The New Tools Revolutionising *Vibrio* Science

Craig Baker-Austin[1*], Joaquin Trinanes[2,3,4], and Jaime Martinez-Urtaza[1,5]

[1]Centre for Environment, Fisheries and Aquaculture (CEFAS), Weymouth, Dorset DT4 8UB, United Kingdom

[2]CRETUS, University of Santiago de Compostela, Campus Universitario Sur, Santiago de Compostela 15782, Spain

[3]Atlantic Oceanographic and Meteorological Laboratory, National Oceanic and Atmospheric Administration, 4301 Rickenbacker Causeway, Miami, FL 33149, USA

[4]Cooperative Institute for Marine and Atmospheric Studies, Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149, USA.

[5]Department of Genetics and Microbiology, Facultat de Biociències-Edifici C, Campus Universitat Autònoma de Barcelona (UAB), 08193 Bellaterra, Barcelona, Spain

Key words: *Vibrio*, next generation sequencing, remote sensing, data visualisation

*Author for editorial correspondence

Dr Craig Baker-Austin
Centre for Environment, Fisheries and Aquaculture (CEFAS),
Weymouth,
Dorset,
DT4 8UB,
United Kingdom
craig.baker-austin@cefas.co.uk
Tel : +44 (0)1305 206619
Fax : +44 (0)1305 206600

For publication in *Environmental Microbiology*

OPINION

As microbiologists we live in exciting times. A variety of technical and conceptual developments, particularly in the last decade have revolutionized the field of microbiology, redrawing the landscape, and entirely redefining what is possible. Perhaps this paradigm shift is no more apparent than in the study of vibrios. The family *Vibrionaceae* are almost unique as a group of bacteria to study in microbiology: they are genomically, phylogenetically and functionally diverse yet a distinct group of environmental bacteria encompassing important human and animal pathogens as well as non-pathogenic species such as ecologically critical symbionts. Sensitive to physiochemical stimulii, they are amongst the fasting replicating bacteria studied, capable of responding almost immediately to favorable environmental conditions such as those afforded by climate warming. Characterized by an unusual double chromosome and frequently carrying numerous cryptic plasmids – their genomes are often pockmarked with insertion elements, transposons, prophages and integrases – paying testament to past genomic promiscuity. With a strong affinity for environmental niches in freshwater and marine systems, they are amongst the most numerous bacteria present in our oceans, coasts and freshwater environments. As such they offer something for almost anyone interested in microbiology and represent an excellent example of field of microbiology that has benefitted hugely by advances across a gamut of disciplines – not just microbiological - but encompassing genomics, genetics, oceanography, ecological, earth observations sciences, and data visualization, among others. We'll briefly outline some of the most exciting, innovative and translational scientific advances that are currently being applied to these ecologically, environmentally and clinically important bacteria.

Perhaps the most obvious advance in microbiology in the last decade has been the revolution in genomics and whole genome sequencing. Where once deciphering a microbial genome represented a time consuming, expensive and technically demanding ordeal, a standard (~4 Mb) bacterial genome can now be sequenced cheaply within hours, and then

assembled and analyzed with limited expertise. This rapid step change in the capacity to sequence and subsequently scrutinize microbial genomes has completely redefined our understanding of vibrios, and in particular pathogenic species such as *V. cholerae, V. vulnificus* and *V. parahaemolyticus.* Early sequencing efforts focused on these bacterial species revealed key aspects related to their genome structure, virulence capabilities and their evolutionary provenance (Heidelberg *et al.*, 2000; Chen *et al.*, 2003; Makino *et al.*, 2003), but recent studies have allowed us to delve further and deeper into the genomic architecture to allow us to ask key biological, ecological, epidemiological and evolutionary questions. A variety of studies have now successfully applied next generation sequencing approaches to study the source of outbreaks (Orata, Keim and Boucher, 2014), their evolutionary trajectory to infer routes of transmission (Martinez-Urtaza *et al.*, 2017) as well as retrospectively reconstructing the dynamics of *Vibrio* outbreaks (Mutreja *et al.*, 2011; Domman *et al.*, 2017).

It cannot be understated how this has changed how we understand these bacteria: these advances have helped to provide a clearer picture of how bacteria spread, emerge and causes disease - often in areas where infections have never previously been reported (Colwell, 1996). The use of genomic sequencing has also allowed us to pose more ecological questions related to these bacteria in the environment, such as apparent cooperation between species (Cordero *et al.*, 2012), spatial and temporal resource partitioning among *Vibrionaceae* strains (Hunt *et al.*, 2008), and the spread of functional genes in *Vibrio* communities (Hehemann *et al.*, 2016). These studies using oceanic assemblages of vibrios reveal a startling diversity and complexity – which we are only now starting to comprehend. Alongside being typically faster, cheaper high-throughput sequencing typically provides greater resolution and granularity than traditional subtyping methods such as PCR, multi-locus sequence typing, pulsed field gel electrophoresis and serotyping, which were the approaches of choice to distinguish and differentiate strains in the past. In much the same

way, the ability to sequence RNA using these high-throughput approaches and study the comparative expression of genes in both clinical and environmental settings has greatly expanded our understanding of how these bacteria can harmlessly exist in seawater, but also to initiate infections in both humans and animals (Livny *et al.*, 2014; Williams *et al.*, 2014). Where the advent of high throughput sequencing has expanded these horizons, it has also heralded the need for "big data" approaches to adequately scrutinize these complex genomic datasets and to allow us to ask fundamental biological questions.

To this end, there are now easy to use online tools to deposit *Vibrio* genomic data alongside continually-updated and publicly available depositories that utilize powerful analytic and visualization outputs. Currently, these can be used to assess the relationships between strains (e.g. environmental or clinical strains), their evolutionary history and even how they have evolved (and how fast) over time. For instance, because mutation is the fuel of evolutionary change and because we can infer this rate of change over time, where metadata is available regarding a specific sequenced strain, this can now be used to build a clearer understanding of how bacteria have emerged, spread and evolved over time – fundamental biological questions that were impossible to explore even a decade ago. Online tools such as NextStrain (https://nextstrain.org/) and Microreact (https://microreact.org/) which offer open-source and real-time tracking of key pathogens with easy to use data visualization are excellent examples of approaches that have transformed how these bacteria can be studied.

Because most clinically-relevant *Vibrio* species are sensitive to temperature and salinity amongst other physio-chemical parameters (Vezzulli *et al.*, 2016; Baker-Austin *et al.*, 2017), earth observation systems are now being routinely used to track, understand and predict potential outbreaks (Semenza et al., 2017). These advances cannot be understated – vibrios were the first non-vector borne group of pathogens studied in this manner using

remote sensing methods (Lobitz *et al.*, 2000; Martinez-Urtaza *et al.*, 2008, 2010), and the first pathogen group where risks can be ascribed in real-time (Semenza et al., 2017). Indeed, epidemiological surveillance data now makes it possible to retrospectively explore the role of climate and environmental perturbations in modulating disease risk by allowing the pinpointing in time and space of different outbreaks, facilitating a more cohesive understanding of the factors driving disease transmission than is achieved during real time observations (Figure 1). The potential to be able to monitor in real-time, and potentially predict the occurrence of disease outbreaks is a critically important tool in managing the risks associated with pathogenic vibrios. Similarly, predictive models based on remote sensing data have been used successfully elsewhere to assess risk associated with this group of bacteria (Lobitz *et al.*, 2000; Baker-Austin *et al.*, 2013; Grimes *et al.*, 2014). These tools are being continually refined, improved and finessed. Improvements in sensor technology - providing hyperspectral imaging at higher spatial and radiometric resolutions; satellite constellations - usually working together to provide improved coverage and backup coupled to computing and networking capacity improvements will improve these methodologies further still, allowing greater resolution and granularity of risk mapping. We believe that routine earth observation monitoring of marine environments, and in particular marine systems with low salinity offers a cost-effective, easy to understand and actionable approach for risk assessment purposes for these emerging pathogens. These are clearly tools that will become more important in the future, given continued and rapid coastal warming, coupled to an increasing global population and expansion in seafood and shellfish consumption (Baker-Austin *et al.*, 2017).

Where Next?

The pace of technological change being applied to study vibrios is truly breath-taking, and so too are the creative approaches that researchers are using to study these bacteria across

different environmental, ecological, genomic and clinical landscapes. Currently, one of the most exciting technological advances is in the use of long-read sequencing methods such as nanopore to quickly and easily sequence bacterial genomes (Allué-Guardia *et al.*, 2018) as well as utilise it as a screening approach in metagenomic studies (Acharya *et al.*, 2019). Up until now, the vast majority of the sequencing-based studies analysed the genetic diversity, population structure and evolution of pathogenic bacteria and were based on genomic data generated from cultured strains.

The recent improvement of analytical approaches to sequencing metagenomic DNA will facilitate the generation of whole genome sequences of bacterial populations without culturing - providing a more realistic proxy of the composition and diversity of *Vibrio* populations in clinical and environmental samples. Furthermore, as metagenomic analysis genomic data are generated directly from samples, genomes can be reconstructed rapidly and can be integrated with data from other sources to produce an accurate picture of the ecological and epidemiological landscape. Advances in sequencing technologies are currently moving so fast that approaches capable of generating genomic data and analysis tools to scrutinise these datasets can be completed in almost real time.

Likewise, the use of nanopore technologies is also revolutionising the field of studying infectious diseases and microbiology with technological development facilitating the analysis of sequencing data simultaneously to the sequencing process - making results available instantaneity via web-based applications (Figure 1). Alongside being rapid, these methods are portable, easy to use and have been successfully applied to study other pathogens such as Ebola in rural outbreak situations (Quick *et al.*, 2016). Clearly, such methods may prove invaluable in gathering data during rapidly unfolding situations such as cholera outbreaks, where genomic information could be used to detect, characterise and

inform risk-based interventions. However, the ability to analyse outbreaks in real-time is often hampered by the rapid availability of epidemiological data coupled to a central data depository for analysis and data visualisation purposes.

New approaches developed for other microbial risks perhaps show us the way forward. The online interactive dashboard, hosted by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University to visualise and track reported cases of the ongoing coronavirus disease pandemic (COVID-19) in real time (Dong, Du and Gardner, 2020) are an excellent example of how this can be achieved. From our experience it appears that a key data gap and hurdle that needs to overcome is in the rapid capturing and sharing of data between researchers interested in vibrios, in particular transnational epidemiological datasets which are critical to study these bacteria when outbreaks do occur. Until recently, silos of different datasets have been analysed in separated contexts with specific tools for each – e.g. genomics, remote sensing, ecology, climatology, oceanography, environmental microbiology etc.

New developments in computational science such as artificial intelligence are already making possible the analysis of massive amounts of complex information. The combination of these repositories of data under a "big data" framework will provide new insights on the evolution and environmental drivers of disease, and will also allow the development of a new generation of tools integrating multifaceted and composite data viewing approaches to allow a more comprehensive picture of Vibrios and their surrounding context. Whilst these new and shiny techniques may be exciting they do not replace the need for some of the more mundane but often overlooked bench science used to study these bacteria – e.g. the ability to isolate strains from complex matrices, basic genetic manipulation tools and the use of microscopy to study physiology of these bacteria, among many others. Discoveries that are currently

emerging – such as from the application of genomewide analysis – will generate new conceptual frameworks that will need to be substantiated further by experimental work, and the Vibrio community will need to come back to the laboratory to test the hypotheses formulated from these genomic datasets. Many of the exciting methods and tools to study Vibrios are available here, and now, and they are already fundamentally changing our collective understanding in these bacteria. The successful integration of various disciplines including microbiological, genomic, epidemiological, climatic and ocean sciences is paramount and cannot be understated.

References

Acharya, K. *et al.* (2019) 'A comparative assessment of conventional and molecular methods, including MinION nanopore sequencing, for surveying water quality', *Scientific Reports*. Springer US, 9(1), pp. 1–11. doi: 10.1038/s41598-019-51997-x.
Allué-Guardia, A. *et al.* (2018) 'Closed genome sequence of Vibrio cholerae O1 El Tor Inaba strain A1552', *Genome Announcements*, 6(9), pp. 4–6. doi: 10.1128/genomeA.00098-18.
Baker-Austin, C. *et al.* (2013) 'Emerging Vibrio risk at high latitudes in response to ocean warming', *Nature Climate Change*, 3(1). doi: 10.1038/nclimate1628.
Baker-Austin, C. *et al.* (2017) 'Non-Cholera vibrios: the microbial barometer of climate change', *Trends in Microbiology*. Elsevier, 25(1), pp. 76–84. doi: 10.1016/j.tim.2016.09.008.
Chen, C. Y. C.-Y. C.-Y. *et al.* (2003) 'Comparative Genome Analysis of Vibrio vulnificus,\ra Marine Pathogen', *Genome Research*, 13(12), pp. 2577–2587. doi: 10.1101/gr.1295503.7.
Colwell, R. R. (1996) 'Global climate and infectious disease: the cholera paradigm.', *Science (New York, N.Y.)*, 274(5295), pp. 2025–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8953025 (Accessed: 31 October 2017).
Cordero, O. X. *et al.* (2012) 'Antibiotic Production and Resistance', *Science (New York, N.Y.)*, 337(September), pp. 1228–1231. doi: 10.1126/science.1219385.
Domman, D. *et al.* (2017) 'Integrated view of Vibrio cholerae in the Americas.', *Science (New York, N.Y.)*. United States, 358(6364), pp. 789–793. doi: 10.1126/science.aao2136.
Dong, E., Du, H. and Gardner, L. (2020) 'An interactive web-based dashboard to track COVID-19 in real time.', *The Lancet. Infectious diseases*. Elsevier Ltd, 3099(20), pp. 19–20. doi: 10.1016/S1473-3099(20)30120-1.
Grimes, D. J. *et al.* (2014) 'Viewing marine bacteria, their activity and response to environmental drivers from orbit: satellite remote sensing of bacteria.', *Microbial ecology*, 67(3). doi: 10.1007/s00248-013-0363-4.
Hehemann, J. H. *et al.* (2016) 'Adaptive radiation by waves of gene transfer leads to fine-scale resource partitioning in marine microbes', *Nature Communications*, 7. doi: 10.1038/ncomms12860.
Heidelberg, J. F. *et al.* (2000) 'DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae', 406(August), pp. 477–484.
Hunt, D. E. *et al.* (2008) 'Resource partitioning and sympatric differentiation among closely related bacterioplankton.', *Science (New York, N.Y.)*. United States, 320(5879), pp. 1081–1085. doi: 10.1126/science.1157890.
Livny, J. *et al.* (2014) 'Comparative RNA-Seq based dissection of the regulatory networks and environmental stimuli underlying Vibrio parahaemolyticus gene expression during infection', *Nucleic Acids Research*, 42(19), pp. 12212–12223. doi: 10.1093/nar/gku891.

Lobitz, B. *et al.* (2000) 'Climate and infectious disease: use of remote sensing for detection of Vibrio cholerae by indirect measurement.', *Proceedings of the National Academy of Sciences of the United States of America*, 97(4), pp. 1438–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10677480 (Accessed: 31 October 2017).
Makino, K. *et al.* (2003) 'Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism distinct from that of V. cholerae', *The Lancet*, 361(9359), pp. 743–749. doi: 10.1016/S0140-6736(03)12659-1.
Martinez-Urtaza, J. *et al.* (2008) 'Emergence of Asiatic Vibrio diseases in South America in phase with El Niño.', *Epidemiology*, 19(6), pp. 829–837. doi: 0.1097/EDE.0b013e3181883d43.
Martinez-Urtaza, J. *et al.* (2010) 'Climate anomalies and the increasing risk of Vibrio parahaemolyticus and Vibrio vulnificus illnesses', *Food Research International*. Elsevier Ltd, 43(7), pp. 1780–1790. doi: 10.1016/j.foodres.2010.04.001.
Martinez-Urtaza, J. *et al.* (2017) 'Genomic Variation and Evolution of Vibrio parahaemolyticus ST36 over the', *American Society for Microbiology*, 8(6), pp. 1–17. doi: 10.1128/mBio.01425-17.
Mutreja, A. *et al.* (2011) 'Evidence for several waves of global transmission in the seventh cholera pandemic', *Nature*. Nature Publishing Group, 477(7365), pp. 462–465. doi: 10.1038/nature10392.
Orata, F. D., Keim, P. S. and Boucher, Y. (2014) 'The 2010 Cholera outbreak in Haiti: how science solved a controversy', *PLOS Pathogens*. Public Library of Science, 10(4), p. e1003967. doi: 10.1371/JOURNAL.PPAT.1003967.
Quick, J. *et al.* (2016) 'Real-time, portable genome sequencing for Ebola surveillance', *Nature*. Nature Publishing Group, 530(7589), pp. 228–232. doi: 10.1038/nature16996.
Semenza, Jan C *et al.* (2017) 'Environmental Suitability of Vibrio Infections in a Warming Climate: An Early Warning System.', *Environmental health perspectives*. United States, 125(10), p. 107004. doi: 10.1289/EHP2198.
Semenza, Jan C. *et al.* (2017) 'Environmental suitability of vibrio infections in a warming climate: An early warning system', *Environmental Health Perspectives*, 125(10), pp. 1–12. doi: 10.1289/EHP2198.
Vezzulli, L. *et al.* (2016) 'Climate influence on *Vibrio* and associated human diseases during the past half-century in the coastal North Atlantic', *Proceedings of the National Academy of Sciences*, 113(34), pp. E5062–E5071. doi: 10.1073/pnas.1609157113.
Williams, T. C. *et al.* (2014) 'Transcriptome sequencing reveals the virulence and environmental genetic programs of vibrio vulnificus exposed to host and estuarine conditions', *PLoS ONE*, 9(12). doi: 10.1371/journal.pone.0114376.

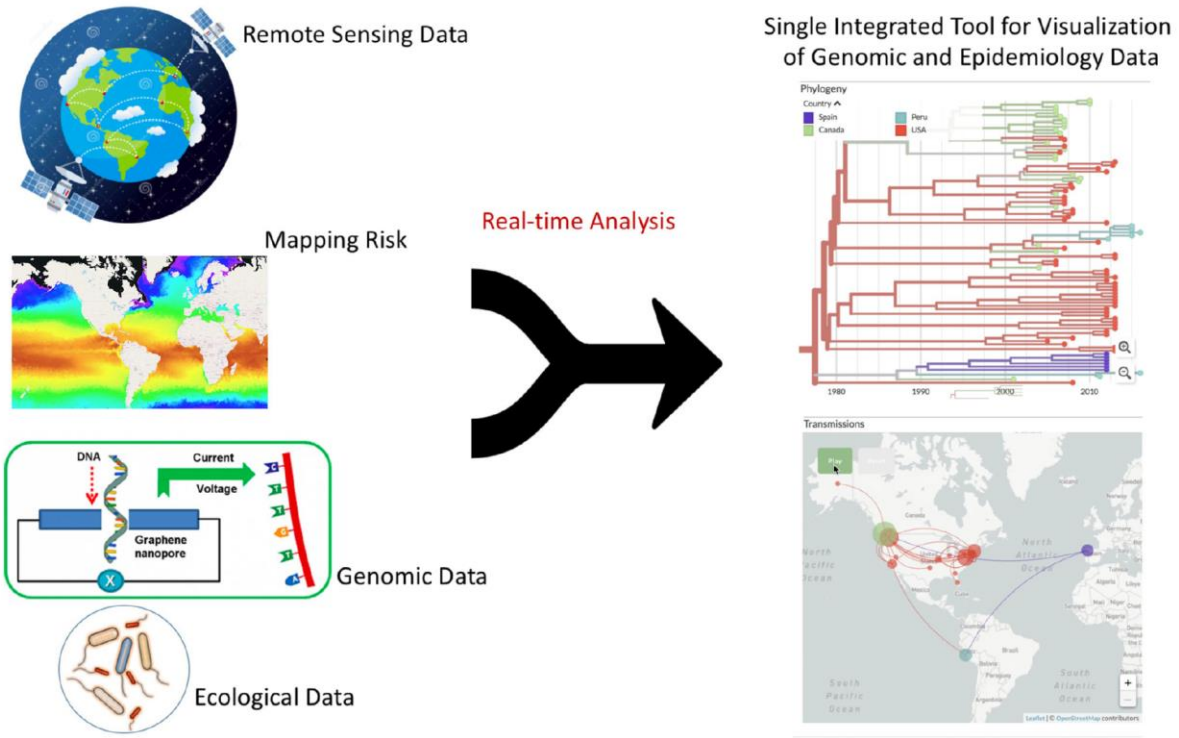

Figure 1. The integration of disciplines such as remote sensing, risk mapping, whole genome sequencing, microbial ecology and enhanced data visualisation techniques has provided a far more cohesive understanding of processes such as the pandemic spread of pathogenic vibrios.