

GENOME WATCH

Overtake in reverse gear

Georgios S. Vernikos

This month's Genome Watch discusses the applications of whole-genome comparative genomics and reverse vaccinology for two life-threatening pathogens

The extent of intra-species diversity in bacterial populations was underlined recently by a study that focused on whole-genome-sequence comparisons of eight *Streptococcus agalactiae* isolates¹: the results revealed that the pan-genome² — the genome of a whole bacterial species that consists of core genes and dispensable genes that are partly shared — might be much larger than the genome of a single isolate.

Whole-genome comparative genomics has the potential to shed light on mechanisms of genome evolution, including horizontal gene transfer, recombination, gene duplication and gene loss, that shape the structure of microbial populations. Insights gained from analyses of core and dispensable genes can be used to develop better treatments and more efficient and broadly applicable vaccines³. Moreover, comparative genomics makes it possible for us to map the times of gene-transfer events in the evolutionary history of bacteria onto the evolutionary history of their hosts, which might reveal how the interactions between key gene-transfer events in the evolution of pathogens⁴ and behavioural or demographic changes in their host populations⁵ lead to the emergence of novel pathogens.

One of the most successful methods for providing protection against pathogens is

vaccination, a treatment that originally relied on processing biological agents to balance reduced virulence against immunogenicity. In conventional vaccine development, pathogenic strains are grown by sequential passages *in vitro* to develop live attenuated (or killed) strains that are harmless to the host but retain the ability to trigger a protective immune response. Alternative approaches have involved using antigens as a basis for subunit vaccines. Although promising, conventional vaccine approaches are not applicable to pathogens that cannot be grown *in vitro* (for example, hepatitis B and C viruses) or to pathogens in which immunodominant cellular components resemble components of human tissues (for example, the serogroup B meningococcus; MenB). Moreover, conventional vaccine approaches are time consuming (5–15 years) and can only identify and exploit antigens that are highly expressed and immunogenic during disease.

Recently, a different method that uses a bottom-up (rather than a top-down), genomic (instead of cellular) approach has been successfully applied to the development of vaccines against pathogens that were previously recalcitrant to such development. The only requirement for this new process is the genome sequence (or sequences) of the target pathogen. Such genome sequences are used as the input material for *in silico* algorithms that make predictions about putative antigens that are likely to be successful vaccine candidates. Because this approach uses the genome sequence rather

than the cell as the starting material, it has been named reverse vaccinology^{6,7}. Reverse vaccinology is fast (1–2 years, depending on the availability of high-throughput screening systems); can identify virtually all potential antigens, irrespective of their concentration, time of expression and immunogenicity; and can be used against all pathogens, including those that cannot be grown *in vitro*. However, this methodology cannot currently be used to develop vaccines that are based on non-protein-coding antigens, such as lipopolysaccharides (TABLE 1).

Reverse vaccinology has recently been successfully applied to the development of universal vaccines against group B *Streptococcus*⁸ (GBS) and vaccine candidates against MenB⁹. Reverse vaccinology, which is now a routine approach, has also been applied to other life-threatening pathogens, including staphylococci and streptococci⁷. There are two versions of reverse vaccinology: classical reverse vaccinology, which uses only a single genome sequence⁹ to make predictions about putative vaccine candidates, and comparative (or pan-genomic) reverse vaccinology, which compares the genomes of several closely and distantly related strains⁸. Comparative reverse vaccinology is generally used to develop universal, rather than strain- or serovar-specific vaccines. The biggest advantage of reverse vaccinology is speed, which is probably of vital importance for tackling rapidly emerging, life-threatening diseases; for example, after the

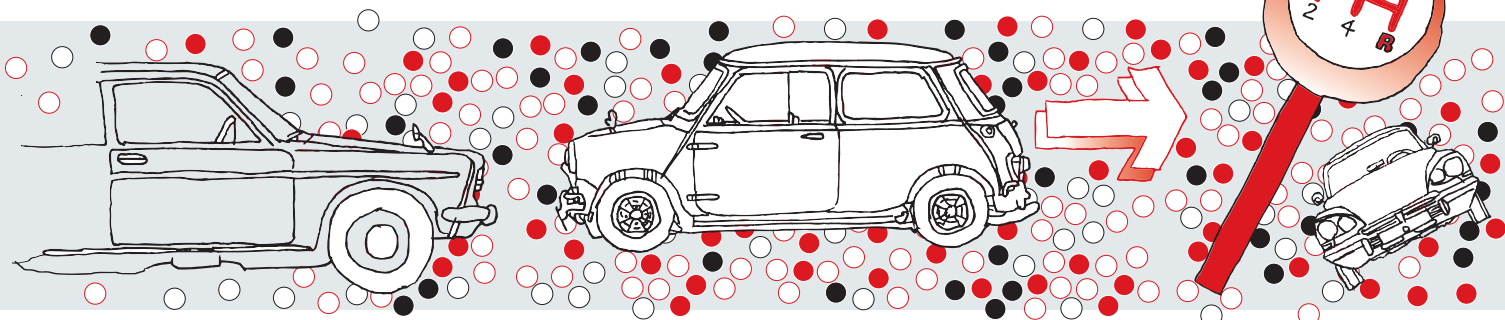


Table 1 | Comparison of conventional and reverse vaccinology*

Conventional vaccinology	Reverse vaccinology
Uses the most abundant antigens during disease	Analyses all antigens produced, whether abundant or not
Organism must be cultivable	Organism does not need to be cultivable
Animal models required	Animal models required
High-throughput expression and analysis are not required	High-throughput expression and analysis are important
Antigen identification is usually time consuming	Antigen identification is rapid
Only structural proteins are considered	Non-structural proteins are analysed
Polysaccharides and lipopolysaccharides can be used as antigens	Non-protein antigens, such as polysaccharides and lipopolysaccharides, cannot be identified

*Table modified, with permission, from REF. 6 © (2000) Elsevier Science.

several human tissues; and second, major protein-based antigens show high sequence variability and offer protection against only the immunizing (or homologous) strains. Pizzo *et al.*⁹ were the first to successfully implement the concept of reverse vaccinology by using the complete genome sequence of a virulent MenB strain as input to prediction algorithms for the identification of putative vaccine candidates. In less than 2 years, the authors had predicted 600 putative surface-exposed proteins, more than half of which were cloned and expressed in *E. coli* and purified for use in the immunization of mice. A quarter of these proteins were novel antigens that were exposed on the surface of the MenB cell, and almost one-third (25 proteins) induced a bactericidal antibody response. The novelty of this approach was that the identified antigens were well conserved at the sequence level, and so were ideal for the development of vaccines that offer protection against a wide range of homologous or even heterologous MenB strains.

Reverse vaccinology is a promising method for the high-throughput discovery of putative population-wide, rather than strain- or serotype-specific, vaccine candidates that have the potential to mirror the variability, dynamics and diversity of entire microbial populations.

The next big challenge is to devise the most efficient algorithmic pipelines to standardize this approach and make it high-throughput.

Georgios S. Vernikos is at the Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.
e-mail: microbes@sanger.ac.uk

doi:10.1038/nrmicro1898

Published online 8 April 2008

1. Tettelin, H. *et al.* Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc. Natl Acad. Sci. USA* **102**, 13950–13955 (2005).
2. Medini, D., Donati, C., Tettelin, H., Massignani, V. & Rappuoli, R. The microbial pan-genome. *Curr. Opin. Genet. Dev.* **15**, 589–594 (2005).
3. Demuth, A. *et al.* Pathogenomics: an updated European Research Agenda. *Infect. Genet. Evol.* 18 Jan 2008 (doi:10.1016/j.meegid.2008.01.005).
4. Parkhill, J. *et al.* Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**, 848–852 (2001).
5. Thomson, N. R. *et al.* *Chlamydia trachomatis*: genome sequence analysis of lymphogranuloma venereum isolates. *Genome Res.* **18**, 161–171 (2008).
6. Rappuoli, R. Reverse vaccinology. *Curr. Opin. Microbiol.* **3**, 445–450 (2000).
7. Rappuoli, R. & Covacci, A. Reverse vaccinology and genomics. *Science* **302**, 602 (2003).
8. Maione, D. *et al.* Identification of a universal Group B streptococcus vaccine by multiple genome screen. *Science* **309**, 148–150 (2005).
9. Pizzo, M. *et al.* Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* **287**, 1816–1820 (2000).

DATABASES

Entrez Genome Project: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj>
[Escherichia coli](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj) | [Neisseria meningitidis](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj) | [Streptococcus agalactiae](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

recent emergence of severe acute respiratory syndrome (SARS), the genome sequence of the responsible coronavirus was rapidly made available (less than a month after the first suggestion that this type of virus might be responsible for the disease), which enabled rapid identification of putative vaccine candidates⁷.

GBS is one of the most common causative agents of life-threatening infections, including meningitis, pneumonia and septicaemia in newborn babies, in the developed world. GBS is a commensal Gram-positive bacterium that is carried by up to one-third of healthy individuals and only occasionally causes disease. Vaccine development using reverse vaccinology against GBS is a promising defence against this neonatal killer. Although conjugate vaccines against some GBS serotypes have been developed and tested, they only provide protection against homologous serotypes and not necessarily against other serotypes of different geographical distributions⁸. This was reflected by an analysis of numerous GBS genomes which suggested that the pan-genome was 'open'; even if hundreds of GBS genomes are sequenced, the pan-genome will never be fully sampled¹. Therefore, a comparative reverse-vaccinology approach is crucial for the development of a universal GBS vaccine. Maione *et al.*⁸ compared the genome sequences of 8 GBS isolates, and predicted a set of 1,811 core genes that are shared by all strains and a dispensable gene pool of 765 genes of limited phylogenetic distribution. These 2 gene pools were independently used as inputs to *in silico* methods to predict 396 core and 193 dispensable genes that encode putative surface-exposed proteins: half of the predicted proteins were successfully expressed in *Escherichia coli*, purified and used to immunize female mice. This reverse-vaccinology approach yielded four putative vaccine candidates that significantly increased the survival rate of challenged neonatal mice. The most important contribution of this study was the fact that although none of the four antigens

represented a universal vaccine, owing to either the limited phylogenetic distribution of the corresponding gene or the minimal surface accessibility of the antigen, a combination of the four antigens provided protection against all nine major GBS serotypes. This also implies that conventional vaccine development or even a classical reverse-vaccinology approach (using a single genomic sequence) would probably have failed to identify all four antigens.

“There are two versions of reverse vaccinology: classical reverse vaccinology, which uses only a single genome sequence to make predictions about putative vaccine candidates, and comparative (or pan-genomic) reverse vaccinology, which compares the genomes of several closely and distantly related strains.”

Neisseria meningitidis, a Gram-negative betaproteobacterium that colonizes the mucosal surface of the nasopharynx of healthy individuals, is another commensal that only occasionally invades the blood and the cerebrospinal fluid to cause septicaemia and meningitis. There are five pathogenic *N. meningitidis* serogroups, namely A, B, C, Y and W135, which are classified according to their capsular polysaccharides (CPSs). Polysaccharide conjugate vaccines are available against all of the serotypes except for B (MenB), which is responsible for one-third of all meningococcal disease in the United States and for 45–80% of the cases in Europe. There are two main reasons why no successful MenB vaccine has been developed so far, both of which are limitations of conventional vaccine development: first, the CPS that produced successful vaccines against the other four serotypes is chemically identical in MenB to a self-antigen that is present in