genomes is necessary for a complete understanding of pathogen biology, as well as to understand better the diversity of bacterial species.

Lessons Learned While Exploiting the Pan-Genome of GBS

The determination of a species pan-genome also allows us to infer important practical information for vaccine design. Although the core genes represent the most desirable source for the selection of conserved and therefore potentially universally applicable vaccine candidates, they are also more likely to be immunologically silent in any successful pathogen. The group of dispensable genes, by contrast, might be an invaluable source of novel antigens that, although only present in a subgroup of strains, might encode important virulence-associated functions and might be exploited in appropriate combinations to elicit a broad immune response.

In particular, the pan-genomic concept in GBS was exploited for the discovery of an effective vaccine for cross-strain reactivity. A multigenome reverse vaccinology approach was undertaken (44), in which the core and dispensable genomes of the GBS pan-genome, resulting from analysis of the eight representative genome sequences, were mined for vaccine candidates. Among the predicted surface-exposed proteins, 396 were core genes and 193 were variable genes. The subsequent ability of the in vitro expressed candidates to confer protection in in vivo screening against a large panel of isolates resulted in the identification of a candidate universal GBS vaccine. This consists of a combination vaccine of four proteins, which is able to cover a wide range of strains. The important novelty of this study is that none of these antigens could be classified as universal, because only one of them is from the core genome (showing negligible surface accessibility in some strains), and the other three are encoded by dispensable genes and were therefore absent in a fraction of the tested strains.

Therefore, the analysis of multiple genomes of GBS revealed tremendous diversity and identified candidates that are not shared by all the strains sequenced but provide general protection when combined. Intriguingly, it was found that all three dispensable proteins are components of pilus-like structures with an important role in the virulence in GBS (59). These structures, over four times the length of the bacterium itself, had remained completely elusive to researchers after decades of work in the field.

Therefore, up-front genome comparisons from strains representative of the genetic diversity is a powerful tool for the selection of broadly protective combination of proteins and may be instrumental in formulation in universal vaccines against pathogens with highly diverse circulating strains. The natural next step to achieve a more comprehensive and epidemiologically related picture of bacterial populations will be population vaccinology, leading to the formulation of vaccines from a collection of proteins that, together, protect against the major circulating populations of a pathogen (60). To accomplish this, a more comprehensive and in-depth understanding of population structure is required. For many years it has become increasingly apparent that biology is no longer a linear science, but with the recent explosion in information due to metagenomics, we are beginning to understand we are at the tip of the iceberg. Understanding the variation that exists within populations seems miniscule when we think of the vast variation that is emerging through metagenomic sampling of global microbial populations. Over six million new proteins were predicted as the fruits of a single metagenomics project sampling global sea water (61), nearly doubling the total number of proteins known to date.

FUNCTIONAL GENOMICS

''Postgenomic'' methods (e.g., genomic microarray-based methods and proteomics) have changed the way investigators approach the classical questions: scientists now address questions on a whole-cell or system-wide basis, in contrast to the classical reductionist approaches. Genomics empowers the use of highly parallel methodologies that allow investigators to study all the genes or all the proteins of a pathogen in the context of a host or under various physiological or genetic states of interest (28). Functional genomics approaches are complementary to in silico antigen discovery. These include the large-scale analysis of gene transcription by DNA microarrays, the identification of the whole set of proteins encoded by an organism (proteomics) by two-dimensional gel electrophoresis and mass spectrometry, as well as using these protein reagents to create protein chip technologies to monitor immunological responses in human sera. Furthermore, the high-throughput capacity of these techniques facilitates the quantification of expressed genes in a comprehensive genome-wide framework.

Before genomics, in vivo expression technologies (IVETs) (62) and signature-tagged mutagenesis (STM) (63) used promoter-trapping methods or inhibition of gene function to analyze genes highly expressed in vivo or important for infection. Needless to say, both technologies have greatly benefited from the availability of genome sequences, although the previous knowledge of the genome sequence is not strictly necessary for their application. By combining whole-genome microarrays and comprehensive ordered libraries of mutants, high-throughput functional screens can now be achieved on a genomic scale (64).

Global Genomic Profiling Using Microarrays

Global genomic profiling of gene expression using ordered DNA or oligonucleotide microarrays is a very powerful technology, which can be exploited in many different ways to further the study of genes that are involved in microbial pathogenesis. For vaccine discovery programs, it is of key importance to know what genes are expressed during host infection and what proteins can elicit an antibody response in humans. Microarray-based expression studies provide a strong contribution to the understanding of how a pathogen orchestrates responses to the host environment. DNA microarrays can be used to obtain a global profile of genes of a pathogenic microorganism whose expression is upregulated during infection of animals or of in vitro models. The major challenges in performing expression analysis in vivo or in vitro include the efficient recovery of RNA and the choice of an appropriate model system and/or experimental systems. The transcriptional changes in *N. meningitidis* were investigated from meningococci incubated in human serum as well as adherent to human epithelial and endothelial cells (65). The authors of this study discovered a wide range of surface proteins that are induced under in vivo conditions and that could represent novel candidates for a protein-based vaccine for meningococcal disease.