Rev. sci. tech. Off. int. Epiz., 2007, **26** (1), 49-67

Genomics and vaccine development

C.G. Gay [1], R. Zuerner [2], J.P. Bannantine [2], H.S. Lillehoj [3], J.J Zhu [4], R. Green [1] & P.-P. Pastoret [5]

(1) United States Department of Agriculture, Agricultural Research Service, National Program Staff, Animal Production and Protection, Beltsville, MD, United States of America
(2) United States Department of Agriculture, Agricultural Research Service, National Animal Disease Center, Ames, IA, United States of America
(3) United States Department of Agriculture, Agricultural Research Service, Parasitic Diseases Laboratory, Beltsville, MD, United States of America
(4) United States Department of Agriculture, Agricultural Research Service, Foreign Animal Diseases Laboratory, Plum Island, Orient Point, NY, United States of America
(5) World Organisation for Animal Health, 12, rue de Prony, 75017 Paris, France

Summary

The current explosion in new high-throughput technologies arising from microbial and animal genomics studies is enabling the analysis of the genome, transcriptome, and proteome and offers the opportunity to gain a better understanding of the molecular pathways underlying pathogen biology, the host immune system, and host–pathogen interactions. These new tools can be applied to veterinary pathogens to overcome some of the current hurdles in the discovery of highly effective vaccines for farmed livestock and poultry.

Keywords

Animal genomics – Immunogenomics – Microarray – Microbial genomics – Vaccines – Vaccinogenomics.

Introduction to microbial genomics

The field of microbial genomics provides exciting new opportunities in the control and prevention of a wide range of veterinary diseases. Genomics, and the functional analysis of genomic data, are leading to novel approaches for vaccine discovery, and improved methods for diagnosis and epidemiology. Genomes of several viral and bacterial pathogens that impact veterinary medicine have been sequenced. Each of these studies has provided new information and unique views into viral and bacterial pathogenesis. In this introduction, the authors provide a brief overview of how bacterial genomic sequences are deduced and how genes are identified from these data. In the following section, they provide a brief overview of some bacteria of importance to veterinary medicine for which genomic sequences have been determined and they describe how genomic data can be exploited to understand the ecology and epidemiology of pathogenic bacteria, a critical element for disease control and prevention.

Currently, two basic approaches are used for determining the sequence of bacterial genomes. Both methods use a 'shotgun' approach, whereby random segments of the genome are sequenced. In the traditional method, plasmid libraries of cloned DNA fragments are constructed, and portions of the cloned DNA adjacent to plasmid vector sequences are determined by primer extension reactions (28). In an alternative pyrosequencing method, short DNA fragments are attached to microbeads, the fragments amplified, and a series of extension reactions are done that record the sequence of each fragment (60). Both techniques have strengths and weaknesses. The traditional method is labour intensive, but yields individual sequencing reads of 800 to 1000 bp in length that are paired with the complementary sequences obtained from the opposite end of the cloned fragment being physically linked. This facilitates construction of a scaffold on which

the entire sequence assembly can be constructed. In contrast, the pyrosequencing method is rapid, generating about 2×10^7 bp of data in a single 5.5 h run, but each sequence is limited to about 100 bp and is not physically linked to other fragments (60). The output from these two methods is processed by various computer programs that identify regions that have identical and overlapping sequence, and these are assembled into a series of contiguous blocks of genomic sequence. These contiguous segments of the genome are referred to as contigs, and the next step in the assembly process is to join adjacent contigs together until the chromosome or plasmid sequence (commonly referred to as a replicon because it is an autonomous replicating unit) is completely connected. Polymerase chain reactions (PCR) and primer walking of selected templates are used to improve sequence quality, and should ultimately yield a single contig per replicon.

Genomic data are processed by a variety of software programs that help identify individual genes, and translate them into the predicted protein products. Different proteins with a common function often share segments with a similar sequence of amino acids. Protein segments having shared sequences, and presumably similar functions are referred to as motifs. For example, the presence of arginine-glycine-aspartic acid (RGD) motifs in proteins facilitates binding with integrins, a feature that can be exploited by pathogens such as *Mycobacteria*, which uses the RGD-containing protein encoded by the gene *iipA* for macrophage invasion (32). The presence of common sequence motifs among lipoproteins and other proteins secreted from the cytoplasm is useful for identifying potential membrane and secreted proteins. The accumulation of this information establishes a framework for subsequent biochemical and pathogenesis studies that can lead to characterisation of previously unidentified virulence factors and antigens.

Using genomics to understand the ecology and epidemiology of infectious diseases

The sequencing and analysis of microbial genomes is fundamentally changing our understanding of the ecology and epidemiology of important pathogenic microbes and providing new insights into predictive biology and the discovery of effective countermeasures for disease control and eradication.

The genomes of some veterinary bacterial pathogens have now been sequenced, including those of several important zoonotic agents. Comparative analysis of *Mycobacterium* spp. revealed evidence of genome reduction in *M. bovis*, the cause of tuberculosis in cattle, with a general loss of functional redundancy (33). The more recent analysis of *M. avium* subspecies *paratuberculosis*, the cause of Johne's disease in ruminant species, helped identify potential targets for diagnostic tests and yielded new data regarding metabolism (54). Leptospirosis is caused by diverse pathogenic members of the genus *Leptospira*, and four strains of *Leptospira* have been sequenced, including two *L. interrogans* strains and two strains of *L. borgpetersenii* (11, 69, 87). Comparative analysis revealed substantial genetic differences that probably affect how these two *Leptospira* species are transmitted between animals (11). Comparative analysis of three *Brucella* species, *B. abortus* (40), *B. melitensis* (23), and *B. suis* (75), suggests that changes in the function of transcriptional regulatory proteins and expression of outer membrane proteins have led to differences in host specificity (14). *Bordetella avium*, a pathogen of poultry, is distantly related to the mammalian pathogens *Bo. bronchiseptica*, *Bo. pertussis*, and *Bo. parapertussis* (73) and comparison of the genomic sequences have shown distinct evolutionary patterns of adaptation to avian vs. mammalian hosts (14). Among the *Bordetella* species that infect mammals, host restriction appears to be a result of gene loss (92); these studies provided a platform that was useful in understanding the evolutionary changes that have occurred in *Leptospira* (11).

Development of an annotated genomic sequence establishes a framework through which targets for epidemiological analysis can be identified. Genomic sites that contain tandem repeated sequences often vary in the number of copies of the core repeating sequence among different strains due to errors that occur during DNA replication. Because changes in copy numbers within different variable nucleotide tandem repeats (VNTR) accumulate at independent rates, simultaneous analysis of multiple VNTRs provides a powerful method for differentiating similar strains of bacteria. Development of VNTR-based tools using PCR for epidemiological studies of several bacterial pathogens has been made possible through access to genomic sequences, and is especially useful in characterising organisms that are otherwise difficult to differentiate, including *Brucella* (7, 53, 105), *Mycobacteria* (54), and *Leptospira* (59, 90, 96, 97) species.

Another application of genomic sequencing data is the development of microarrays that provide hybridisation targets representing the entire genome, all placed on a microscope slide. Microarrays allow investigators to assess genetic variation between isolates and characterise global patterns of gene expression. For microarray analysis, RNA or DNA samples are differentially tagged with chemical labels and used to hybridise with DNA targets on the array. Unhybridised material is removed by washing and the retained, tagged samples are modified with a chemical that fluoresces when excited by lasers in a specialised instrument. The intensity of each spot, representing a

hybridisation target, usually a specific gene, is measured and compared to control samples to determine either genetic diversity (DNA input) or differential gene expression (RNA input). For example, genetic differences among *M. avium* subspecies *paratuberculosis* were identified by microarray studies (61, 76), resulting in valuable information on bacterial adaptation to different mammalian hosts. Microarray analysis of *Pasteurella multocida* gene expression during growth in chickens revealed a subset of genes induced by infection that are also expressed in response to iron limitation (6). Similar studies using *in vitro* models have also helped characterise changes in gene expression (e.g. temperature response in *L. interrogans* [57] and invasion of bovine epithelial cells by *M. avium* subsp. *paratuberculosis* [76]) and are helping to expand our understanding of how bacteria respond and adapt to growth in the natural host. A key point in using microarrays to study gene expression is that many putative genes identified by genomic analysis encode proteins of unknown function. By identifying genes that respond to environmental stimuli rather than selecting genes based on a bias formed by presumed function, it may be possible to identify bacterial proteins essential for survival in the host. This information is critical for rational selection of proteins for development as subunit vaccines.

Genomic analysis is also helping to develop a more comprehensive understanding of signals used to direct proteins to extracytoplasmic locations, including the outer membrane. Outer membrane proteins (OMPs) are often considered ideal vaccine candidates, and improved methods for identifying protein motifs that direct proteins to the outer membrane are essential to assign presumptive locations of proteins with unknown function. This problem is of particular importance in identifying putative OMPs in spirochetes, a distinct group of bacteria with an unusual cell wall/cell membrane structure. Models for predicting OMPs based on other bacteria were potentially misleading when applied to spirochetes. Availability of the genomic sequences for several pathogenic spirochetes enabled Setubal *et al.* (94) to develop an algorithm with improved predictive power to identify potential OMPs in this group of bacteria. This information is being used to help select and analyse potential vaccine candidates for a wide variety of spirochete diseases, including leptospirosis (31).

Viruses, due in part to their small size, are more easily compared using genomic approaches than bacteria, and new studies are providing useful information on strain variation. For example, comparison of the genomes from 45 strains of variola (smallpox) virus provided improved epidemiological analysis which would be invaluable in the case of virus release, and helped to identify proteins that may affect virulence (27). A similar comparison of 103 foot and mouth disease virus (FMDV) isolates revealed highly conserved, and presumably essential regions of the genome (13). Analysis of these FMDV isolates also found evidence for recombination, leading to increased diversity (13), potentially confounding epidemiological analysis and resulting in the discovery of vaccines that may be effective under experimental conditions but ineffective in the field.

These analyses illustrate how genomic sequencing is increasing our understanding of the interaction of important pathogenic microbes with their environment and facilitating the identification of relevant targets for designing vaccines that are effective under field conditions. In the next section, the authors describe approaches for analysing genomic sequences to enable the rational design of new vaccines and identify sensitive diagnostic targets.

Whole genome analysis of pathogens in vaccine discovery

Although there are comparatively far fewer completely sequenced genomes for bacterial pathogens of farmed livestock than there are for human pathogens, several completed sequences of these pathogens have recently become available for studies in vaccine development. Cattle pathogens such as *Bacillus anthracis* (83), *Mycobacterium avium* subspecies *paratuberculosis* (54), *Brucella abortus* (40) and *Leptospira* (11, 69, 87), as well as swine and poultry pathogens including *Pasteurella multocida* (62) and *Bordetella avium* (92) are among the veterinary pathogens with published sequences. Thus far, very few vaccine developments flowing from this genomic information have been published owing to the recent completion of the sequencing projects. However, two veterinary vaccines, identified through genomic approaches, have shown promise and are worthy of mention.

Eimeria is a protozoan parasite that causes coccidiosis in livestock and is especially costly to the poultry industry, with estimated annual worldwide losses at US$ 800 million (106). Efforts to develop a vaccine against this parasite have been difficult not only because of the several thousand genes and hence potential antigens encoded by this parasite, but also because the majority of candidate molecules are immunogenic, but few of these same molecules are immunoprotective (stimulate a protective immune response) (3). By using linkage analysis of DNA markers combined with an understanding of the parasite infection cycle, investigators have identified four key regions within the *Eimeria* genome capable of stimulating protective immunity (2, 3). Some progress has also been made with *Brucella abortus*, a facultative intracellular pathogen that causes abortions in livestock. Researchers analysed sequence data of *Brucella* and found the *exsA* gene (88). Further bioinformatics analysis showed

ABC transporter motifs present within the gene product, which suggest polysaccharide transport functions are critical to *Brucella* virulence. A deletion mutant of *exsA* was constructed and this attenuated strain showed protection in mice (88).

Examples of genome-wide applications with human pathogens are more common and provide a pathway that studies involving veterinary pathogens will no doubt follow. For example, Group A *Streptococcus* (GAS) has long been known to cause a variety of human diseases that range from pharyngitis to skin invasion/infection (19, 86), but the molecular basis for the phenotypic differences underlying these widely varying disease presentations were unknown until recently. By comparing complete transcriptomes of a GAS pharyngeal isolate and GAS skin isolate, investigators found 89 genes differentially expressed, 24 of which were virulence genes (103). These investigators followed up on this observation by completely re-sequencing the pharyngeal isolate and found a 7-bp frameshift mutation in a two-component regulator encoded by *covRS*. This mutation resulted in a truncation of the CovS protein, a histidine sensor (103). This mutation was exclusively responsible for the diverse disease phenotypes of each GAS isolate and has resulted in solid vaccine leads and intervention strategies.

Perhaps one of the best-known examples of genomic application to vaccine development is with *Neisseria meningitides* Group B. Investigators sequenced the complete genome of this human pathogen and recombinantly expressed 350 genes predicted to encode surface-exposed or secreted proteins (77). Immunological assays quickly trimmed the list of vaccine candidates to seven, of which some were later shown to be protective in animal studies (104). Thus, the term 'reverse vaccinology' was coined to reflect how genomic approaches allow for the design of vaccines starting from the prediction of all antigens *in silico* (performed by computer simulation), independently of their abundance and without the need to grow the micro-organism *in vitro* (82).

Accordingly, the genomic approach that can most readily be applied to vaccine discovery is the creation of subunit vaccines. In most cases, the production of killed or whole-cell protein preparations and attenuated live vaccine strains does not need genomic technology, but rather a protein chemist or microbiologist. However, the identification of suitable antigens is crucial to successful vaccine development based on subunit approaches. Using a combination of proteomics, genomics and bioinformatics, investigators can quickly narrow the list from thousands of genes down to a few dozen vaccine candidates (Figure 1). The genomics/proteomics methods define the coding capacity, and then the bioinformatic analyses trim off pseudogenes (nonfunctional or noncoding) and sequences similar to human proteins and make predictions about secreted and surface-located proteins. Then it is back to the laboratory to recombinantly produce these proteins and determine their immunogenicity (Fig. 1), although bioinformatics can make limited predictions along these lines as well. From these exercises emerge the short list of solid vaccine candidates to test in an animal model for protection.

Genomics also has the capability to make DNA vaccination studies much more efficient. Before the genome sequence was available for *Mycobacterium avium* subspecies *paratuberculosis*, DNA vaccination was attempted for this cattle pathogen using the expression library immunisation procedure (47). This study revealed two pools of DNA that were shown to be protective in mice and limited efforts were made to identify the relevant DNA in those pools (44). Random expression library immunisation was used because the genome sequence was not available at the project's inception. This random cloning method meant that the majority of clones would be in the opposite orientation relative to the coding strand or would be out of frame with the coding sequence. Therefore, many additional clones were needed to make the library truly representative of every coding sequence in the genome. An approximate total of 16,500 clones were used to immunise mice in that study (44). With the genome sequence now complete (54), a directed expression library immunisation project, in which each clone faithfully represented a single coding sequence, could be initiated. This method has the advantage that fewer clones are needed, making resulting clone pools less complex and there is no 'garbage' or nonfunctional clones such as those in the opposite orientation or out of frame. For such a study, only 4,350 clones would be needed because that is the total number of genes present in the *M. avium* subspecies *paratuberculosis* genome (54). An added benefit is that fewer mice would be needed to test the clone pools.

Genomic approaches can also identify the best targets for knockout mutations that enable engineering of attenuated vaccine strains. However, as yet, there are no published studies for bacterial pathogens that demonstrate a genome-wide approach that can identify a target, knockout this target, and show both attenuation and protection in an animal host. Rather, the literature reports studies in which genomics has been used to genetically define a known vaccine strain. The *Salmonella typhimurium* vaccine strain is protective in mice and lacks the transcriptional regulator RfaH. Use of whole genome microarrays identified the RfaH-dependent genes giving investigators insight into the mechanism of attenuation for this vaccine strain (68).

The most famous example of the use of genomics to define an attenuated bacterial vaccine strain is *M. bovis* BCG (Bacillus Calmette-Guérin – named after the the French scientists Calmette and Guérin), which is the most widely used global vaccine to prevent human tuberculosis (TB).

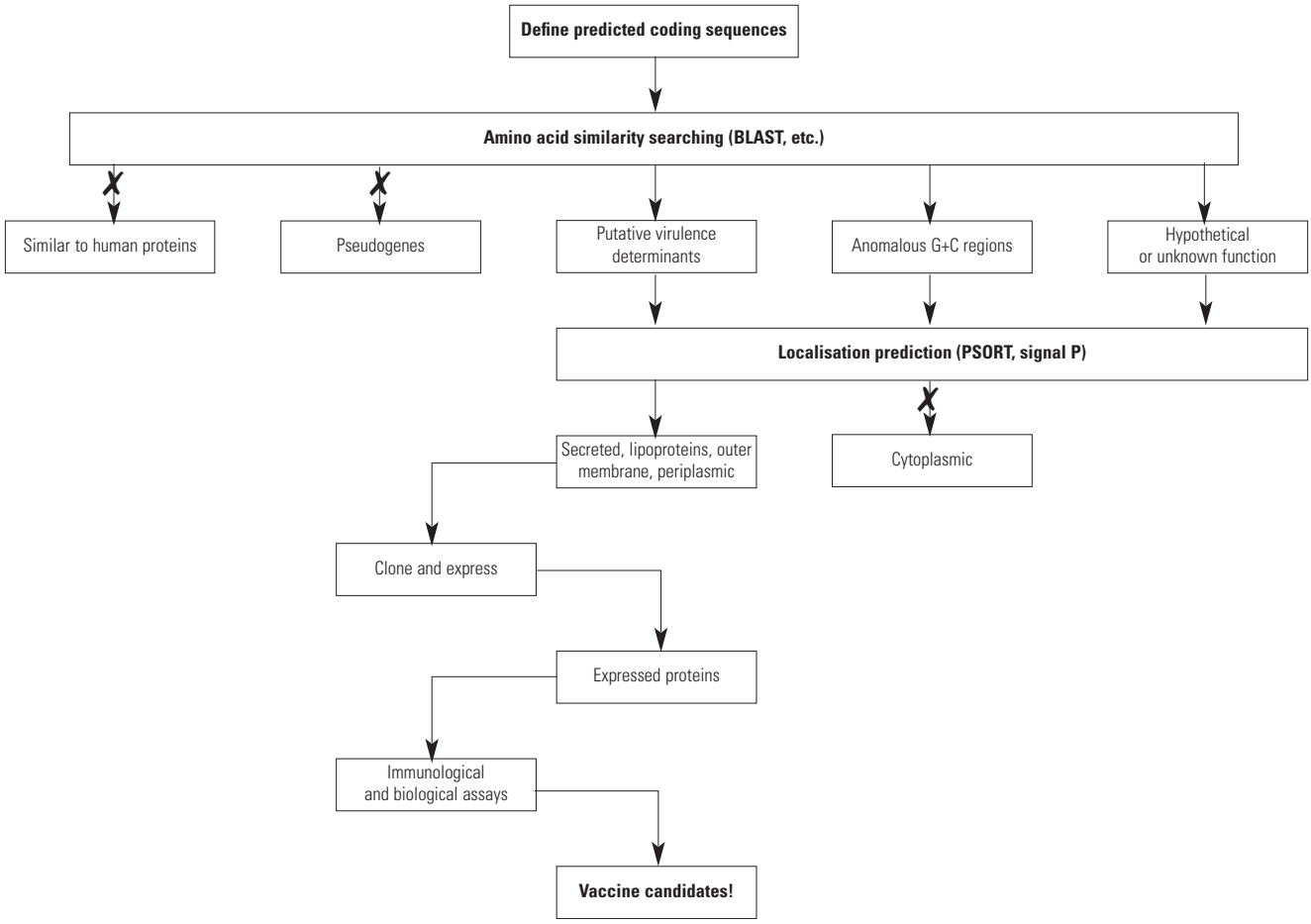

Fig. 1
Schematic flow diagram showing the genomic approach to obtaining vaccine candidates for use in subunit vaccine approaches
The initial step involves annotating the genome to define its coding capacity and hence all potential antigens. Bioinformatics similarity searches should then be performed to discard pseudogenes and anything resembling human proteins. The remaining list of genes is then cloned and expressed and analysed immunologically for vaccine candidates

Over 3 billion individuals have been vaccinated with BCG without major side effects (4). The BCG vaccine strain was derived long ago from a fully virulent isolate of *M. bovis* by prolonged serial passage of the bacterium resulting in its attenuation (66). However, the molecular basis for this attenuation was never understood until the complete genome sequences of *M. tuberculosis*, the causative agent of TB, and *M. bovis* BCG became available (16, 29, 33). Genomic comparison of these two species revealed one region of deletion in BCG, termed RD1 (8, 79). This region contains the well-known antigen ESAT-6, a secreted antigenic target that strongly induces Th1 immune responses (100).

More than any other benefit, whole genome analysis of pathogens enables the targeted selection of protective immunogens encoded by the disease-causing pathogen. This allows investigators to move away from empirical approaches in vaccine development towards a more focused, logical development and discovery of protective DNA segments and proteins. In the next section, the authors describe the applications of bioinformatics in the design of the ideal vaccine.

Bioinformatics and computational vaccinology

Designing an ideal vaccine depends greatly on several factors associated with targeted pathogens and host responses, including knowledge at the molecular level of the immune response, pathogenesis, host–pathogen interaction, and genetic and physiological variation among animals and pathogens. Recently discovered genome sequences of food animals and pathogens together with rapid advances in biotechnology will allow us to collect an unprecedented amount of information on hosts and pathogens that may have significant implications for

vaccine discovery. However, transforming this information into practical understanding requires intensive data-mining using sophisticated computational and bioinformatic tools. Highly intensive computation using high-speed central processing unit, multi-thread, and 64-bit technologies have greatly facilitated this process. Using computational approaches in vaccine design has become known as 'computational vaccinology' (30).

Applying bioinformatics algorithms to facilitate vaccine design is a very powerful approach that is changing many paradigms of vaccine discovery (81, 91, 93). As discussed in the previous section, a new approach is reverse vaccinology, the process of using *in silico* analysis of genetic information instead of pathogens themselves as the starting point (81). This approach has resulted in several successful vaccines that conventional methods would have taken much longer or failed to produce (81, 93). Thus, we can now use a genome-based approach in reverse vaccinology where the genome sequence of a pathogen is screened with bioinformatic tools to identify open reading frames that may encode candidate proteins. Proteins predicted to be surface-exposed or secreted are considered as vaccine candidates for further laboratory testing. Some proteins having structures similar to known toxins can also be included in the candidate list. If the genome sequences of different strains (virulent and avirulent) or serotypes are available, a pan-genome approach can also be used to identify candidate vaccines by comparative genomics. The applications of these approaches in vaccine development have been reported (77, 86).

If candidate antigens are identified, peptide vaccines can be developed based on the epitopes of the antigens. 'Immunoinformatics' – the new science of epitope prediction – applies bioinformatics to the design of peptide vaccines (50). Antigen processing and presentation in the adaptive immune response are well-known at the molecular level. B-cell epitopes can be either linear or discontinuous amino acid residues dependent on the conformation of protein antigens (surface accessibility), whereas T-cell epitopes are short linear peptides that are processed by proteases and presented by class I and II major histocompatibility complex (MHC) molecules. These epitopes can be mapped using laboratory procedures, which are costly and labour intensive. The epitopes can also be predicted using various bioinformatic algorithms. Currently, T-cell epitopes are more predictable than B-cell epitopes due to the linear nature of the former. The prediction of T-cell epitopes can be based on anchor motifs in the binding pockets of MHC molecules (71), or on training sets of laboratory tested data, using statistical methods such as a hidden Markov model or machine learning methods, e.g. artificial neural networks and support vector machines (10). A protein called 'transporter associated with antigen presentation' (TAP) selectively transports endogenous antigenic peptides into the endoplasmic reticulum (ER) for class I MHC antigen presentation. This selectivity can be taken into consideration in the prediction of class I MHC epitopes (58, 108).

In contrast to T-cell epitopes, B-cell epitopes remain much less predictable (5, 30). Recently, using recurrent neural network (89), machine learning classifiers (98, 99), and structural-energetic analysis (12) improved the prediction of continuous B-cell epitopes, whereas the combination of protein 3D structures and statistics has been used to predict discontinuous B-cell epitopes (42). Although the technical difficulties of predicting B-cell epitopes remain to be overcome, combining laboratory and bioinformatic analysis, such as phage display and mimotope analyses, can increase the accuracy of predicting continuous and linear epitopes (65, 78). Mimotopes were first described as peptides that mimic native epitopes of foot and mouth disease virus and can bind to the same antibody as native antigens (34, 35). Candidate vaccines can be identified based on mimotopes that can induce antibody capable of binding to native antigens of pathogens (74). This approach may be useful for developing multi-epitope vaccines to fight against pathogens with several serotypes, such as foot and mouth disease virus.

One of the challenges of epitope-based vaccines is population coverage due to MHC polymorphism. Different MHC molecules display distinct peptide-binding specificity (84, 85). However, it has been shown that certain MHC alleles share overlapping peptide-binding specificity and the alleles can be grouped into supertypes based on their common binding specificity (95). Predicting peptides that bind to MHC supertypes for vaccine development can avoid the complication of MHC polymorphism. MHC alleles can also be grouped into supertypes based on the bioinformatics analysis of MHC protein structures and sequences (24), and supertypic MHC ligands can be predicted for multi-epitope vaccine development to increase population coverage (84). It has been estimated that targeting only 3 to 6 class I HLA alleles should cover ~90% of the human population because of linkage disequilibrium in the MHC loci (39). MHC genes are also tightly linked in food animals (51, 52).

Another application of bioinformatics in vaccine development is the interpretation of data collected with functional genomics approaches to gain detailed understanding of the immune response, pathogenesis, and host–pathogen interaction. The knowledge obtained can be implemented in vaccine design. DNA microarray and proteomic analyses are two common approaches used in the studies of functional genomics, measuring transcript and protein expression levels, respectively. Because gene expression levels are collected in a genome-scale, the data

must be stored in databases in order to be managed and analysed effectively. The data also contain a large portion of technical variation introduced by laboratory procedures. The variation must be removed or minimised by data normalisation before statistical analysis (80). Because multiple statistical tests are used in the data analysis, significant thresholds must be adjusted to balance between false positives and false negatives in detecting differentially expressed genes (25). Differential gene expression can be further analysed to infer biological conclusions based on known molecular pathways and gene functions (20). Bioinformatic analysis will play a very important role in animal health by generating the detailed knowledge needed for rational vaccine development.

In summary, bioinformatics has become an additional powerful approach in vaccine design. The impact of the application of bioinformatics on rational vaccine design will be very significant in the future as research in this field progresses. Short synthetic peptides have been considered to be the next generation vaccines (93); however, there are several technical difficulties in using peptides as vaccines (41). Many of the obstacles could be overcome by bioinformatic approaches. Currently, there are many challenges confronting animal health in the areas of disease prevention and eradication. Bioinformatics may allow us to take all relevant information into consideration, including the genetic diversity of hosts and pathogens, to formulate vaccines that have broader effects regardless of these variations. Combining genomics and biotechnologies, bioinformatics can provide us with the detailed knowledge needed for vaccine development. However, the tools and infrastructure to facilitate these applications in animal health have yet to be fully developed. The next section provides an update on the animal genome initiatives.

Animal genomics

In the past two decades, molecular biology has changed the face of agricultural animal research, primarily in the arena of genomics and the relatively new offshoot areas of functional genomics, proteomics, transcriptomics, metabolomics and metagenomics. Development of genetic and physical genome maps in the past 15 years has given rise to the possibility of being able to understand the molecular nature of the genetic component of phenotypic variation. While quantitative geneticists have been successful in improving production traits, genomic technology has potential to lead to more accurate and rapid animal improvement, especially for phenotypic traits that are difficult to measure.

In the mid-1980s, a new window of opportunity opened in livestock production science. In 1986, the term 'genomics' was coined to name a new journal in which science generated from the new technologies that had been developed and applied to the study of mammalian DNA (principally for the Human Genome Initiative) could be published. Technologies that were being used at that time included such things as the application of bacterial restriction endonucleases for rudimentary visualisation of differences in the sequence of DNA, in particular chromosomal locations through 'restriction mapping'. This was followed quickly by the development of the polymerase chain reaction (PCR) in 1987, which opened up an entirely new world for the study of differences in the DNA sequence of animals. Coupled with the discovery of short tandem repeat DNA markers, PCR became a powerful tool that quickly allowed the development of genetic maps of the livestock genomes, primarily based on linkage of microsatellite DNA markers. These maps were developed and published in the early 1990s.

By the time genetic linkage maps were in place, a number of research groups around the world had developed resource family populations that were being employed to identify regions of the genome appearing to harbour genes giving rise to phenotypic variation in complex economically important traits (so-called Quantitative Trait Loci [QTL]). Once DNA markers anchoring these QTL regions were identified, it was postulated that 'marker-assisted selection' could be used to make directed genetic change in the desired traits using this technology.

By the end of the 20th Century, it was recognised that more genomic tools and resources were necessary for the fulfilment of the promise of livestock and poultry genomics. Although a large number of putative QTL had been identified for a wide spectrum of traits, only a handful of causative mutations had been elucidated through this approach. In all of these successful cases, the fine mapping of the identified genes had relied on comparative mapping approaches to make use of the denser information available in the human, mouse, and rat maps. Despite having some genomic resources, such as bacterial and yeast artificial chromosome libraries, it became clear that without the availability of the whole genome sequence as a scaffold from which to work, the time and expense of fine QTL mapping was much greater than initially envisioned. Fortunately, new high-throughput technologies were being developed that made the sequencing of whole genomes more practical, efficient, and cost effective. The human genomics research community quickly recognised this opportunity and the government and privately funded human genome sequencing projects were launched.

As the 21st Century began and the human genome moved toward an initial draft sequence, other new technologies also became available that allowed livestock and poultry researchers to move into large-scale gene expression studies for the first time. By coupling expressed sequence

tags (ESTs) with new microarray technologies, researchers were able to visualise changes in levels of expression of hundreds or thousands of genes in specific tissues under a wide variety of conditions. This began to broaden genomics research into the functional realm and initiated open discussions on how genomics might be used to bridge various disciplines into a 'systems biology' framework.

Recently, the agricultural research community has been able to capitalise on the infrastructure built by the human genome project (17, 46) by sequencing two of the major livestock genomes (*Gallus domesticus* [45, 107] and *Bos taurus* [36]). The 2006 calendar year marked a major milestone in the history of agricultural animal research since annotated draft genome sequences were completed for chickens and cattle and sequencing was initiated for the porcine and equine genomes. We now have in place a powerful toolbox for understanding the genetic variation underlying economically important and complex phenotypes.

Developed concomitantly with these genome projects has been a suite of associated tools, including:

– EST libraries

– bacterial artificial chromosome maps

– integrated physical and linkage maps

– full-length complementary DNA (cDNA) libraries

– microarrays or gene chips

– identification and validation of a large number of single nucleotide polymorphism markers.

Currently, major efforts are underway to develop haplotype maps of these genomes in order to fine map QTL and enable whole genome selection for quantitative traits (48).

While the maturing field of livestock genomics has been largely centred on improvement of production traits up to the present time, it is widely recognised that the highest potential of these technologies resides in difficult to measure and expensive traits such as efficiency of nutrient utilisation and resistance to disease. In particular, genomics holds great promise for unravelling the interactions between various hosts and pathogens (38). Understanding host–pathogen interactions at the molecular level will increase our understanding of viral and bacterial pathogenesis and the mechanisms pathogens use to evade host immune responses, both of which are paramount to the discovery of the ideal vaccine for control and eradication of animal diseases. The next section describes the role of functional genomics and the application of microarray technologies to understand host–pathogen interactions at the genomics level.

Host–pathogen interactions at the genomics level

Recent progress in sequencing the genomes of microbial pathogens and their hosts is providing sophisticated strategies for unravelling the biological complexity of host–pathogen interactions (18). Elucidating these interactions at the molecular level, however, remains largely unrealised because understanding of gene function lags behind gene expression analyses obtained through high-throughput, large-scale functional genomics approaches. Nonetheless, functional genomics is rapidly revolutionising the analysis of whole genome responses of pathogens and hosts. This will lead to a better understanding of disease processes, the mechanisms through which pathogens evade host immunity and the genetic basis of host–pathogen interactions, which will ultimately result in the discovery of novel vaccines. Collectively, the integration of these approaches in vaccine research (vaccinogenomics) is likely to fundamentally change the way scientists approach the challenges of discovering safe and effective vaccines.

DNA microarray technologies allow high-throughput measurement of global gene transcription patterns on a whole-genome or tissue-specific basis, thereby enabling the investigation of the transcriptional status of complex biological systems underlying host–pathogen interactions. Specifically, genomic technologies combined with immunology (immunogenomics) permit in-depth analysis of complex immunological processes based on large-scale whole genome approaches. Unlike conventional methods of differential gene expression (e.g. SAGE [serial analysis of gene expression] and differential display) that enable functional annotation of sequenced genomes, DNA microarray hybridisation analysis stands out for its simplicity, comprehensiveness, data consistency, speed, and high-throughput methodologies.

Global profiling of host and pathogen gene expression is an attractive approach to identifying the novel genes involved in disease processes since, in general, genes are transcribed only when and where their function is required. Thus, determining the conditions under which a given gene is expressed allows inferences to be made about its function. For example, this approach has led to the annotation of the function of multiple microbial genomes, probing a microbe's physiological state, identifying candidate virulence factors, pharmacogenomics (drug-specific signature gene expression), and molecular genotyping (molecular diagnostics for genotyping polymorphisms in related pathogens). Similarly, host genomic analyses have led to a better understanding of the response to pathogenesis, the development of diagnostic gene expression profiles, the dissecting of the genetic basis of disease susceptibility, and the characterisation of genetic polymorphisms associated with diseases.

Gene expression profiling

A variety of human DNA and oligonucleotide microarrays are commercially available. The most commonly used host microarrays are largely composed of ESTs. DNA arrays have become popular because they are generally considered to be easier to use than other gene expression profiling methods, and they allow the simultaneous quantification of thousands of genes from multiple samples. DNA array technologies rely on nucleic acid hybridisation between labelled free targets derived from a biologic sample, and an array of DNA fragments (the probes, representing genes of interest) tethered to a solid surface (9, 101, 102). The targets, often produced by reverse transcription of messenger RNA (mRNA) and simultaneous labelling of the corresponding cDNAs, form a complex mixture of fragments that hybridise with their cognate probes during the assay. The signal generated on each probe reflects the mRNA expression level of the corresponding gene in the sample. After detection, quantification, and integration of signals with specialised software, intensities are normalised for technical deviations, providing a gene expression profile for each sample that may be compared with the profiles of other samples. Standard, robust statistical methods are required for assigning significance values to gene expression measurements and to infer meaningful information. Although most global gene expression analyses have used some form of clustering algorithm to find genes coregulated across the dataset, under a primary assumption that shared gene expression often implies shared function, more sophisticated data mining techniques and specialised analysis tools may be needed to extract meaningful biological insights.

When applied to host–pathogen interaction studies, gene expression profiling has been commonly used to analyse altered expression patterns during disease states, thereby elucidating the mechanisms of disease and pinpointing the functional pathways involved in the host response to infection. Furthermore, comparison of temporal gene expression patterns during microbial infections has facilitated novel gene discovery for use in candidate vaccines and biotherapeutics. One underlying assumption of DNA microarrays is that genes are preferentially expressed when their functions are required. Given this assumption is correct, application of expression profiling in host–pathogen studies allows one to examine the functions of the genes of both hosts and pathogens: by using pathogen arrays, one can monitor the expression of microbial genes, characterise the functions of unknown genes, identify virulence-associated genes, measure physiological adaptations under various environmental conditions, and evaluate the effects of drugs and vaccines. Similarly, by using host gene microarrays, one can explore host responses at the level of gene expression and provide a molecular description of the events that follow infection. Host profiling may also identify gene expression signatures unique for each pathogen and in genetically disparate hosts, thus providing novel tools for diagnosis, prognosis, and clinical management of infectious diseases. Together, this information will guide the future design of a new generation of molecular vaccines.

Microarray applications in host–pathogen interaction studies

Strategies to investigate host–pathogen interactions using high-throughput gene expression analysis have been described utilising various *in vitro* and *in vivo* models with whole genomic or tissue-specific microarrays. The main objective of these studies is to identify groups of genes that are involved in the activation or repression of key regulatory pathways of interest. Additionally, high-throughput gene expression arrays allow one to investigate the temporal sequences of induction or repression of transcription, a prerequisite for determining the order of events following host–pathogen interaction. In most cases involving complex disease processes, it is difficult to investigate all of the interacting factors *in vivo*. Thus, in order to reduce the complexity of whole animals, and to facilitate the interpretation of genomic data, *in vitro* systems have been exploited (e.g. homogeneous cell lines that are relevant to the type of study), the results of which are compared to the results obtained with *in vivo* studies. In either case, the interpretation of gene expression changes will be challenging, and it is important that the results of microarrays are validated using other methods, such as reverse transcription PCR or Northern blotting.

In vitro studies

The first reported application of whole genome expression arrays to analyse host–pathogen interactions used primary human fibroblast cells infected with human cytomegalovirus (CMV) (109). RNA samples collected at 40 min, 8 h, and 24 h after CMV infection were used to interrogate gene chips containing oligonucleotides corresponding to >6,600 human mRNAs (GeneChip microarray, Affymetrix, Santa Clara, California, USA). At 40 min post-infection, 27 mRNAs showed significant alterations in expression, and at 8 h and 24 h, the number of altered genes increased to 93 and 364, respectively. These high numbers of genes were in contrast to previous results obtained by differential display that identified 15 interferon-inducible genes activated by CMV.

Although CMV replicates in many different cell types and the response may be different from those seen in primary human fibroblasts, it can be speculated that many of the genes identified using the GeneChip array are involved in early response of host cells to this virus. Therefore, it is not

surprising that data analysis using GeneChip software showed that substantial transcription changes began very early after infection involving the activation of many early transcription factors and proinflammatory signalling molecules, including cytokines, chemokines, stress-inducible proteins, and interferon-inducible proteins. These results were consistent with the expected host cellular response to CMV infection. In particular, CMV modulated the expression of genes involved in the production of prostaglandin E2 from arachidonic acid, indicating that prostaglandin E2-mediated inflammation is part of the host response to CMV infection. The data also revealed altered expression of immune-related genes. For example, upregulation of HLA-E mRNA by a factor of 6 was observed. Interestingly, genes encoding HLA-A, HLA-D, and HLA-G were not changed. HLA-E is a nonclassical class 1 major histocompatibility molecule whose expression has been associated with pathogen evasion of host NK cell recognition. Thus, identification of key host genes whose functions provide tantalising hints of potential mechanistic roles in disease processes underscores the utility of gene array technologies in the study of disease pathogenesis.

In a second example of the use of whole genome expression arrays, gene expression analysis was used to investigate proinflammatory responses of human intestinal epithelial cells infected with *Salmonella* (26) and human promyelocytic cells infected with *Listeria monocytogenes* (15). In both cases, genes involved in the early proinflammatory response to intracellular pathogens were significantly induced, including IL-1β, intracellular adhesion molecule-1 (ICAM-1), and macrophage inflammatory protein-1β (MIP-1β). Moreover, an interesting contrast was noted between the ability of the two microbes to induce host genes. Thus, *Salmonella* induced apoptosis-promoting genes whereas anti-apoptosis genes were modulated by *L. monocytogenes*. These studies typify the power of expression profiling in detecting different virulence strategies that microbes employ in host–pathogen interactions at the molecular and cellular levels.

Macrophages are important cells of the host immune system and play an important role in dictating the quantity and quality of immunity to microbial pathogens. Thus, gene expression profiling of macrophages has been used to characterise host immune responses. For example, RNA samples from an established macrophage cell line (HD11) infected with the intracellular pathogens *Salmonella* or *Eimeria* were used to investigate the underlying mechanisms of host innate immunity against these micro-organisms using a microarray containing approximately 5,000 macrophage ESTs (55). Analysis of the transcriptional profiles of HD11 cells infected with *S. enteritidis* at 2 h, 5 h, and 24 h identified 338 genes that exhibited at least 2-fold increased or decreased expression. Among these genes, the chemokine ah294 consistently showed highest expression at all time points examined; ah294 is a CC chemokine that activates innate immune responses and prevents the apoptosis of virus-infected macrophages. Other immune-related genes with enhanced gene expression following *Salmonella* infection included immune-responsive protein 1, interleukin-6 (IL-6), inducible T-cell costimulator, anti-apoptotic NR13, matrix metalloproteinase-9 (MMP-9), and glutamate-cysteine ligase (GCLM). By contrast, genes associated with cell adhesion and cell proliferation were downregulated following *S. enteritidis* infection. In the case of *Eimeria*-infected HD11 cells, upregulated expression of several important immune effector genes was reported, including the proinflammatory cytokine IL-1β, the chemokines ah221 and MIP-1β, and osteopontin. Interleukin-1β is secreted by macrophages and other cells upon activation by a variety of different stimuli and, in turn, induces the expression of other chemokines, thereby amplifying the immune response. Among these, MIP-1β and K203 belong to the CC chemokine family and are involved in the recruitment of macrophages to sites of infection.

In a related study, gene expression analysis with the macrophage microarray was used to characterise the innate immunity of three antigenically distinct species of *Eimeria*, namely, *E. acervulina*, *E. maxima*, and *E. tenella* (22). All of these species of *Eimeria* elicited similar gene expression response profiles, characterised by pronounced induction of many common genes involved in innate immunity. In particular, a set of 25 core response genes was identified. In addition, 60-67 genes that were unique to the individual *Eimeria* species were induced or repressed. In summary, while a shared similarity in transcript quality existed among the three *Eimeria* micro-organisms, differences were evident in the magnitude, direction, and timing of the immune responses to each individual species.

Another example is that of Marek's disease virus (MDV), a herpesvirus which causes T-cell lymphoma by infecting $CD4^+$ T cells and inducing immunosuppression. Herpesvirus of turkeys (HVT) has been successfully used as a vaccine to prevent Marek's disease in chickens. To investigate the mechanism of this protective response, expression gene analysis of chicken embryonic fibroblasts infected with the HVT vaccine was performed (49). Transcript levels upregulated by HVT included those encoded by genes known to be induced by interferon, as well as others modulating protein kinases and scaffolding proteins of signal transduction cascades. Many of these genes are known to function at critical steps in the host protective response to viral infection. In summary, all of the studies mentioned above are illustrative of the power of using new high-throughput molecular/genetic tools to investigate complex interactions between hosts and pathogens.

In vivo studies

Influenza A/Texas/36/91 virus causes a human-like influenza syndrome in pigtailed macaques and this animal model has been successfully used to study influenza virus infection at the genetic level (1). Transcriptional analysis of lung and tracheobronchial lymph nodes of pigtailed macaques infected with a genetically reconstructed strain of human influenza H1N2 A/Texas/36/91 virus was carried out to study host–virus interactions and to compare the antiviral response of macaques and humans. A commercially available human cDNA array (Agilent Technologies, Palo Alto, California, USA) containing duplicate spots of 13,026 unique clones was used in this study. Significant transcriptional activation of inflammatory cells with the activation of interferon, B cell, and apoptotic pathways accompanied by overt clinical signs was observed in the lungs of H1N2-infected macaques, which coincided with gross and histopathological signs of inflammation and tissue damage. The results of this cDNA microarray study provided insights into the molecular and cellular mechanisms associated with local innate immunity to influenza virus which were consistent with clinical signs of disease. Furthermore, gene expression profiling of influenza-infected lungs revealed new views of the role of cytotoxic T cells and natural killer cells in clearing influenza virus from the lung.

Another example of the use of functional genomics studied at the *in vivo* levels concerns avian coccidiosis due to infection of the gut with *Eimeria* parasites. The immune response to *Eimeria* is complex and involves many different types of locally situated intestinal intraepithelial lymphocytes (IELs) (21, 55). Different species of *Eimeria* show preferential invasion of distinct sites in the intestine and induce a species-specific host immune response. Two major species of *Eimeria*, *E. maxima* and *E. acervulina*, preferentially invade and develop in the jejunum and duodenum, respectively. To investigate local host immune responses induced by *Eimeria* infection, global transcriptional changes in IELs induced by oral inoculation of chickens with *E. acervulina* or *E. maxima* were monitored using a cDNA microarray containing 400 unique immune-related genes (63, 64). RNA samples from the jejunum and duodenum were obtained at 4 different time points following primary and secondary infections in order to characterise response kinetics. The results demonstrated that multiple immune-related gene transcripts were significantly upregulated or downregulated following primary or secondary infection with *E. acervulina* or *E. maxima*. In general, infection by either parasite resulted in the altered expression of more genes in naïve hosts than in immune hosts, and *E. acervulina* induced more changes compared with *E. maxima*. On the other hand, similar changes in the levels of several cytokine mRNAs were observed in both *Eimeria* species following primary infection. Also identified was a set of transcripts whose expression was commonly enhanced or repressed in the intestinal IELs of chickens infected with either parasite.

A third example of the use of gene microarrays to study host–pathogen interactions *in vivo* involves MDV. Liu *et al.* (56) used expression profiling to investigate the underlying genetic basis for disease resistance to MDV using two genetically disparate avian hosts. Transcriptional differences seen between two inbred chicken lines (lines 6 and 7), which were MDV resistant and susceptible respectively, provided insights for these investigators into the mechanisms of disease resistance. Furthermore, the nature of host proteins that interacted with specific MDV proteins was identified using a supplementary approach based on a yeast two-hybrid assay. Specifically, the growth hormone gene (GH1) was identified as a candidate gene associated with MDV resistance and further studies indicated that GH1 variation correlated with a number of Marek's disease-associated traits.

The long-term goals of using functional genomics and microarray technologies in infectious disease studies include obtaining a detailed molecular understanding of host–pathogen interactions and identifying critical target molecules and pathways for better diagnosis and design of preventive measures. Importantly, applying an integrated systems biology approach using diverse techniques such as immunome, proteome, *in vitro* and *in vivo* transcriptome analyses, comparative genomics analyses and bioinformatics analyses (as described by Musser and DeLeo [67]) will yield new insights into microbial pathogenesis and the host response. This will enable the identification of potential candidate vaccine and therapeutic targets more quickly and efficiently than otherwise possible by conventional approaches. The final section of this article provides a peek into the future application of genomics for selecting good responders to vaccination.

Selection of good responders to vaccination

An important application of animal genomics will be the evaluation of genetic influences on individual animal responses to vaccination. Veterinarians involved in vaccine clinical trials have long observed disparity in the response of individual animals to infection and vaccination in well characterised animal challenge models. These empirical observations have highlighted the need for sound biometric analyses as well as robust regulatory standards such as good laboratory practices (GLP) and good clinical practices (GCP) to eliminate experimental and

environmental biases in clinical trials. With the elimination of these biases remains the effect of host genetics on the actual safety and efficacy profile of vaccines in various livestock and poultry animal populations.

Most immunogenetics studies in livestock and poultry species have focused on disease resistance (a good review of these efforts can be found in a previous issue of the OIE *Scientific and Technical Review* [43]). Scientific studies providing evidence that an individual animal's genotype may predetermine immunological responses to vaccination are more limited. One landmark study by Newman *et al.* demonstrated in large half-sibling families differences in antibody responses induced in cattle by vaccination with *B. abortus* Strain 19, a live attenuated bacterial vaccine (70). The data were analysed using a parametric statistical model that incorporated the effects of sire, bovine major histocompatibility complex (BoLA) types, and parameters related to experimental design. Variation between individual animals was not only significant but the study also identified individual animals and families with high or low antibody production phenotypes. In several cases, these traits were significantly correlated with individual bulls, suggesting the existence of sire effects, or individual BoLA types.

Elizabeth Glass at the Roslin Institute in the United Kingdom reported that BoLA haplotypes are associated with FMDV-specific T-cell and antibody responses (37). In a fully pedigreed cattle population genotyped with 186 microsatellite markers derived from a cross between two extremes of cattle, Holtstein dairy and Charolais beef cattle, a first cohort of females immunised with a 40-mer FMDV peptide in Freund's incomplete adjuvant demonstrated a wide variation of immune responses ranging from complete non-responders to very high responders. Of all the immune responses measured, significant sire effects were seen for INF-γ, IgG_2, and IgG_1:IgG_2 ratio, suggesting that genetic influences other than the MHC genes may be regulating host responses to the FMDV peptide.

In another study (72), a commercial bovine respiratory syncytial virus (BRSV) vaccine was tested in the same Holstein-Charolais crossbred study population described above. BRSV-specific IgG antibody responses associated with protection were measured by ELISA. The analysis included the separation of heritable factors (e.g. breed-cross and sire effects) from non-heritable factors (e.g. year of birth, age and sex effects) to quantify their respective contributions to the variation in antibody response. Although this study could not determine any breed differences between Holstein and Charolais calves, the results established a significant calf–sire heritable influence on BRSV-specific IgG antibody levels.

These studies suggest that the heritability of complex traits such as vaccine responsiveness is polygenic and unlikely to be under the control of a single gene. When considering the complexities of host–pathogen interactions, it is expected that the many genes that control vaccine responses will be highly variable and individual genes will potentially display polymorphisms that collectively will determine the level of vaccine responsiveness in individual animals.

Conclusion

Genomic-based approaches are driving fundamental changes in our understanding of microbiology. Comparative analysis of microbial strains is providing new insights into pathogen evolution, virulence mechanisms, and host range specificity. Most importantly, gene discovery and genetic variations can now be used in genotyping analyses and the rational design of vaccines.

New research strategies employing high-throughput gene expression analysis are providing novel platforms for more comprehensive understanding of host–pathogen interactions. In particular, functional genomics is rapidly revolutionising the analysis of whole genome responses of host and pathogens, which will ultimately lead to a better understanding of disease processes and the mechanisms through which pathogens evade host immunity; identification of the genetic basis of host–pathogen interactions; and discovery of novel vaccines, drugs, and biotherapeutics.

Ultimately, we will be able to monitor the two way conversation between hosts and pathogens with the rapidly developing public database of the completely annotated genomic sequence datasets of many hosts and pathogens, the use of sequence-based high-throughput expression profiling technologies, and integrated bioinformatic tools to analyse and interpret genomic data. Through these multiple and combined approaches, we will obtain a complete picture of infectious diseases, microbial pathogenesis and protective host immune mechanisms using an integrated systems biology that will be crucial in developing a new generation of intervention strategies against pathogens infecting humans and animals. Microarray-based technologies for studying genome-wide transcriptional profiling hold exceptional promise for infectious diseases studies, since transcriptional control plays a key role in host–pathogen interactions. Rapidly advancing microarray technology platforms (expression profiling) will allow greater flexibility by providing this technology with increasing array element densities, better detection sensitivities, and more highly cost-efficient protocols. Future challenges for microarray researchers

will include developing databases and algorithms to manage and analyse the vast genomic-scale datasets and extracting meaningful biological information from them.

Vaccinogenomics, the integration of pathogen and host genomics in vaccine research, is likely to revolutionise the way scientists approach the challenges of discovering safe and effective vaccines. The availability of the genomics tools described in this review provides unprecedented opportunities for the rational design of highly effective veterinary vaccines. Identifying genes and genetic variances that control mechanisms of immune evasion, disease resistance, and vaccine responsiveness will in the future fundamentally change vaccine discovery research and enable vaccinologists to design vaccines to control and eradicate pathogens in targeted animal populations. For example, the use of chicken lines with defined genetic backgrounds in modern production systems provides unique opportunities for applying vaccinogenomic approaches to enable the development of vaccines that perform consistently under field conditions. Paradoxically, the heterogeneity found in outbred livestock populations may also present opportunities for vaccinogenomics by enabling marker-assisted selection of good responders to vaccination. Ultimately, genetic markers of protective immunity may one day lead to practical applications in selective breeding programmes to significantly increase disease resistance in farmed livestock and poultry, thereby improving animal welfare and the safety of our food supply.

La génomique et la mise au point de vaccins

C.G. Gay, R. Zuerner, J.P. Bannantine, H.S. Lillehoj, J.J Zhu, R. Green & P.-P. Pastoret

Résumé
Grâce au développement spectaculaire des nouvelles technologies à haut-débit dérivées de l'étude des génomes microbiens et animaux, il est devenu possible d'analyser le génome, le transcriptome et le protéome, ce qui ouvre de nouvelles perspectives pour mieux comprendre les processus moléculaires à l'œuvre dans la biologie des agents pathogènes, dans le système immunitaire de l'hôte et dans les interactions hôte-agent pathogène. L'application de ces nouveaux outils au domaine vétérinaire devrait permettre de surmonter certains obstacles parmi ceux qui freinent encore la mise au point de vaccins performants destinés au bétail et aux volailles.

Mots-clés
Génomique animale – Génomique microbienne – Immunogénomique – Microdamier – Vaccin – Vaccinogénomique.

Genómica y desarrollo de vacunas

C.G. Gay, R. Zuerner, J.P. Bannantine, H.S. Lillehoj, J.J Zhu, R. Green & P.-P. Pastoret

Resumen
Gracias al asombroso desarrollo de las nuevas tecnologías de alto potencial derivadas del estudio de la genómica microbiana y animal, actualmente se está analizando el genoma, el transcriptoma y el proteoma. Esos estudios posibilitarán una mejor comprensión de las vías moleculares de la biología de los agentes patógenos, el sistema inmunitario de los huéspedes y las interacciones entre huéspedes y patógenos. La aplicación de esas nuevas herramientas a los agentes patógenos de los animales debería permitir superar algunos de los obstáculos actuales al descubrimiento de vacunas eficaces para el ganado y las aves de corral de criadero.

Palabras clave
Genómica animal – Genómica aplicada a las vacunas – Genómica de la inmunidad – Genómica microbiana – Micromatrices – Vacunas.

References

1. Baskin C.R., Garcia-Sastre A., Tumpey T.M., Bielefeldt-Ohmann H., Carter V.S., Nistal-Villan E. & Katze M.G. (2004). – Integration of clinical data, pathology, and cDNA microarrays in influenza virus-infected pigtailed macaques (*Macaca nemestrina*). *J. Virol.*, **78** (19), 10420-10432.

2. Blake D.P., Hesketh P., Archer A., Carroll F., Smith A.L. & Shirley M.W. (2004). – Parasite genetics and the immune host: recombination between antigenic types of *Eimeria maxima* as an entree to the identification of protective antigens. *Molec. biochem. Parasitol.*, **138** (1), 143-152.

3. Blake D.P., Shirley M.W. & Smith A.L. (2006). – Genetic identification of antigens protective against coccidia. *Parasite Immunol.*, **28 (7),** 305-314.

4. Bloom B.R. & Murray C.J. (1992). – Tuberculosis: commentary on a reemergent killer. *Science*, **257** (5073), 1055-1064.

5. Blythe M.J. & Flower D.R. (2005). – Benchmarking B cell epitope prediction: underperformance of existing methods. *Protein Sci.*, **14** (1), 246-248.

6. Boyce J.D., Wilkie I., Harper M., Paustian M.L., Kapur V., & Adler B. (2002). – Genomic scale analysis of *Pasteurella multocida* gene expression during growth within the natural chicken host. *Infect. Immun.*, **70** (12), 6871-6879.

7. Bricker B.J., Ewalt D.R. & Halling S.M. (2003). – Brucella 'HOOF-Prints': strain typing by multi-locus analysis of variable number tandem repeats (VNTRs). *BMC Microbiol.*, **3**, 15.

8. Brosch R., Gordon S.V., Buchrieser C., Pym A.S., Garnier T., and Cole S.T. (2000). – Comparative genomics uncovers large tandem chromosomal duplications in *Mycobacterium bovis* BCG Pasteur. *Yeast*, **17** (2), 111-123.

9. Brown P.O. & Botstein D. (1999). – Exploring the new world of the genome with DNA microarrays. *Nature Genet.*, **21** (1 suppl.), 33-37.

10. Brusic V., Bajic V.B. & Petrovsky N. (2004). – Computational methods for prediction of T-cell epitopes – a framework for modelling, testing, and applications. *Methods*, **34** (4), 436-443.

11. Bulach D.M., Zuerner R.L., Wilson P., Seemann T., McGrath A., Cullen P.A., Davis J., Johnson M., Kuczek E., Alt D.P., Peterson-Burch B., Coppel R.L., Rood J.I., Davies J.K. & Adler B. (2006). – Genome reduction in *Leptospira borgpetersenii* reflects limited transmission potential. *Proc. natl Acad. Sci. USA*, **103** (39), 14560-14565.

12. Caoili S.E. (2006). – A structural-energetic basis for B-cell epitope prediction. *Protein Pept. Lett.*, **13** (7), 743-751.

13. Carrillo C., Tulman E.R., Delhon G., Lu Z., Carreno A., Vagnozzi A., Kutish G.F. & Rock D.L. (2005). – Comparative genomics of foot-and-mouth disease virus. *J. Virol.*, **79** (10), 6487-6504.

14. Chain P.S., Comerci D.J., Tolmasky M.E., Larimer F.W., Malfatti S.A., Vergez L.M., Aguero F., Land M.L., Ugalde R.A. & Garcia E. (2005). – Whole-genome analyses of speciation events in pathogenic *Brucellae*. *Infect. Immun.*, **73** (12), 8353-8361.

15. Cohen P., Bouaboula M., Bellis M., Baron V., Jbilo O., Poinot-Chazel C., Galiegue S., Hadibi E.H. & Casellas P. (2000). – Monitoring cellular responses to *Listeria monocytogenes* with oligonucleotide arrays. *J. biol. Chem.*, **275** (15), 11181-11190.

16. Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. 3rd, Tekaia F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver K., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton J., Squares R., Squares S., Sulston J.E., Taylor K., Whitehead S. & Barrell B.G. (1998). – Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, **393** (6685), 537-544.

17. Collins F.S., Green E.D., Guttmacher A.E., & Guyer M.S. (2003). – A vision for the future of genomics research. *Nature*, **422** (6934), 835-847.

18. Cummings C.A. & Relman D.A. (2000). – Using DNA microarrays to study host-microbe interactions. *Emerg. infect. Dis.*, **6** (5), 513-525.

19. Cunningham M.W. (2000). – Pathogenesis of group A streptococcal infections. *Clin. Microbiol. Rev.*, **13** (3), 470-511.

20. Dahlquist K.D., Salomonis N., Vranizan K., Lawlor S.C. & Conklin B. R. (2002). – GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nature Genet.*, **31** (1), 19-20.

21. Dalloul R.A. & Lillehoj H.S. (2006). – Poultry coccidiosis: recent advancements in control measures and vaccine development. *Expert. Rev. Vaccines*, **5** (1), 143-163.

22. Dalloul R.A., Bliss T.W., Hong Y.H., Ben-Chouikha I., Park D.W., Keeler C.L. & Lillehoj H.S. (2007). – Unique responses of the avian macrophage to different species of *Eimeria*. *Molec. Immunol.*, **44** (4), 558-566.

23. DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujer C., Los T., Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G., Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E., Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.J., Haselkorn R., Kyrpides N. & Overbeek R. (2002). – The genome sequence of the facultative intracellular pathogen *Brucella melitensis*. *Proc. natl Acad. Sci. USA*, **99** (1), 443-448.

24. Doytchinova I.A., Guan P. & Flower D.R. (2004). – Identifiying human MHC supertypes using bioinformatic methods. *J. Immunol.*, **172** (7), 4314-4323.

25. Draghici S. (2003). – Data analysis tools for DNA microarrays. Chapman & Hall/CRC, Boca Raton.

26. Eckmann L., Smith J.R., Housley M.P., Dwinell M.B. & Kagnoff M.F. (2000). – Analysis by high density cDNA arrays of altered gene expression in human intestinal epithelial cells in response to infection with the invasive enteric bacteria *Salmonella*. *J. biol. Chem.*, **275** (19), 14084-14094.

27. Esposito J.J., Sammons S.A., Frace A.M., Osborne J.D., Olsen-Rasmussen M., Zhang M., Govil D., Damon I.K., Kline R., Laker M., Li Y., Smith G.L., Meyer H., Leduc J.W. & Wohlhueter R.M. (2006). – Genome sequence diversity and clues to the evolution of variola (smallpox) virus. *Science*, **313** (5788), 807-812.

28. Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M., *et al.* (1995). – Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, **269** (5223), 496-512.

29. Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W., Jacobs Jr W.R., Venter J.C. & Fraser C.M. (2002). – Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J. Bacteriol.*, **184** (19), 5479-5490.

30. Flower D.R. (2003). – Databases and data mining for computational vaccinology. *Curr. Opin. Drug Discov. Devel.*, **6** (3), 396-400.

31. Gamberini M., Gomez R.M., Atzingen M.V., Martins E.A., Vasconcellos S.A., Romero E.C., Leite L.C., Ho P.L. & Nascimento A.L. (2005). – Whole-genome analysis of *Leptospira interrogans* to identify potential vaccine candidates against leptospirosis. *FEMS Microbiol. Lett.*, **244** (2), 305-313.

32. Gao L.Y., Pak M., Kish R., Kajihara K. & Brown E.J. (2006). – A mycobacterial operon essential for virulence *in vivo* and invasion and intracellular persistence in macrophages. *Infect. Immun.*, **74** (3), 1757-1767.

33. Garnier T., Eiglmeier K., Camus J.C., Medina N., Mansoor H., Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S., Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R., Parkhill J., Barrell B.G., Cole S.T., Gordon S.V. & Hewinson R.G. (2003). – The complete genome sequence of *Mycobacterium bovis*. *Proc. natl Acad. Sci. USA*, **100** (13), 7877-7882.

34. Geysen H.M., Meloen R.H. & Barteling S.J. (1984). – Use of peptide synthesis to probe viral antigens for epitopes to a resolution of a single amino acid. *Proc. natl Acad. Sci. USA*, **81** (13), 3998-4002.

35. Geysen H.M., Barteling S.J. & Meloen R.H. (1985). – Small peptides induce antibodies with a sequence and structural requirement for binding antigen comparable to antibodies raised against the native protein. *Proc. natl Acad. Sci. USA*, **82** (1), 178-182.

36. Gibbs R.A., Weinstock G., Kappes S.M., Schook L.B., Skow L. & Womack J. (2002). – Bovine genomic sequencing initiative: De-humanizing the cattle genome. Available at:http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/BovSeq.pdf.

37. Glass E.J. (2004). – Genetic variation and responses to vaccines. *Anim. Hlth Res. Rev.*, **5** (2), 197-208.

38. Green R.D., Qureshi M.A. & Long J.A. (2005). – Charting the road map for long-term USDA efforts in agricultural animal genomics: summary of the USDA Animal Genomics Workshop, September 2004. USDA Mimeo (18 pp.).

39. Gulukota K. & DeLisi C. (1996). – HLA allele selection for designing peptide vaccines. *Genet. Anal.*, **13** (3), 81-86.

40. Halling S.M., Peterson-Burch B.D., Bricker B.J., Zuerner R.L., Qing Z., Li L.L., Kapur V., Alt D.P. & Olsen S.C. (2005). – Completion of the genome sequence of *Brucella abortus* and comparison to the highly similar genomes of *Brucella melitensis* and *Brucella suis*. *J. Bacteriol.*, **187** (8), 2715-2726.

41. Hans D., Young P.R. & Fairlie D.P. (2006). – Current status of short synthetic peptides as vaccines. *Med. Chem.*, **2** (6), 627-646.

42. Haste Andersen P., Nielsen M. & Lund O. (2006). – Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci.*, **15** (11), 2558-2567.

43. Hawken R.J., Beattie C.W. & Schook L.B. (1998). – Resolving the genetics of resistance to infectious diseases. *In* Genetic resistance to animal diseases. *Rev. sci. tech. Off. int. Epiz.*, **17** (1), 17-25.

44. Huntley J.F., Stabel J.R., Paustian M.L., Reinhardt T.A. & Bannantine J.P. (2005). – Expression library immunization confers protection against *Mycobacterium avium* subsp. *paratuberculosis* infection. *Infect. Immun.*, **73** (10), 6877-6884.

45. International Chicken Genome Consortium (2004). – Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, **432**, 695-716.

46. International HapMap Consortium (2005). – A haplotype map of the human genome. *Nature*, **437**, 1299-1320.

47. Johnston S.A. & Barry M.A. (1997). – Genetic to genomic vaccination. *Vaccine*, **15** (8), 808-809.

48. Kappes S.M., Green R.D. & Van Tassell C.P. (2006). – Sequencing the bovine genome and implications for future research. *In* Proc. 8th World Congress on Genetics Applied to Livestock Production. Belo Horizonte, Brazil.

49. Karaca G., Anobile J., Downs D., Burnside J. & Schmidt C.J. (2004). – Herpesvirus of turkeys: microarray analysis of host gene responses to infection. *Virology*, **318** (1), 102-111.

50. Korber B., LaBute M. & Yusim K. (2006). – Immunoinformatics comes of age. *PLoS comput. Biol.*, **2** (6), e71.

51. Kroemer G., Guillemot F. & Auffray C. (1990). – Genetic organization of the chicken MHC. *Immunol. Res.*, **9** (1), 8-19.

52. Kumanovics A., Takada T. & Lindahl K.F. (2003). – Genomic organization of the mammalian MHC. *Annu. Rev. Immunol.*, **21**, 629-657.

53. Le Fleche P., Jacques I., Grayon M., Al Dahouk S., Bouchon P., Denoeud F., Nockler K., Neubauer H., Guilloteau L.A. & Vergnaud G. (2006). – Evaluation and selection of tandem repeat loci for a *Brucella* MLVA typing assay. *BMC Microbiol.*, **6**, 9.

54. Li L., Bannantine J.P., Zhang Q., Amonsin A., May B.J., Alt D., Banerji N., Kanjilal S. & Kapur V. (2005). – The complete genome sequence of *Mycobacterium avium* subspecies *paratuberculosis*. *Proc. natl Acad. Sci. USA*, **102** (35), 12344-12349.

55. Lillehoj H.S., Kim C.H., Keeler Jr. C.L. & Zhang S. (2007). – Immunogenomic approaches to study host immunity to enteric pathogens. *Poult. Sci.* (in press).

56. Liu H.C., Cheng H.H., Tirunagaru V., Sofer L. & Burnside J. (2001). – A strategy to identify positional candidate genes conferring Marek's disease resistance by integrating DNA microarrays and genetic mapping. *Anim. Genet.*, **32** (6), 351-359.

57. Lo M., Bulach D.M., Powell D.R., Haake D.A., Matsunaga J., Paustian M.L., Zuerner R.L. & Adler B. (2006). – Effects of temperature on gene expression patterns in *Leptospira interrogans* serovar Lai as assessed by whole-genome microarrays. *Infect. Immun.*, **74** (10), 5848-5859.

58. McSparron H., Blythe M.J., Zygouri C., Doytchinova I.A. & Flower D.R. (2003). – JenPep: a novel computational information resource for immunobiology and vaccinology. *J. chem. Inf. Comput. Sci.*, **43** (4), 1276-1287.

59. Majed Z., Bellenger E., Postic D., Pourcel C., Baranton G. & Picardeau M. (2005). – Identification of variable-number tandem-repeat loci in *Leptospira interrogans* sensu stricto. *J. clin. Microbiol.*, **43** (2), 539-545.

60. Margulies M., Egholm M., Altman W.E., Attiya S., Bader J.S., Bemben L.A., Berka J., Braverman M.S. Chen Y.J., Chen Z., Dewell S.B., Du L., Fierro J.M., Gomes X.V., Godwin B.C., He W., Helgesen S., Ho C.H., Irzyk G.P., Jando S.C., Alenquer M.L., Jarvie T.P., Jirage K.B., Kim J.B., Knight J.R., Lanza J.R., Leamon J.H., Lefkowitz S.M., Lei M., Li J., Lohman K.L., Lu H., Makhijani V.B., McDade K.E., McKenna M.P., Myers E.W., Nickerson E., Nobile J.R., Plant R., Puc B.P., Ronan M.T., Roth G.T., Sarkis G.J., Simons J.F., Simpson J.W., Srinivasan M., Tartaro K.R.,

Tomasz A., Vogt K.A., Volkmer G.A., Wang S.H., Wang Y., Weiner M.P., Yu P., Begley R.F. & Rothberg J.M. (2005). – Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437** (7057), 376-380.

61. Marsh I.B., Bannantine J.P., Paustian M.L., Tizard M.L., Kapur V. & Whittington R.J. (2006). – Genomic comparison of *Mycobacterium avium* subsp. *paratuberculosis* sheep and cattle strains by microarray hybridization. *J. Bacteriol.*, **188** (6), 2290-2293.

62. May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S. & Kapur V. (2001). – Complete genomic sequence of *Pasteurella multocida*, Pm70. *Proc. natl Acad. Sci. USA*, **98** (6), 3460-3465.

63. Min W., Lillehoj H.S., Kim S., Zhu J.J., Beard H., Alkharouf N. & Matthews B.F. (2003). – Profiling local gene expression changes associated with *Eimeria maxima* and *Eimeria acervulina* using cDNA microarray. *Appl. Microbiol. Biotechnol.*, **62** (4), 392-399.

64. Min W., Lillehoj H.S., Ashwell C.M., van Tassell C.P., Dalloul R.A., Matukumalli L.K., Han J.Y. & Lillehoj E.P. (2005). – Expressed sequence tag analysis of *Eimeria*-stimulated intestinal intraepithelial lymphocytes in chickens. *Molec. Biotechnol.*, **30** (2), 143-150.

65. Moreau V., Granier C., Villard S., Laune D. & Molina F. (2006). – Discontinuous epitope prediction based on mimotope analysis. *Bioinformatics*, **22** (9), 1088-1095.

66. Mostowy S., Tsolaki A.G., Small P.M. & Behr M.A. (2003). – The *in vitro* evolution of BCG vaccines. *Vaccine*, **21** (37-30), 4270-4274.

67. Musser J.M. & DeLeo F.R. (2005). – Toward a genome-wide systems biology analysis of host–pathogen interactions in group A *Streptococcus*. *Am. J. Pathol.*, **167** (6), 1461-1472.

68. Nagy G., Danino V., Dobrindt U., Pallen M., Chaudhuri R., Emody L., Hinton J.C. & Hacker J. (2006). – Down-regulation of key virulence factors makes the *Salmonella enterica* serovar Typhimurium rfaH mutant a promising live-attenuated vaccine candidate. *Infect. Immun.*, **74** (10), 5914-5925.

69. Nascimento A.L., Ko A.I., Martins E.A., Monteiro-Vitorello C.B., Ho P.L., Haake D.A., Verjovski-Almeida S., Hartskeerl R.A., Marques M.V., Oliveira M.C., Menck C.F., Leite L.C., Carrer H., Coutinho L.L., Degrave W.M., Dellagostin O.A., El-Dorry H., Ferro E.S., Ferro M.I., Furlan L.R., Gamberini M., Giglioti E.A., Goes-Neto A., Goldman G.H., Goldman M.H., Harakava R., Jeronimo S.M., Junqueira-de-Azevedo I.L., Kimura E.T., Kuramae E.E., Lemos E.G., Lemos M.V., Marino C.L., Nunes L.R., de Oliveira R.C., Pereira G.G., Reis M.S., Schriefer A., Siqueira W.J., Sommer P., Tsai S.M., Simpson A.J., Ferro J.A., Camargo L.E., Kitajima J.P., Setubal J.C. & Van Sluys M.A. (2004). – Comparative genomics of two *Leptospira interrogans* serovars reveals novel insights into physiology and pathogenesis. *J. Bacteriol.*, **186** (7), 2164-2172.

70. Newman M.J., Truax R.E., French D.D., Dietrich M.A., Franke D. & Stear M.J. (1996). – Evidence for genetic control of vaccine-induced antibody responses in cattle. *Vet. Immunol. Immunopathol.*, **50** (1-2), 43-54.

71. Nielsen M., Lundegaard C., Worning P., Hvid C.S., Lamberth K., Buus S., Brunak S. & Lund O. (2004). – Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. *Bioinformatics*, **20** (9), 1388-1397.

72. O'Neill R.G., Woolliams J.A., Glass E.J., Williams J.L. & Fitzpatrick J.L. (2006). – Quantitative evaluation of genetic and environmental parameters determining antibody response induced by vaccination against bovine respiratory syncytial virus. *Vaccine*, **24** (18), 4007-4016.

73. Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N., Harris D.E., Holden M.T., Churcher C.M., Bentley S.D., Mungall K.L., Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A., Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I., Chillingworth T., Collins M., Cronin A., Davis P., Doggett J., Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K., Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C., Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K., Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K., Unwin L., Whitehead S., Barrell B.G. & Maskell D.J. (2003). – Comparative analysis of the genome sequences of *Bordetella pertussis, Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat. Genet.*, **35** (1), 32-40.

74. Partidos C.D. & Steward M.W. (2002). – Mimotopes of viral antigens and biologically important molecules as candidate vaccines and potential immunotherapeutics. *Comb. Chem. high Throughput Screen.*, **5** (1), 15-27.

75. Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F., Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J., Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R., Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E., Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L., Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M. & Fraser C.M. (2002). – The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts. *Proc. natl Acad. Sci. USA*, **99** (20), 13148-13153.

76. Paustian M.L., Kapur V. & Bannantine J.P. (2005).– Comparative genomic hybridizations reveal genetic regions within the *Mycobacterium avium* complex that are divergent from *Mycobacterium avium* subsp. *paratuberculosis* isolates. *J. Bacteriol.*, **187** (7), 2406-2415.

77. Pizza M., Scarlato V., Masignani V., Giuliani M.M., Arico B., Comanducci M., Jennings G.T., Baldi L. Bartolini E., Capecchi B., Galeotti C.L., Luzzi E., Manetti R., Marchetti E., Mora M., Nuti S., Ratti G., Santini L., Savino S., Scarselli M., Storni E., Zuo P., Broeker M., Hundt E., Knapp B., Blair E., Mason T., Tettelin H., Hood D.W., Jeffries A.C., Saunders N.J., Granoff D.M., Venter J.C., Moxon E.R., Grandi G. & Rappuoli R. (2000). – Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science*, **287** (5459), 1816-1820.

78. Pizzi E., Cortese R. & Tramontano A. (1995). – Mapping epitopes on protein surfaces. *Biopolymers*, **36** (5), 675-680.

79. Pym A.S., Brodin P., Brosch R., Huerre M. & Cole S.T. (2002). – Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. *Molec. Microbiol.*, **46** (3), 709-717.

80. Quackenbush J. (2002). – Microarray data normalization and transformation. *Nature Genet.*, **32** (Suppl.), 496-501.

81. Rappuoli R. (2000). – Reverse vaccinology. *Curr. Op. Microbiol.*, **3** (5), 445-450.

82. Rappuoli R. (2001). – Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine*, **19** (17-19), 2688-2691.

83. Read T.D., Salzberg S.L., Pop M., Shumway M., Umayam L., Jiang L., Holtzapple, E., Busch, J.D. *et al.* (2002). – Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science*, **296** (5575), 2028-2033.

84. Reche P.A. & Reinherz E.L. (2003). – Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. *J. molec. Biol.*, **331** (3), 623-641.

85. Reche P.A. & Reinherz E.L. (2005). – PEPVAC: a web server for multi-epitope vaccine development based on the prediction of supertypic MHC ligands. *Nucleic Acids Res.*, **33** (web server issue), W138-142.

86. Reid S.D., Green N.M., Buss J.K., Lei B. & Musser J.M. (2001). – Multilocus analysis of extracellular putative virulence proteins made by group A *Streptococcus*: population genetics, human serologic response, and gene transcription. *Proc. natl Acad. Sci. USA*, **98** (13), 7552-7557.

87. Ren S.X., Fu G., Jiang X.G., Zeng R., Miao Y.G., Xu H., Zhang Y.X., Xiong H., Lu G., Lu L.F., Jiang H.Q., Jia J., Tu Y.F., Jiang J.X., Gu W.Y., Zhang Y.Q., Cai Z., Sheng H.H., Yin H.F., Zhang Y., Zhu G.F., Wan M., Huang H.L., Qian Z., Wang S.Y., Ma W., Yao Z.J., Shen Y., Qiang B.Q., Xia Q.C., Guo X.K., Danchin A., Saint Girons I., Somerville R.L., Wen Y.M., Shi M.H., Chen Z., Xu J.G. & Zhao G.P. (2003). – Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. *Nature*, **422** (6934), 888-893.

88. Rosinha G.M., Freitas D.A., Miyoshi A., Azevedo V., Campos E., Cravero S.L., Rossetti O., Splitter G. & Oliveira S.C. (2002). – Identification and characterization of a *Brucella abortus* ATP-binding cassette transporter homolog to *Rhizobium meliloti* ExsA and its role in virulence and protection in mice. *Infect. Immun.*, **70** (9), 5036-5044.

89. Saha S. & Raghava G.P. (2006). – Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*, **65** (1), 40-48.

90. Salaun L., Merien F., Gurianova S., Baranton G. & Picardeau M. (2006). – Application of multilocus variable-number tandem-repeat analysis for molecular typing of the agent of leptospirosis. *J. clin. Microbiol.*, **44** (11), 3954-3962.

91. Scarselli M., Giuliani M.M., Adu-Bobie J., Pizza M. & Rappuoli R. (2005). – The impact of genomics on vaccine design. *Trends Biotechnol.*, **23** (2), 84-91.

92. Sebaihia M., Preston A., Maskell D.J., Kuzmiak H., Connell T.D., King N.D., Orndorff P.E., Miyamoto D.M. *et al.* (2006). – Comparison of the genome sequence of the poultry pathogen *Bordetella avium* with those of *B. bronchiseptica*, *B. pertussis*, and *B. parapertussis* reveals extensive diversity in surface structures associated with host interaction. *J. Bacteriol.*, **188** (16), 6002-6015.

93. Serruto D. & Rappuoli R. (2006). – Post-genomic vaccine development. *FEBS Lett.*, **580** (12), 2985-2992.

94. Setubal J.C., Reis M., Matsunaga J. & Haake D.A. (2006). – Lipoprotein computational prediction in spirochaetal genomes. *Microbiology*, **152** (Pt 1), 113-121.

95. Sidney J., del Guercio M.F., Southwood S., Engelhard V.H., Appella E., Rammensee H.G., Falk K., Rotzschke O., Takiguchi M., Kubo R.T. *et al.* (1995). – Several HLA alleles share overlapping peptide specificities. *J. Immunol.*, **154** (1), 247-259.

96. Slack A.T., Dohnt M.F., Symonds M.L. & Smythe L.D. (2005). – Development of a Multiple-Locus Variable Number of Tandem Repeat Analysis (MLVA) for *Leptospira interrogans* and its application to *Leptospira interrogans* serovar Australis isolates from Far North Queensland, Australia. *Ann. clin. Microbiol. Antimicrob.*, **4** (10), 10.

97. Slack A.T., Symonds M., Dohnt M. & Smythe L. (2006). – An improved multiple-locus variable number of tandem repeats analysis for *Leptospira interrogans* serovar Australis: a comparison with fluorescent amplified fragment length polymorphism analysis and its use to redefine the molecular epidemiology of this serovar in Queensland, Australia. *J. med. Microbiol.*, **55** (Pt II), 1549-1557.

98. Sollner J. (2006). – Selection and combination of machine learning classifiers for prediction of linear B-cell epitopes on proteins. *J. mol. Recognit.*, **19** (3), 209-214.

99. Sollner J. & Mayer B. (2006). – Machine learning approaches for prediction of linear B-cell epitopes on proteins. *J. mol. Recognit.*, **19** (3), 200-208.

100. Sorensen A.L., Nagai S., Houen G., Andersen P. & Andersen A.B. (1995). – Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. *Infect. Immun.*, **63** (5), 1710-1717.

101. Southern E., Mir K. & Shchepinov M. (1999). – Molecular interactions on microarrays. *Nature Genet.*, **21** (1 suppl.), 5-9.

102. Stoughton R.B. (2005). – Applications of DNA microarrays in biology. *Annu. Rev. Biochem.*, **74**, 53-82.

103. Sumby P., Whitney A.R., Graviss E.A., DeLeo F.R. & Musser J.M. (2006). – Genome-wide analysis of group a streptococci reveals a mutation that modulates global phenotype and disease specificity. *PLoS Pathog.*, **2** (1), e5.

104. Welsch J.A., Moe G.R., Rossi R., Adu-Bobie J., Rappuoli R. & Granoff D.M. (2003). – Antibody to genome-derived neisserial antigen 2132, a *Neisseria meningitidis* candidate vaccine, confers protection against bacteremia in the absence of complement-mediated bactericidal activity. *J. infect. Dis.*, **188** (11), 1730-1740.

105. Whatmore A.M., Shankster S.J., Perrett L.L., Murphy T.J., Brew S.D., Thirlwall R.E., Cutler S.J. & MacMillan A.P. (2006). – Identification and characterization of variable-number tandem-repeat markers for typing of *Brucella* spp. *J. clin. Microbiol.*, **44** (6), 1982-1993.

106. Williams R. B. (1999). – A compartmentalised model for the estimation of the cost of coccidiosis to the world's chicken production industry. *Int. J. Parasitol.*, **29** (8), 1209-1229.

107. Wong G.K., Liu B., Wang J., Zhang Y., Yang X., Zhang Z., Meng Q., Zhou J., Li D., Zhang J., Ni P., Li S., Ran L., Li H., Zhang J., Li R., Li S., Zheng H., Lin W., Li G., Wang X., Zhao W., Li J., Ye C., Dai M., Ruan J., Zhou Y., Li Y., He X., Zhang Y., Wang J., Huang X., Tong W., Chen J., Ye J., Chen C., Wei N., Li G., Dong L., Lan F., Sun Y., Zhang Z., Yang Z., Yu Y., Huang Y., He D., Xi Y., Wei D., Qi Q., Li W., Shi J., Wang M., Xie F., Wang J., Zhang X., Wang P., Zhao Y., Li N., Yang N., Dong W., Hu S., Zeng C., Zheng W., Hao B., Hillier L.W., Yang S.P., Warren W.C., Wilson R.K., Brandstrom M., Ellegren H., Crooijmans R.P., van der Poel J.J., Bovenhuis H., Groenen M.A., Ovcharenko I., Gordon L., Stubbs L., Lucas S., Glavina T., Aerts A., Kaiser P., Rothwell L., Young J.R., Rogers S., Walker B.A., van Hateren A., Kaufman J., Bumstead N., Lamont S.J., Zhou H., Hocking P.M., Morrice D., de Koning D.J., Law A., Bartley N., Burt D.W., Hunt H., Cheng H.H., Gunnarsson U., Wahlberg P. *et al.* (2004). – A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature*, **432** (7018), 717-722.

108. Zhang G.L., Petrovsky N., Kwoh C.K., August J.T. & Brusic V. (2006). – PREDTAP: a system for prediction of peptide binding to the human transporter associated with antigen processing. *Immunome Res.*, **2**, 3.

109. Zhu H., Cong J.P., Mamtora G., Gingeras T. & Shenk T. (1998). – Cellular gene expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. *Proc. natl Acad. Sci. USA*, **95** (24), 14470-14475.