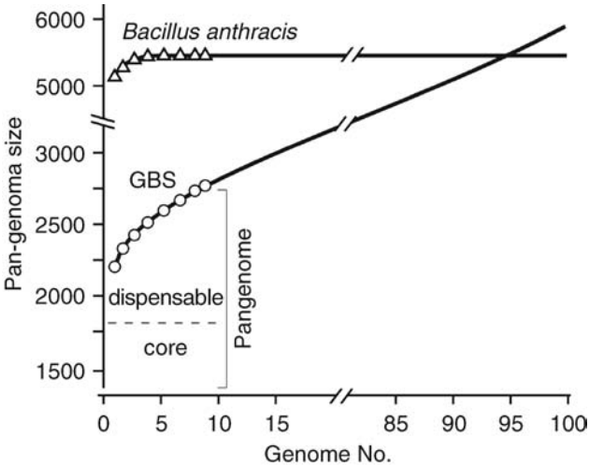

Figure 2 The pan-genome is the global gene repertoire of a bacterial species and depends on analysis of the number of available genome sequences. The size of the pan-genomes of *Streptococcus agalactiae* (GBS) and *Bacillus anthracis* are shown as a function of the number of sequenced strains. The curves represent a mathematical extrapolation of the data to a large number of strains. The pan-genome is made up of the core genome (the pool of genes shared by all the strains of the same bacterial species) and the dispensable genome (the pool of genes present in some, but not all, strains of the same bacterial species). Analysis of eight GBS genome sequences indicates that the pan-genome contains 2713 genes, of which 1806 belong to the core genome and 907 to the dispensable genome (56). Either the size of a species pan-genome can grow with the number of sequenced strains and is nominated an ''open'' pan-genome (the GBS pan-genome), or the size can quickly saturate to a limiting value and is nominated a ''closed'' pan-genome (the *B. anthracis* pan-genome). GBS pan-genome is predicted to grow by an average of 33 new genes every time a new strain is sequenced (56). The final number of genes in the pan-genome of a species with an open pan-genome may be several orders of magnitude larger than the number of genes in an individual genome. *Abbreviation*: GBS, group B *Streptococcus*.

universal protein-based vaccines can be developed against encapsulated bacteria that are usually targeted by conjugate vaccines. In addition, one further advantage of reverse vaccinology is that all the antigens are produced as soluble recombinant proteins in *E. coli*, thus supporting the straightforward development of a suitable manufacturing process for large-scale production.

After the initial success of reverse vaccinology in solving the problem of antigen selection for MenB, reverse vaccinology has been applied to the search for antigens and virulence factors in many other pathogenic bacteria including *Bacillus anthracis* (40,41), *Porphyromonas gingivalis* (42), *Streptococcus pneumoniae* (43), *Streptococcus agalaticae* (44), *Chlamydia pneumoniae* (45), and *Brucella melitensis* (46). Thus, the reverse vaccinology approach appears to be applicable to a range of pathogens and, in principle, also to eukaryotic parasites, for which genome sequence and suitable in vitro or in vivo models are available.

Epitope Predictions and Immunomics

For a vaccine to be effective, it must invoke a strong response from both T cells and B cells; therefore, epitope mapping is a central issue in vaccine design (47,48). In the past, scientists isolated proteins from whole cells and then digested the protein antigens to find smaller fragments or epitopes that stimulated the T-cell and B-cell response. The set of pathogen epitopes that interface with the host immune system is now known as the ''immunome.'' A key focus of immunomics has been the development of algorithms for the design and discovery of new vaccines. T-cell epitope-mapping algorithms are based on straightforward mathematical analyses of the patterns of amino acids that occur in peptides bound to (and presented in the context of) human leukocyte antigen (HLA) by antigen-presenting cells (47). Because the epitope peptide is bound in a linear form to HLA, the interface between ligand and T cell can be modeled with breathtaking accuracy. In contrast, B-cell epitope-mapping algorithms have lagged behind T-cell epitope-mapping algorithms, and few B-cell epitope-mapping algorithms are in current use.

In a whole-genome approach, similar to reverse vaccinology, a pathogen's entire proteome in silico can be analyzed using T-cell epitope-mapping tools and further in vitro evaluation to discover new protein candidates for vaccines. The concept has been described as fishing for antigens using epitopes as bait. A number of recent papers point a clear path from the genome to the immunome, bringing us closer to understanding just what information about a pathogen is required for effective host immune defense (48–50). Large-scale screening of pathogens for T- and B-cell epitopes from the National Institute of Allergy and Infectious Diseases (NIAID) categories A to C pathogens and other pathogens of global importance is also currently under way (51). This study includes a database of the complete systematic genome mapping of pathogens such as *B. anthracis*, *M. tuberculosis*, *Clostridium tetani*, *Francisella tularensis*, *Y. pestis*, and viruses including smallpox, flaviviruses, arenaviruses, rabies, influenza, and hepatitis A. Immunomics results in the expansion of a number of different proteins that can be screened for vaccine development while narrowing this search to regions of the proteins that are likely to induce an immune response (52).

Researchers are now implementing these combined methods to scan genomic sequences for vaccine components. There are however some limitations to the ''genome-to-vaccine'' approach. One obvious point of contention is the choice of the genome/genomes to mine. While often the choice of which genome to mine was essentially the first genome that was sequenced, this may not always represent an adequate representative organism. Today multiple genomes from each species are increasingly becoming available, opening the era of the pan-genomic reverse vaccinology, which will be dealt with further below. Another limitation of genome-derived vaccines is the inability to identify nonprotein antigens including polysaccharides or CD1-restricted antigens such as glycolipids, or the inability to identify posttranslational modifications on the selected protein antigens. Furthermore, there is a lack of algorithms that can be used to make a good correlation between antigens and their likely efficacy in protection especially for humoral immune responses (neutralizing antibodies), although such tools are now being developed. In addition, development of effective vaccines through genome mining is dependent on the available methods to measure in vitro efficacy and correlates of protection and not simply on identifying epitopes that