

[ScienceWatch Home](#)[Inside This Month...](#)[Interviews](#)[Featured Interviews](#)[Author Commentaries](#)[Institutional Interviews](#)[Journal Interviews](#)[Podcasts](#)[Analyses](#)[Featured Analyses](#)[What's Hot In...](#)[Special Topics](#)[Data & Rankings](#)[Sci-Bytes](#)[Fast Breaking Papers](#)[New Hot Papers](#)[Emerging Research Fronts](#)[Fast Moving Fronts](#)[Research Front Maps](#)[Current Classics](#)[Top Topics](#)[Rising Stars](#)[New Entrants](#)[Country Profiles](#)[About Science Watch](#)[Methodology](#)[Archives](#)[Contact Us](#)[RSS Feeds](#)

# scienceWATCH<sup>®</sup>.com

TRACKING TRENDS & PERFORMANCE IN BASIC RESEARCH

[Interviews](#)[Analyses](#)[Data & Rankings](#)

2009 : January 2009 - Author Commentaries : Stephen Porcella - Interview

## AUTHOR COMMENTARIES - 2009

January 2009



### Stephen Porcella

Featured Scientist from *Essential Science Indicators*<sup>SM</sup>

According to a recent analysis of Essential Science Indicators data from Thomson Reuters, the work of Dr. Steve Porcella has entered the top 1% in the field of Microbiology. Dr. Porcella's current record in this field includes 15 papers cited a total of 648 times between January 1, 1998 and August 31, 2008. His overall record for this period in the database includes 37 papers cited a total of 1,167 times.

Dr. Porcella is the Chief of the Research Technologies Section at Rocky Mountain Labs, which is part of the National Institute of Allergy and Infectious Diseases (NIAID), and oversees the RTS Genomics Unit within that section.

**In the interview below, he talks with ScienceWatch.com about his highly cited work.**

### **SW:** Would you tell us a bit about your educational background and research experiences?

I have a Ph.D. in Microbiology from the Division of Biological Sciences, working with Dr. Ralph Judd at the University of Montana. I was fortunate to perform some of my graduate research at Rocky Mountain Labs (RML), NIAID, NIH, where I worked on *Neisseria gonorrhoea* and worked with a "new at the time" slab acrylamide 370S DNA sequencer. A post-doctoral appointment in the Department of Microbiology at University of Texas Southwestern Medical School in Dallas led to my work on *Treponema pallidum* and *Borrelia burgdorferi* and management of a 377 DNA sequencer. A second post-doctoral appointment under Dr. John Swanson at RML allowed me to collaborate and work extensively with Dr. Tom Schwan on *Borrelia hermsii*, *Borrelia burgdorferi*, and *Borrelia turicatae*. All of my postdoctoral work involved genomics approaches, protein work, bioinformatics, and nucleic acid work.

Also at RML, but while under Dr. James Musser, now at the Methodist Hospital Research Institute in Houston, I worked on *Streptococcus pyogenes*. Dr. Musser tasked me with setting up an ambitious Genomic sequencing, Microarray (spotted), and q-PCR, and technology group at RML. After Dr. Musser's departure, this technology group officially became the RML Research Technologies Section (RTS), part of the intramural NIAID Research Technologies Branch under Branch Chief Dr. Robert Hohman. Today I am the Section Chief of the RTS, which now comprises three technology Units—Genomics, Electron Microscopy, and Flow Cytometry. We have over 20 individuals in the RTS, ranging from highly skilled and trained technical staff members to student researchers.

I spend much of my time working closely with the Genomics Unit, where the molecular challenges are enormous, the workload always growing, and the range of projects very diverse. The RTS Genomics

Unit currently provides technology support for high throughput (HT), high-density microarrays (all custom and commercial Affymetrix products), HT-DNA sequencing (capillary, 454, SOLiD), HT-Robotics, HT-Q-PCR, HT phenotype microarray, HT patient cohort and pathogen genotyping, and bioinformatics for all of the above.

Our goal within the RTS is to provide advanced technology support, from experimental design to publishable figures to all Intramural investigators within the NIAID. These scientists are all actively pursuing the NIAID mission of performing basic and applied research contributing to the development of diagnostic reagents, therapeutics, and vaccines against human pathogens and human disease.

*"When the first fluorescent DNA sequencer became available, no one could have guessed how much progress would be made from those early days."*

**SW: What would you say is the main focus of your research, and what drew you to this area?**

My main focus is genomics, both at the host and pathogen level. Many of the investigators that we support are specifically interested in either the host response to infection or the pathogen response during the infectious process, and sometimes both at the same time. We also support those working on genetic or immunological disorders, the latter of which can lead to life-threatening infections. Last, but not least, we also work with labs studying vectors such as fleas, ticks, mosquitoes, and their roles in the transmission of disease. Our support of pathogen work also focuses on analysis of newly discovered pathogen isolates, host genetics, host susceptibility to disease, and pathogen genetics.

We work with many different scientists, labs, models, pathogens, and patient cohorts, across the NIAID Intramural program. Through this diversity, we have developed novel protocols and technologies specific or compatible across many different models/systems. We routinely perform DNA sequencing (single clone to genome level), RNA/DNA isolation, high throughput microarray and the attendant data analysis, high throughput Q-PCR for sample quantitation or microarray dataset validation, human or pathogen genotyping, and complex genomic or statistical microarray bioinformatics.

Several keys to these efforts are as follows:

- 1) Performing pathogen genome sequencing to high quality and deep annotation, while providing a myriad of comparative analytical efforts, all with the goal of efficiently communicating the complexity of the data to the investigators.
- 2) Rapidly incorporating genome-level data into new custom expression arrays where we can quickly perform Comparative Genomic Hybridization (CGH), or determine differential expression profiles under different conditions, over time, or under different treatments. At this time, we have developed 8 custom Affymetrix arrays containing more than 60 different pathogens.
- 3) Understanding the fundamentals of sound, randomized, multiple replicate experimental designs in order to provide the best data possible, removed from batch effects, or other artifactual external influences, is critical. Typically, we prefer six biological replicates per time point or condition, as this gives us the best numbers for statistical significance in the end, but also allows for loss of one to three samples while still maintaining numbers appropriate for analysis. Paramount is that the experiment to be performed will provide the data for the question being asked. Equally important to all of this is communicating effectively with investigators so that both the wet-bench, RTS efforts, and bioinformatics are all using consistent information.
- 4) When accomplished, the processes described above lend themselves to rapid Q-PCR validation efforts where we often see high concordance (p-value statistical correlation) with the chip-based expression data. This latter effort enables researchers to move quickly to publication, or to pursue other experiments for confirmation (protein-based, for example).
- 5) We provide all the support needed to get data into the public domain and fast-track the acquisition of accession numbers for publication.
- 6) Generation of publication quality tables and figures. Because our RTS scientists work with the technology every day, from start to finish, for many years now, they understand the data and its interpretation better than anyone. Therefore, with principal investigator's guidance we are capable of producing any figure or table needed in order to efficiently tell the story in a publication format. This extra effort by my group, which we make as a priority, is the reason why so many collaborative papers are being produced.
- 7) Staying on top of the latest technology. This is a no-brainer, and as an example, we have recently

acquired a SOLiD sequencer and are aggressively pursuing this technology for many of the topics described above.

Finally, what drew my interest to this area? It is simple—early in my graduate career I saw that genomic technologies or approaches could give many answers to phenotypic observations/questions. This concept is especially important for research on intractable, hard-to-study pathogens or host-pathogen inter-relationships. Microarray research is basically an extension of the genomic record, as well as Q-PCR (another tool), genotyping, and phenotype arrays.

Bioinformatics are challenging, but our philosophy has always been, "focus on the downstream bioinformatics, while we are pursuing the implementation of a new technology, with the goal of preventing a bottleneck downstream." It is this approach I believe that has allowed us to rapidly and efficiently get new technologies up and running at high accuracy within reasonable timeframes and meet nearly all investigators needs.

Also, DNA and, to a lesser extent, RNA are fairly easy and stable to work with compared with proteins. Therefore, high throughput methodologies enable rapid and extensive scientific discovery under the right conditions across many different labs, models, and experimental questions. Finally, there is nothing like the big data rush, when you are the first to see those discoveries come and you communicate that to the investigators.

**SW:** Your most-cited original paper in our database is the 2002 *PNAS* article, "Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks." Would you walk our readers through this paper, its goals, findings, and significance?

This was a landmark paper because we in-house sequenced a complex bacterial genome at a time when only big sequencing centers were taking on these labor-intensive projects. At the time, only one other Group A *Streptococcus* (GAS) genome had been sequenced, that of an M1 isolate, genetically distinct from M1 strains commonly responsible for GAS infections. M18 strains had been associated for decades with Acute Rheumatic Fever (ARF) outbreaks in the US. As soon as we finished the M18 genome we made a spotted array of it and hybridized DNAs from 36 M18 strains cultivated from diverse localities. The results noted few genomic gene differences; however, phage and phage-like elements turned out to be the primary source of variation. This array data was further supported by a high throughput, comparative gene sequencing of 500 isolates collected during the same ARF outbreaks.

"My fundamental hope is that computing and bioinformatics keeps pace with the development of new, cutting-edge genomics technologies"

Therefore, horizontal gene transfer events were described as being important if not critical sources of genomic diversity among M18 strains. The data also showed that M18 strains recovered from two ARF outbreaks 12 years apart in Salt Lake City (SLC), were nearly genetically identical and therefore the increase in ARF cases in SLC in 1998-1999 was associated with a resurgence of an M18 clone common in 1987-1988 in the same area.

An interesting point: shortly after this paper was published we started a seven genome-sequencing project. To my knowledge, this was the most anyone anywhere had undertaken in terms of multiple, highly related isolates sequenced at the same time. Cross-contamination was a big fear during the sequencing phase of all isolates, so we were meticulous in protecting against it. A company, Integrated Genomics out of Chicago, was instrumental in accomplishing that goal. Many of the subsequent papers we published on *Streptococcus* were a direct result of that multi-genome sequencing effort.

**SW:** Judging from your papers in our database, you do quite a bit of work with group A *Streptococcus*. What is it about this organism that warrants such interest? What other organisms do you work on?

Currently, my group is not actively doing work on Group A *Streptococcus pyogenes* (GAS); that effort tapered off greatly when Dr. Musser took an appointment at Methodist Hospital. My group does do an enormous amount of work with many different bacterial pathogens described below. However, Group A Streptococcal infections are a huge public health problem in developed and underdeveloped countries around the world. A half million deaths occur worldwide each year due to GAS infections. Globally, greater than 600 million cases per year can be attributed to GAS infections. In the US alone, in 2000, acute pharyngitis was responsible for 11 million office visits. Invasive GAS infections can occur at a rate of approximately 3.5 per 100,000 with roughly 1,500 of those resulting in death. A vaccine remains elusive and more research and funding are needed to reduce the impact of this very important human

pathogen in the US and world.

Pathogen organisms my group has or is currently working on, in terms of providing technical support, are as follows (I won't list all the species or strains); *Borrelia*, *Coxiella*, *Chlamydia*, *Rickettsia*, *Francisella*, *Granulibacter*, *Salmonella*, *Burkholderia*, *Mycobacteria*, *Yersinia*, *Staphylococcus*, *Brucella*, Giardia, Malaria, Langet virus, prions, and a host of other viruses. We are also providing technical support on whole-tissue or flow-sorted cell analysis for human, mouse, rat, hamster, and guinea pig models and we are also spending a lot of work on genetically immuno-diseased individuals, particularly those individuals whose mutations are not known but which may be life-threatening.

**SW: You recently co-authored a paper in *Infection and Immunity*, "The *Chlamydia trachomatis* plasmid is a transcriptional regulator of chromosomal genes and a virulence factor." Would you tell our readers a little bit about this research?**

I frequently use this paper as an example of the type of project that my group regularly contributes to. Normally, *Chlamydia trachomatis* contains a 7.5kb cryptic plasmid of unknown function. Drs. Carlson and Caldwell had an isolate that lacked this plasmid. Growth kinetics, plaquing efficiency, and plaque size all showed no difference between plasmid-minus versus plasmid-containing strains. Failure to accumulate glycogen granules was the only major, observed phenotype of the plasmid-minus strain. Based upon some preliminary Q-PCR, it was known at what time during the infectious cycle glycogen synthase (enzyme involved in glycogen synthesis) was maximally expressed in the wild type relative to the mutant.

Therefore, a microarray experiment was designed to look at six biological replicates of each condition (Wt and MT), including tissue culture cells only (*Chlamydia* is an obligate intracellular parasite) at that time point. One key hurdle to obligate parasites and microarrays is that you are dealing with a mixed sample, usually with host RNA in great abundance relative to the pathogen. Accurate quantitation of the *Chlamydial* RNA, in a host RNA background, is also inherently difficult for the following reason: *Chlamydia* is a biphasic organism, involving an active replication phase and a dormant spore-like phase. No true constitutively expressed gene exists for Q-PCR quantitation methods. Therefore, we extracted RNA and DNA simultaneously (to use the DNA as normalization and genomic equivalents). We tested by Q-PCR 3 different "constitutive-like" genes appropriate for the replicating phase, approximated the amount of pathogen RNA that was there, and ran the chips. We like to use PCA plots of the chip data to demonstrate that replicates group more closely with each other rather than across conditions and that was shown in the paper.

Venn diagrams, quality filters, and statistical analysis showed 29 genes passed all criteria for significant differential expression. Twenty-two of these genes were coded for on the chromosome including glycogen synthase, while the others were on the plasmid. ID-50 experiments in mice confirmed that the plasmid appears to play a significant role, perhaps more so *in vivo*, with regards to virulence and gene expression control. This is a very important study that opens up a wide range of research opportunities on this important human pathogen and it opens other avenues of exploration in other organisms where cryptic plasmids are known to exist.

**SW: How has the scope of our knowledge with regard to bacterial genomes, genomics and new technologies changed over the past decade?**

When the first fluorescent DNA sequencer became available, no one could have guessed how much progress would be made from those early days. The growth and development of genomics technologies has almost been logarithmic in its pace. There have been pitfalls, and manufacturer mistakes, but by and large the strides made have been far-reaching and with great demonstrated impact. A question some ask is, "Are we better off with these technologies in-house and operating 24/7 in support of biomedical research?" Cost, resources, and complexities are valid arguments against them. My group and I believe 100% that the benefit greatly outweighs the drawbacks. The discoveries and efforts associated with these technologies would not have been possible with older methodologies or techniques. My group strongly believes that the best is still to come. With new technologies coming and computing infrastructure catching up, future endeavors look very promising.

A saying we use here is that there is a genomics technology revolution taking place. Never before has such technology been available to answer so many important and wide-ranging questions about pathogenesis and human disease. When you look at the last 200 years of research and compare it to the last 10 years, the recent accomplishments are mind-boggling. We must never lose sight of this recent impact in biomedical research. My group puts a lot of effort into staying on that cutting edge of the revolution. We are taking advantage of the technologies available and those soon to be released. We are aggressively pursuing their full implementation towards the discovery of new diagnostics,

therapeutics, and vaccines in human disease.

### **SW: What are your hopes for this field for the future?**

My fundamental hope is that computing and bioinformatics keeps pace with the development of new, cutting-edge genomics technologies. Another very important factor is the chemistry and high throughput aspects of these technologies. Protocols or techniques can be modified or improved to give better yields, higher-quality data, more reproducible results, higher throughput (96 and 384 or 1536 formats) and hopefully all at lower costs. Chemistry improvements must go hand-in-hand with technology developments and an effort always needs to be made to reduce costs. We are very interested in single molecule sequencing, as current next-generation technologies require substantial input amounts of material. We are also interested in higher density arrays, in smaller formats, again to reduce sample amount, which for much of our work is a constant issue. Better and deeper annotated genomes or genome updates is something that will constantly need to be improved.

Therefore, with regard to arrays and other fixed platform technologies is it imperative that new, streamlined bioinformatics processing for updating that information be available. As you can see, we are interested in high throughput aspects of technologies, all of which can generate enormous amounts of data. Data processing times slow us down in terms of getting that data to the investigators. We are always looking for ways to automate our analysis to maintain a high throughput pipeline.

A key, but overlooked, aspect of this is databases. Quick, inexpensive, easy-to-use databases for sample tracking, data tracking, inventories, data hand-offs, and storage of raw and processed data are greatly needed. Especially attractive are those databases that include some level of automated data analysis, whether it is quality analysis, or even volume tracking. We are always looking for biomedical off-the-shelf databases, but find few opportunities available. We have the experience and capability to make databases in-house that are specific to our needs, but we would rather spend our time instead on data generation, data analysis, and data hand-off to investigators.

Whole-genome, high-density human genotyping platforms with the ability to provide data on SNPs, copy number, methylation, Indels, or splice variants are of great interest to my group. We are always looking for high-density, high-coverage platforms that can accomplish mutation detection.

One last important comment: by working together great things can be accomplished. Our Genomics Unit here has a very team-oriented philosophy. Because of our outstanding RTS staff members and their team-oriented work ethic I believe many good things are still to come in our efforts in biomedical research and scientific discovery. ■

**Stephen F. Porcella, Ph.D.**

**Section Chief**

**Research Technologies Section**

**Genomics, Flow Cytometry, Electron Microscopy Units**

**Research Technologies Branch**

**RML**

**NIAID, NIH**

**Hamilton, MT, USA**

#### **Stephen Porcella's current most-cited paper in *Essential Science Indicators*, with 199 cites:**

Smoot JC, *et al.*, "Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks," *Proc. Nat. Acad. Sci. USA* 99 (7): 4668-73, 1 April 2002. Source: *Essential Science Indicators* from Thomson Reuters.

Keywords: genomics, host, pathogen, host response to infection, life-threatening infections, vectors, microarray analysis, genome sequencing, Group A Streptococcus, pathogenic organisms, Chlamydia, virulence factor, genomics technologies.



PDF

[back to top](#)

2009 : January 2009 - Author Commentaries : Stephen Porcella - Interview

