistome reveals strategies that can be applied in new antibiotic discoveries.

Materials and Methods

Virus Design: Route to airborne mechanism

For the purpose of virus transmission, the major challenges for virus transmission research is needed to elucidate mechanisms for transmission. The focus should be on both gain of function and loss of function approaches. Rather, the loss of function experiments are removing any crucial part and it will stop running. In analogy, mutating a transmissible virus so it no longer transmits is a pointless exercise, giving us none to little mechanistic information; there are a thousand ways to accomplish that. Gain of function experiments that only one or a few parts need adjusting but the key is determining which part(s) out of the possible thousand they are. To investigate which viral parts need before it becomes transmissible, at least two options are available, and both should be followed. We believe that the evolutionary events leading to the pandemic viruses and genetic changes in the avian-origin viral genes made these reasserting viruses airborne. Thus, we can hypothesize based on accumulated data, which viral characteristics would facilitate airborne transmission to be important for virus to become airborne. Therefore, we postulate that the minimal requirements for airborne viruses are several of these kinds of features that attachment to and replication in appropriate cells of the URT, high virus yields in the URT and virus shedding as single particles.

Transmission: Viral determinant characteristics

Human-to-human transmission of viruses can occur through direct or indirect contact and/or via aerosols and respiratory droplets. Opinions differ on the importance of each route, as data have been published support of various routes. However, efficient aerosolization of viral particles is crucial for the transmission efficiency and pandemic potential of viruses. There is no exact particle cut-off size at which transmission changes from exclusively large respiratory droplets to aerosols, but it is generally accepted that transmission occurs through aerosols for infectious particles with a diameter smaller than μm. Larger droplets do not appear to

remain suspended in air and typically travel less than 1 m before settling on environmental surfaces or on the mucosa of close contacts. Smaller particles – below 5μm -, in contrast, settle less rapidly and can therefore travel much further.

Humans exhale respiratory droplets of widely varying quantity and size. Coughing and sneezing has been documented to generate mostly aerosol particles with sizes under 1 μm. There are several pieces of evidence in support of the role of aerosols in virus transmission. A classical study has shown that much lower doses of virus are required to cause disease in human volunteers when infected via aerosols than the doses required for intranasal inoculation. Second, aerosolized influenza virus can remain infectious for prolonged periods of time, in particular at low humidity. Moreover, when virus aerosolization is blocked with UV treatment of upper room air, virus transmission can be abolished.

Consolidation algolithm for antibiotic resistome model

The antibiotic resistome is dynamic and ever expanding, yet its foundations were laid long before the introduction of antibiotics into clinical practice. Here, we revisit our theoretical framework for the resistome concept and consider the many factors that influence the evolution of novel resistance genes, the spread of mobile resistance elements, and the ramifications of these processes for clinical practice. Observing the trends and prevalence of genes within the antibiotic resistome is key to maintaining the efficacy of antibiotics in the clinic. Antimicrobial resistance is a major global health challenge. Metagenomics allows analyzing the presence and dynamics of "resistomes" (the ensemble of genes encoding antimicrobial resistance in a given microbiome) in disparate microbial ecosystems.

However, the low sensitivity and specificity of available metagenomic methods preclude the detection of minority populations (often present below their detection threshold) and/or the identification of allelic variants that differ in the resulting phenotype. Here, we describe a novel strategy that combines targeted metagenomics using last generation in-solution capture platforms, with novel bioinformatics tools to establish a standardized framework that allows both quantitative and qualitative analyses of resistomes.

Results and Discussion

Bioinformatics tools: NGS and Software

Identification and characterization of micro-organisms that cause infections are crucial for successful treatment [25], recovery and safety of patients. Sequence analyses can be used to answer different diagnostic questions, such as the genetic relationship of either bacteria or viruses, the detection of mutations in viral or bacterial genomes leading to resistance against antivirals or antibiotics [26], identification of fungi through sequence analyses of the 18S ribosomal deoxyribonucleic acid (rDNA) of the Internal Transcribed Spacer (ITS) region and identification of bacteria through sequence analyses of the 16S rDNA such as the MiSeq (Illumina) and the Ion PGM™ (ThermoFisher) [27]. This was also confirmed with the clear divergence among *P.*

mirabilis SCDR1 *Proteus mirabilis* species 300 on the proteomic level. Genome sequencing and analysis of the first spontaneous Nano silver resistant bacterium *Proteus mirabilis* strain SCDR1 using the bioinformatics tools for identifying and combating anti-microbial resistance was illustrated in (Figure 1). The biggest challenge concerning the introduction of NGS in the clinical microbiology laboratory is the data analyses. Nonetheless, even with little knowledge of bioinformatics, it is possible to perform NGS data analyses for diagnostic purposes, using the numerous user-friendly software packages available. However, for more in-depth analysis, scientific knowledge is required on the genomic features and the biological background of the micro-organism under investigation (Table 1).
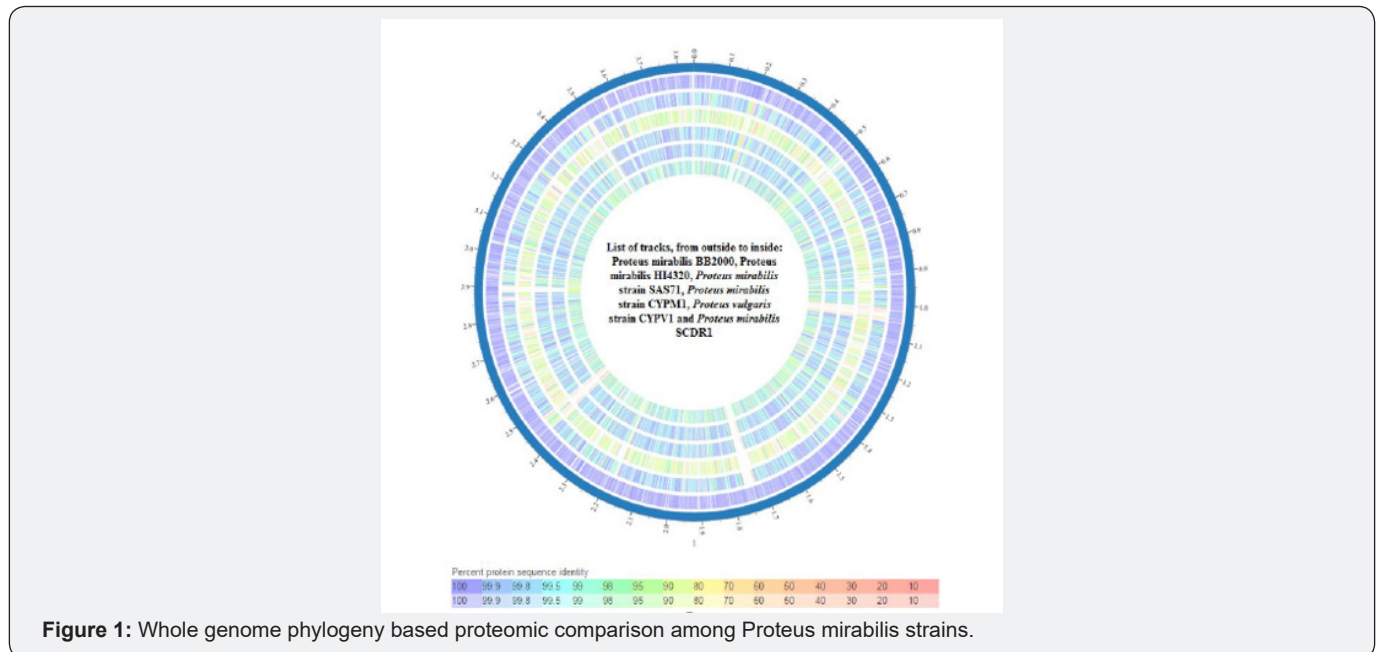

Figure 1: Whole genome phylogeny based proteomic comparison among Proteus mirabilis strains.

Table 1: Software packages frequently used for NGS data analyses through an internet.

Application	Software	Link
Annotation	Prokka	www.vicbioinformatics.com
	RAST	http://rast.nmpdr.org/
Assembly	BioNumerics	http://www.applied-maths.com/
	CLC Genomic Workbench	https://www.qiagenbioinformatics.com/
	SeqSphere	http://www.ridom.de/
	SPAdes	http://bioinf.spbau.ru/spades
	Velvet	www.ebi.ac.uk/~zerbino/velvet
Data quality check	BaseSpace	https://basespace.illumina.com
	BioNumerics	www.applied-maths.com
	CLC Genomic Workbench	www.clcbio.com
	FastQC	www.bioinformatics.babraham.ac.uk
Identification	K-merFinder	www.genomicepidemiology.org
	NCBI BLAST	www.ncbi.nlm.nih.gov/blast
Metagenomics	MEGAN	http://ab.inf.uni-tuebingen.de/software/malt
Phylogeny	FastTree	www.microbesonline.org/fasttree
	RAxML	http://sco.h-its.org/exelixis/software.html
	SeqSphere	www.ridom.de
	SNPTree	www.genomicepidemiology.org
Resistance	ARDB	https://ardb.cbcb.umd.edu
	CARD	https://card.mcmaster.ca
	ResFinder	www.genomicepidemiology.org

SNP calling	BioNumerics	www.applied-maths.com
	CLC Genomic Workbench	www.clcbio.com
	Samtools	www.htslib.org
	SeqSphere	www.ridom.de
Typing (wgMLST)	BIGSdb	http://bigsdb.readthedocs.io
	BioNumerics	www.applied-maths.com
	CLC Genomic Workbench	www.clcbio.com
	EnteroBase	https://enterobase.warwick.ac.uk
	SeqSpere	www.ridom.de
Virulence	VFDB	www.mgc.ac.cn/VFs
	VirulenceFinder	www.genomicepidemiology.org
Visualisation & Comparative study	ACT	www.sanger.ac.uk/science/tools
	Artemis	www.sanger.ac.uk/science/tools
	BRIG	https://sourceforge.net/projects/brig/
	ClustalW	www.genome.jp/tools/clustalw
	DNA plotter	www.sanger.ac.uk/science/tools
	WebACT	www.webact.org

Virus transmission

A virus-spread mechanism is related to inter-flat or interzonal airflow through open windows caused by buoyancy effects. Both on-site measurements and numerical simulations quantify the amount of the exhaust air that exits the upper part of the window of a floor and re-enters the lower part of the open window of the immediately upper floor. Ventilation air could contain up to 7% (in terms of mass fraction) of the exhaust air from the lower floor. The transmission route was reconstructed by epidemiological and genomic data (Figures 2-4). Each node represents a patient, and an arrow indicates a possible transmission event from one patient to another. The blue arrow with solid line represents a direct transmission event supported by both epidemiological data and genetic data, the blue arrow with dash line represents an indirect transmission (e.g. via environment) supported by epidemiological data, and the red arrow indicates the equally parsimonious transmission link which cannot be resolved by neither epidemiological data nor genetic data.
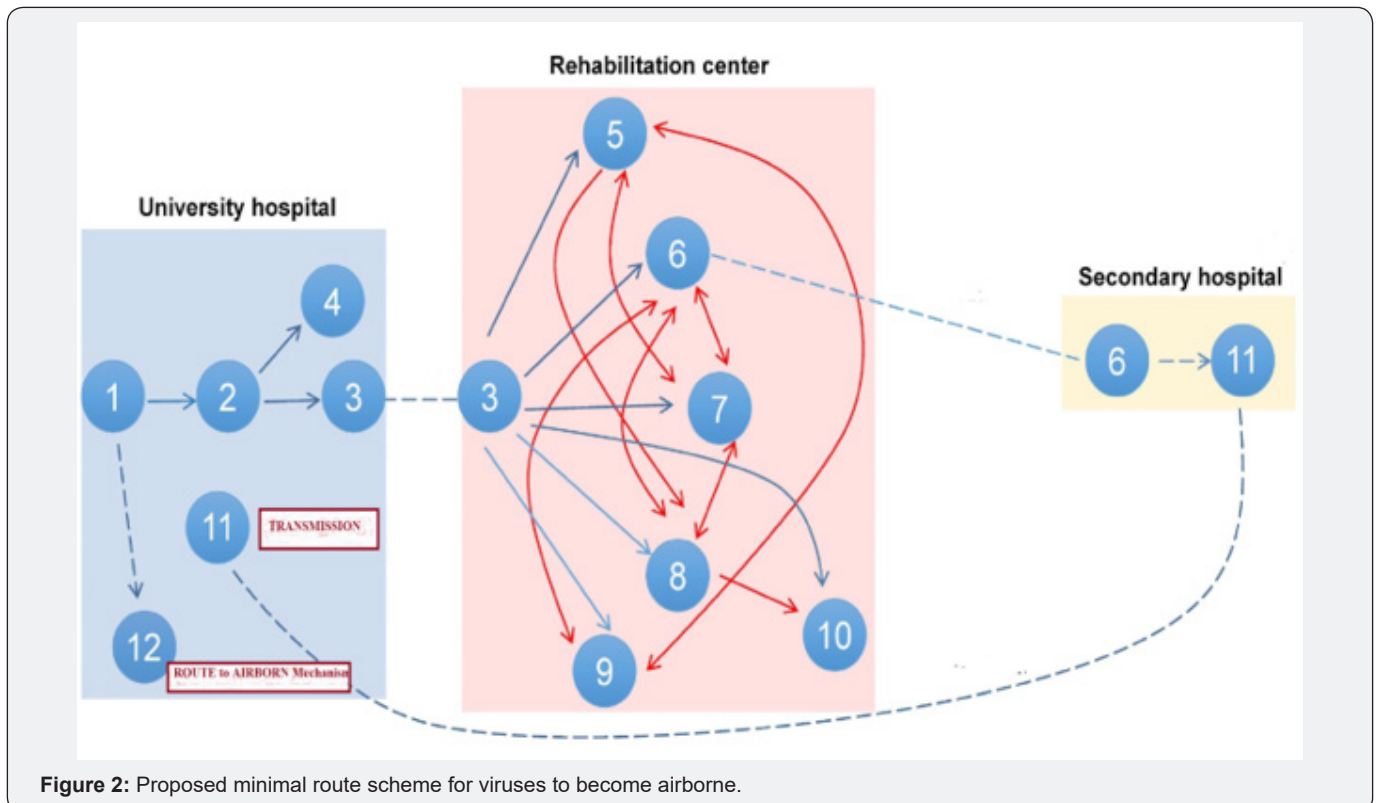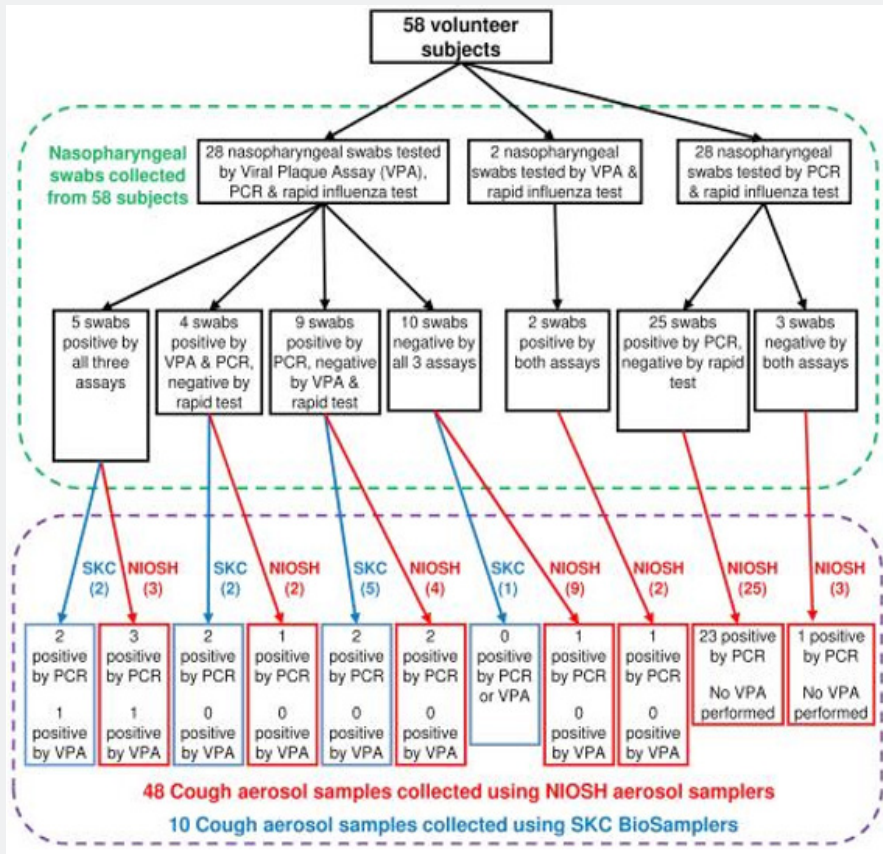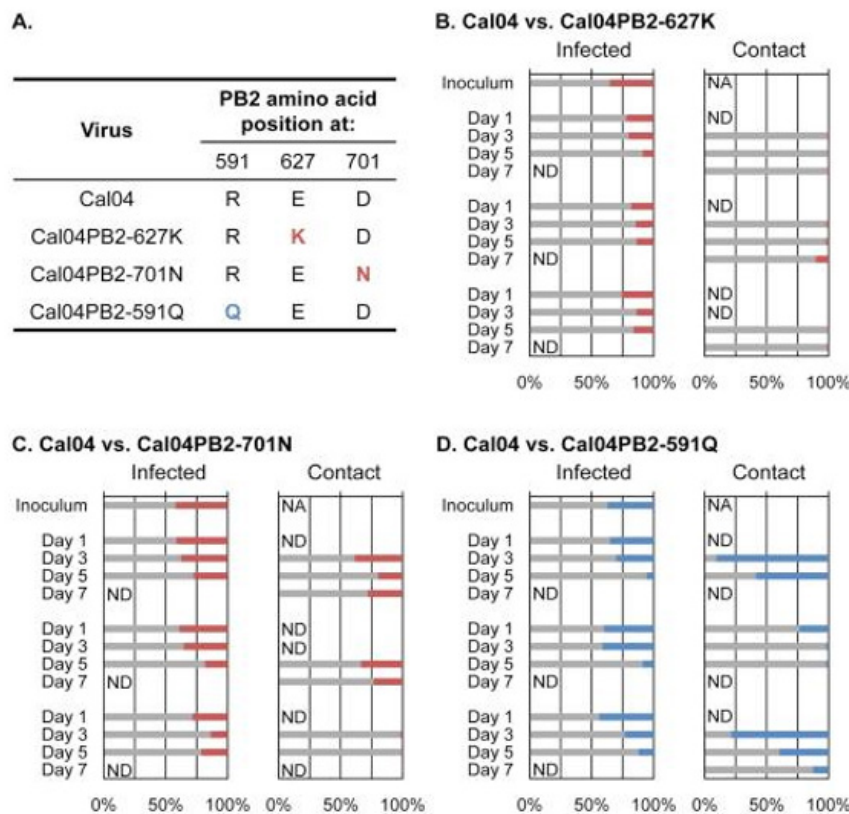

Figure 2: Proposed minimal route scheme for viruses to become airborne.

Figure 3: Flow chart showing patients and tests performed. The numbers in parentheses are the number of aerosol samples collected using SKC BioSamplers and NIOSH aerosol samplers.

Figure 4: Competitive transmission studies in ferrets for wild-type and mutant pandemic H1N1 viruses.

The inter-institutional transfer of the patient is shown by dash lines, on which the distance between institutions is indicated. The red star represents an outbreak at a secondary hospital, but the isolates were unavailable for further research. Shown figures in Figure 3 and 4 are presented its mutants tested in competitive transmission studies. On the other hand, the bacterial bioinfor-

matics database and analysis resource (PATRIC) gene annotation analysis showed that 261 the number of the observed Coding Sequence (CDS) is 4423, rRNA is 10 and tRNA is 71. The unique gene 262 count for the different observed metabolic pathways is 2585 (Figure 5).
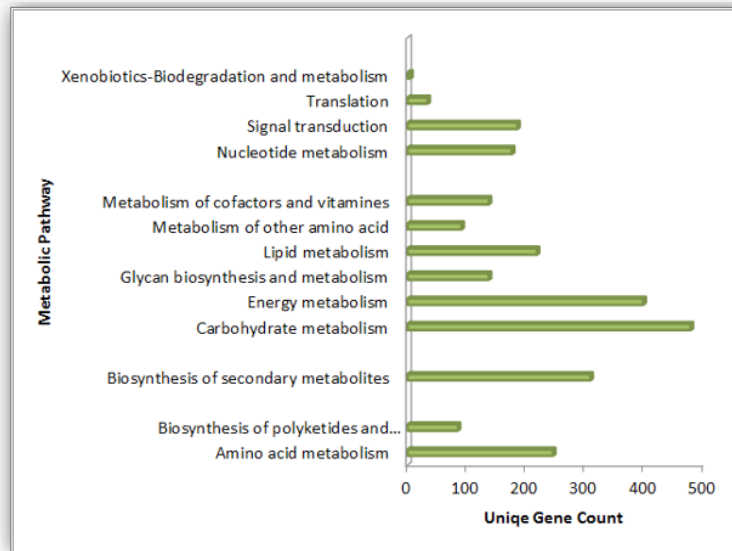

Figure 5: Distribution of unique gene counts amongst different metabolic pathways.

A resistome algorithm model

The small cross-hatched boxes represent the antibiotics and resistance genes of relevance to clinical practice. Respectively, these are a small subset of the world of small bioactive molecules (the parvome), and the world of potential resistance determinants (the resistome). The resistome comprises the genes

that potentially encode resistance to antibiotics. The mobilome comprises the mobile proportion of bacterial genomes. The mobilome and resistome overlap, since many resistance genes are located on mobile elements. Both the resistome and mobilome are a subset of the total coding capacity of prokaryotic cells, the pangenome, which is expressed as the panproteome.
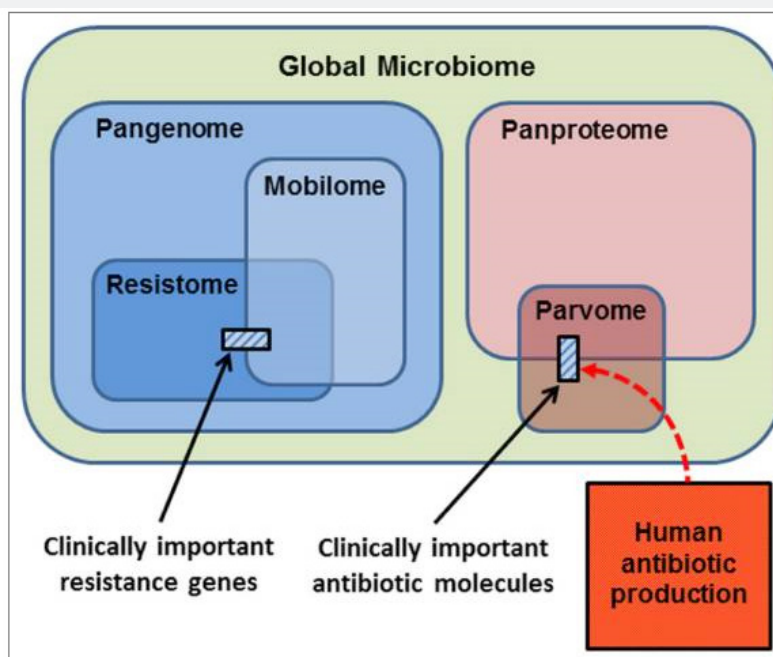

Figure 6: Conceptual representation of the biological molecules of relevance to antibiotic resistance.

Note that only a small proportion of the parvome is utilized by humans for antibiotic purposes, and that the scale of commercial antibiotic production probably overwhelms the natural production of these molecules by the entire global microbiota (Figure 6). A recent study in Figure 7 compared the detection of viruses in known respiratory virus-positive samples and not previously analysed nasopharyngeal swabs by an RNA sequencing-based metagenomics approach with a more conventional molecular method. The data-analyses was performed using Taxonomer, a rapid and interactive, web-based metagenomics data-analyses tool. Overall, the metagenomics approach had a high agreement with the molecular method, detected viruses not targeted by the molecular method, and yielded epidemiologically and clinically relevant sequence information (Figure 7).
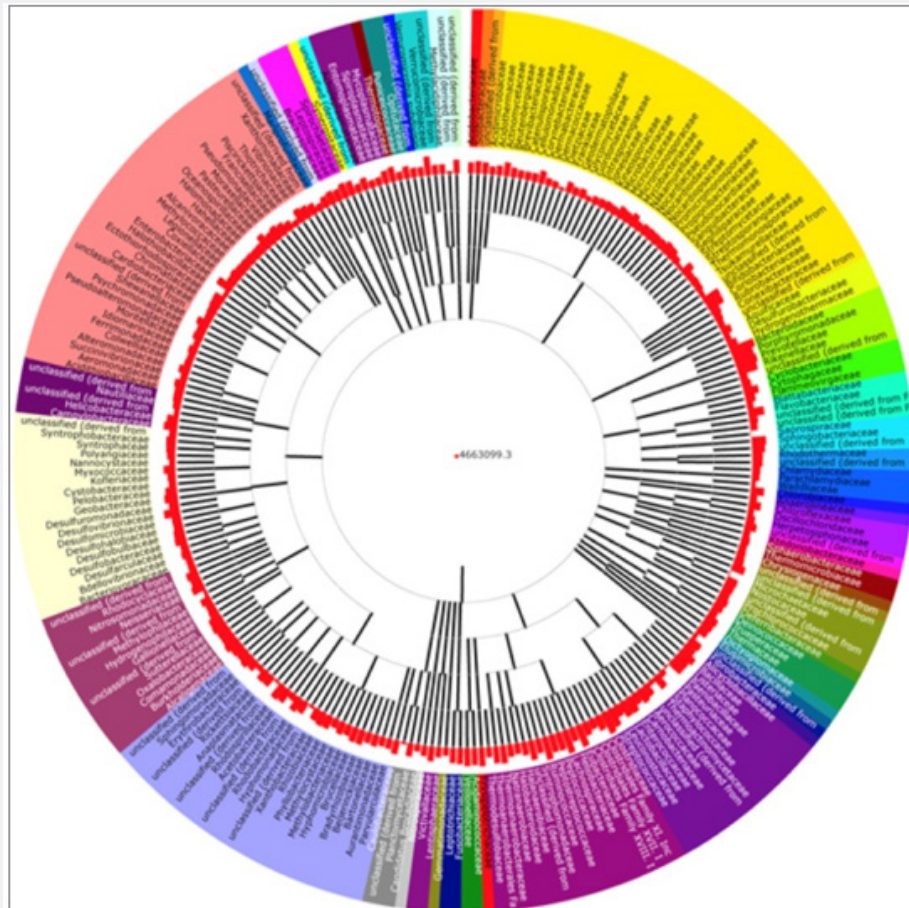

Figure 7: An example of an output of a metagenomics approach of a faecal sample. The different colours represent different bacterial families.

Prediction mathematical model to infection

In this study, we propose that patient-to-patient variability sets a fundamental limit on outcome prediction accuracy for a general class of mathematical models for the immune response to infection. We investigate several systems of Ordinary Differential Equations (ODEs) that model the host immune response to a pathogen load. Advantages of systems of ODEs for investigating the immune response to infection include the ability to collect data on large numbers of 'virtual patients', each with a given set of model parameters, and obtain many time points during the course of the infection. We implement patient-to-patient variability v in the ODE models by randomly selecting the model parameters from distributions with coefficients of variation v that are centered on physiological values.

We use logistic regression with one-versus-all classification to predict the discrete steady-state outcomes of the system. We find that the prediction algorithm achieves near 100% accuracy for $v = 0$, and the accuracy decreases with increasing v for all ODE models studied. The fact that multiple steady-state outcomes can be obtained for a given initial condition, i.e. the basins of attraction overlap in the space of initial conditions, limits the prediction accuracy for $v > 0$. Increasing the elapsed time of the variables used to train and test the classifier, increases the prediction accuracy, while adding explicit external noise to the ODE models decreases the prediction accuracy. Our results quantify the competition between early prognosis and high prediction accuracy that is frequently encountered by clinicians. Clinicians need to predict patient outcomes with high accuracy as early as possible after disease inception (Figure 8). Therefore, the early elucidation of resistance mechanisms using WGS also has implications for the design of clinical trials.
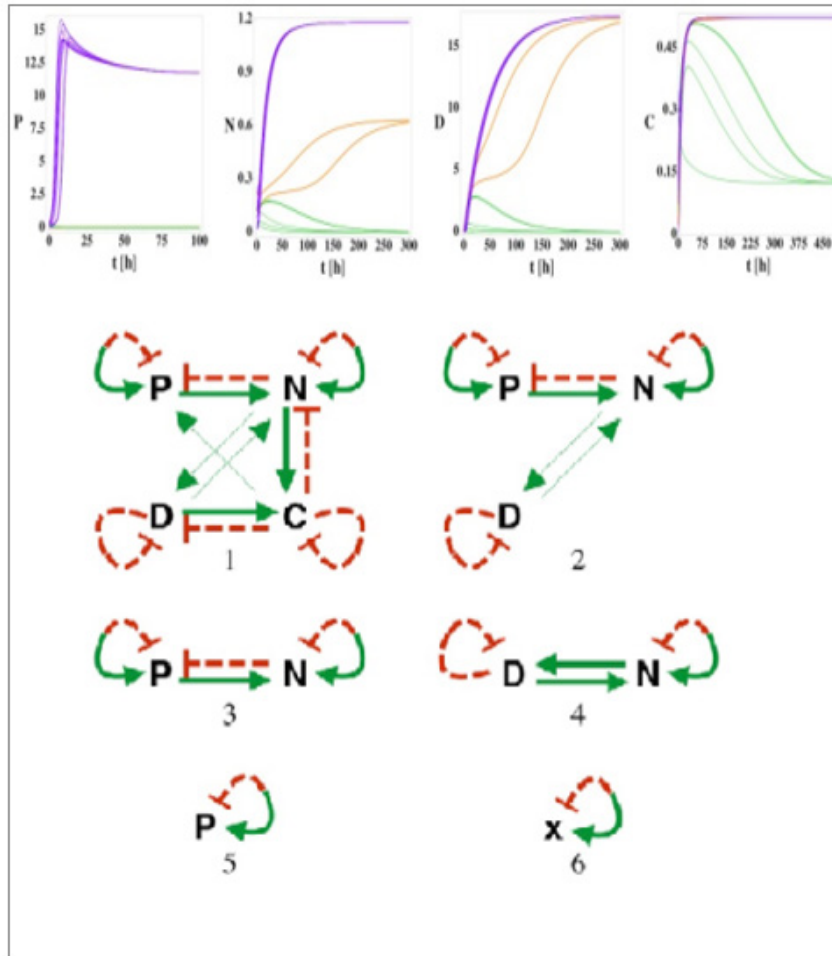
Figure 8: P, N, D, and C versus time for model from 20 random initial conditions with no patient variation (Top) and Interactions between variables for models(bottom).

If resistance mechanisms are discovered that only result in marginally increased Minimal Inhibitory Concentrations (MICs) compared with the wild type MIC distributions, more frequent dosing or higher doses could be employed in clinical trials to overcome this level of resistance. Moreover, the discovery of cross-resistance between agents using WGS can influence the choice of antibiotics that are included in novel regimens. WGS has become an essential tool for drug development by enabling the rapid identification of resistance mechanisms.

Conclusion

Bioinformatics tools such as Next-generation sequencing (NGS) technologies are increasingly being used for diagnosis and monitoring of infectious diseases. Herein, we reviewed the application of NGS in clinical microbiology, focusing on genotypic resistance testing, direct detection of unknown disease-associated pathogens in clinical specimens, investigation of microbial population diversity in the human host, and strain typing applications in clinical virology,

Acknowledgement

References

1. Zhang Y, Manjunath M, Kim Y, Heintz J, Song JS (2018) SequencEnG: an Interactive Knowledge Base of Sequencing Techniques. Bioinformatics. Ning S, Li X (2018) Non-coding RNA Resources. Adv Exp Med Biol 1094: 1-7.

2. Lefranc MP, Ehrenmann F, Kossida S, Giudicelli V, Duroux P (2012) Use of IMGT® Databases and Tools for Antibody Engineering and Humanization. Methods Mol Biol 907: 3-37.

3. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, et al. (2012) Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 380(9859): 2095-128.

4. European Centre for Disease Prevention and Control. (2013) Health Inequalities, the Financial Crisis, and Infectious Disease in Europe. Stockholm: ECDC.

5. Buehler SS, Madison B, Snyder SR, Derzon JH, Cornish NE, et al. (2016) Effectiveness of practices to increase timeliness of providing targeted therapy for inpatients with bloodstream infections: a laboratory medicine best practices systematic review and meta-analysis. Clin Microbiol Rev 29(1): 59-103.

6. Wurtz R, Cameron BJ (2005) Electronic laboratory reporting for the infectious diseases physician and clinical microbiologist. Clin Infect Dis 40(11): 1638-1643.

7. Panackal AA, M'Ikanatha NM, Tsui FC, McMahon J, Wagner MM, et al. (2002) Automatic electronic laboratory-based reporting of notifiable infectious diseases at a large health system. Emerg Infect Dis 8(7): 685-691.

8. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW (2012) Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet 13(9): 601–612.

9. Köser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, et al. (2012) Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS Pathog 8(8): 10-18.

10. Grad YH, Lipsitch M, Feldgarden M, Arachchi HM, Cerqueira GC, et al, (2012) Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. Proc Natl Acad Sci U S A 109(8): 3065-3070.

11. McGann P, Bunin JL, Snesrud E, Singh S, Maybank R, et al. (2016) Real time application of whole genome sequencing for outbreak investigation-what is an achievable turnaround time? Diagn Microbiol Infect Dis 85(3): 277-282.

12. Gordon NC, Price JR, Cole K, Everitt R, Morgan M, et al. (2014) Prediction of *Staphylococcus aureus* antimicrobial resistance by whole-genome sequencing. J Clin Microbiol 52(4): 1182-1191.

13. Su H, Yang X, Wang S, Shi H, Liu X (2018) Effect of annexin II-mediated conversion of plasmin from plasminogen on airborne transmission of H9N2 avian influenza virus. Vet Microbiol 223: 100-106.

14. Zhang N, Li Y (2018) Transmission of Influenza A in a Student Office Based on Realistic Person-to-Person Contact and Surface Touch Behaviour. Int J Environ Res Public Health 15(8): E1699.

15. Donald PR, Diacon AH, Lange C, Demers AM, von Groote-Biddlingmeier F, et al. (2018) Droplets, dust and guinea pigs: an historical review of tuberculosis transmission research, 1878-1940. Int J Tuberc Lung Dis 22(9): 972-982.

16. Mermel LA (2018) The great influenza centennial-what have we learned about the epidemiology and prevention of transmission? Clin Microbiol Infect S1198-743X (18): 30494-4.

17. Baur D, Gladstone BP, Burkert F, Carrara E, Foschi F, et al. (2017) Effect of antibiotic stewardship on the incidence of infection and colonisation with antibiotic-resistant bacteria and Clostridium difficile infection: a systematic review and meta-analysis. Lancet Infect Dis 17(9): 990-1001.

18. Kho AN, Doebbeling BN, Cashy JP, Rosenman MB, Dexter PR, et al. (2013) A regional informatics platform for coordinated antibiotic-resistant infection tracking, alerting, and prevention. Clin Infect Dis 57(2): 254-262.

19. Leopold SR, Goering RV, Witten A, Harmsen D, Mellmann A (2014) Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. J Clin Microbiol 52(7): 2365-2370.

20. Taylor SL, Leong LEX, Mobegi FM, Choo JM, Burr LD, et al. (2018) Understanding the impact of antibiotic therapies on the respiratory tract resistome: a novel pooled-template metagenomic sequencing strategy. Multidisciplinary Respiratory Medicine. 13(Suppl 11): 30.

21. Kang K, Ni Y, Li J, Imamovic L, Sarkar C, et al. (2018) The Environmental Exposures and Inner- and Intercity Traffic Flows of the Metro System May Contribute to the Skin Microbiome and Resistome. Cell Rep 24(5): 1190-1202.

22. Wang Y, Hu Y, Zhu B, Jiao X, Gao GF (2018) Antibiotic resistome in farm animals and their related environments: a review. 34(8): 1226-1233.

23. National Electronic Disease Surveillance System Working Group (2011) National Electronic Disease Surveillance System (NEDSS): a standards-based approach to connect public health and clinical medicine. J Public Health Manag Pract 7(6): 43-50.

24. O Brien TF, Stelling J (2011) Integrated multilevel surveillance of the world's infecting microbes and their resistance to antimicrobial agents. Clin Microbiol Rev 24(2): 281-295.

25. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, et al. (2011) Rapid pneumococcal evolution in response to clinical interventions. Science 331(6016): 430-434.

26. Shao Z, Li W, Ren J, Liang X, Xu L, et al. (2010) Identification of a new Neisseria meningitidis serogroup C clone from Anhui province, China. Lancet 367(9508): 419-423.

27. Zhou H, Gao Y, Xu L, Li M, Li Q, et al. (2012) Distribution of serogroups and sequence types in disease-associated and carrier strains of Neisseria meningitidis isolated in China between 2003 and 2008. Epidemiol. Infect 140(7):1296-1303.

**Your next submission with Juniper Publishers
will reach you the below assets**

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
  (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
https://juniperpublishers.com/online-submission.php