Rev. Sci. Tech. Off. Int. Epiz., 2016, **35** (1), 203-213

New perspectives from genomic analyses of bacterial infectious agents

R.J. Goldstone & D.G.E. Smith*

School of Life Sciences, John Muir Building, Heriot–Watt University, Edinburgh, EH14 4AS, United Kingdom
*Corresponding author: david.smith@hw.ac.uk

Summary
Recent advances in the technologies for genomic sequencing and systems for handling and processing sequencing data have transformed bacterial genomics into a near-routine approach for both small- and large-scale investigations of infectious agents. Nonetheless, the application of genomics – especially larger-scale studies – to animal infectious agents lags behind its application to human pathogens, despite the growing importance of many animal species as food sources. Assiduously conducted genomic studies offer major benefits, not merely by providing a detailed understanding of infectious agents but also through the exploitation of such findings to enable more accurate diagnosis, high-resolution typing and the development of improved interventions. The use of genomics for these and other purposes is likely to grow in future years and it must be anticipated that investigation and characterisation of important animal infectious agents will also gain considerable benefits. Using mainly animal pathogens as examples – including several infectious agents listed by the World Organisation for Animal Health – this paper provides a concise summary of some recent purposes and developments in bacterial genomics analysis.

Keywords
Bacterial infectious agent – Genome-wide association study – Metagenomics – OIE-listed infectious agent – Reverse vaccinology – Single nucleotide polymorphism – Vaccinology – Whole-genome.

Introduction

As frequently stated, advances in genome sequencing capabilities over recent years have accelerated to the extent that the sequencing process itself has become essentially routine, capable of generating high-quality data in hours, with further developments under way to sequence and process data 'in real time'. It could be considered that the biological and analytical aspects of interpreting genome sequencing data now cause greater bottlenecks in the effective use of this rapidly growing resource. Whilst accomplished, user-friendly analysis tools – such as the CLC genomics workbench (CLC Ltd, Aarhus, Denmark) – are available from commercial sources, numerous tools are also downloadable or freely accessible online, such as the National Center for Biotechnology Information (NCBI) genomics workbench or the Galaxy webserver. Moreover, increasing capability in bioinformatics is resulting in many in-house or custom tools being developed and made publicly available. As of January 2016, there were nearly 60,000 prokaryotic genomes publicly available via GenBank and the European Bioinformatics Institute (EBI), with numbers increasing continually. Not surprisingly, this resource represents a vast diversity of bacteria in respect of their phylogeny, gene content and habitat. Yet, despite this wealth of data, animal infectious agents are relatively poorly represented unless they are also zoonotic and/or of biodefence concern (summarised in Table I). With the growing world population, increasing demand for the high-quality protein provided by livestock and the ever-present threat of enzootic and epizootic diseases, it is highly surprising that little attention has been given to applying high-throughput sequencing and population genomics to significant animal infectious agents. It is tempting to postulate that the widespread use of biosecurity measures, antibiotics and/or vaccines to limit the impact of some major animal diseases on production systems has limited the search for improved understanding of animal infectious agents. However, rising global concern about the increase and spread of antibiotic resistance and infectious diseases is likely to drive the need for improved understanding of

doi: http://dx.doi.org/10.20506/rst.35.1.2428

Table I
Summary of genome resources for some of the bacterial infectious agents listed by the World Organisation for Animal Health

Disease	Organism(s)	No. GenBank genomes [a]	Species affected
Anthrax	*Bacillus anthracis*	97	Multi-species
Brucellosis	*Brucella abortus*	152	Multi-species
	Brucella melitensis	63	Multi-species
	Brucella suis	54	Multi-species
Heartwater	*Ehrlichia ruminantium*	3	Multi-species
Paratuberculosis (Johne's disease)	*Mycobacterium avium* subspecies *paratuberculosis*	30	Multi-species
Q fever	*Coxiella burnettii*	24	Multi-species
Tularemia	*Francisella tularensis* subspecies *tularensis*	31	Multi-species
Anaplasmosis	*Anaplasma marginale*	14	Cattle
Genital campylobacteriosis	*Campylobacter fetus* subspecies *venerealis*	8	Cattle
Bovine tuberculosis	*Mycobacterium bovis*	42 [b]	Cattle
Haemorrhagic septicaemia	*Pasteurella multocida* serotypes B:2 & E:2	0	Cattle
Contagious bovine pleuropneumonia	*Mycoplasma mycoides* subspecies *mycoides* SC	5	Cattle
Contagious caprine pleuropneumonia	*Mycoplasma capricolum* subspecies *capripneumoniae*	6	Sheep and goats
Enzootic abortion of ewes; ovine chlamydiosis	*Chlamydophila (Chlamydia) abortus*	3	Sheep and goats
Ovine epididymitis	*Brucella ovis*	16	Sheep and goats
Salmonellosis	*Salmonella enterica* subspecies *enterica* serovar Abortusovis	0	Sheep and goats
Contagious equine metritis	*Taylorella equigenitalis*	4	Equines
Glanders	*Burkholderia mallei*	32	Equines
Chlamydiosis	*Chlamydia psittaci*	49	Avian species
Mycoplasmosis	*Mycoplasma gallisepticum*	12	Avian species
Mycoplasmosis	*Mycoplasma synoviae*	3	Avian species
Fowl typhoid	*Salmonella enterica* subspecies *enterica* serovar Gallinarum	7	Avian species
Pullorum disease	*Salmonella enterica* subspecies *enterica* serovar Pullorum	5	Avian species
European foulbrood	*Melissococcus plutonius*	2	Bees
American foulbrood	*Paenibacillus larvae*	4	Bees
Withering syndrome	*Xenohaliotis californiensis*	0	Molluscs
Necrotising hepatopancreatitis (NHP)	NHPB (NHP bacterium; not currently specified)	0	Crustaceans

a) Number of genomes tabulated from the National Center for Biotechnology Information (NCBI) on 8 June 2015. Numbers may include multiple entries for some bacterial strains. Numbers exclude genomes in other collections and which are not yet in the public domain
b) Numbers for *Mycobacterium bovis* include 16 Bacille Calmette-Guerin (BCG) variants from different laboratory origins

major animal infectious agents (including those of animal origin) at the isolate, strain and population levels in the future.

With improvements in technology, the sequencing of multiple isolates or strains is becoming a very cost-effective method of detailed characterisation, not only for bacteria, which can be readily isolated in culture, but also for highly fastidious, host-dependent or obligately intracellular organisms, such as *Chlamydia pecorum* (1), or *Lawsonia intracellularis* (2). Furthermore, the feasibility of gaining a meaningful genome sequence from organisms within complex ecosystems (e.g. 'metagenomics') or even single cells within infected tissues, without the requirement for cultivation, is improving (3). Although such techniques are rarely employed for bacteria of veterinary relevance, metagenomics has recently been used to investigate the population genome in bovine mastitis (4). It is inevitable that such sequencing will result in increased application to other veterinary diseases and become a much more routinely used procedure.

There is a vast literature in which genomics is exploited to aid our understanding of bacterial infectious agents. The

purpose of this short overview is not to address all relevant aspects of the issue of bacterial genomics; rather, the focus will be placed on the selected successful application of genomics to bacterial infectious agents of animals. These applications include:

– defining 'prototypic' or novel strains of aetiological agents

– characterising uncultivable agents

– defining host-specificity factors or virulence factors

– population genomics/genome-wide association studies

– the identification of vaccine candidates

– identifying targets, e.g. single nucleotide polymorphisms (SNPs) for typing, tracking and tracing

– charting the evolution and adaptation of infectious agents through short and long time frames.

The applications cited above are not mutually exclusive – genomics studies often bridge two or more of these categories, as will be apparent in the examples below. Another notable development in genomics is the application of high-throughput approaches to monitor gene expression and the role of genes in infection. This can be termed 'functional genomics' and is exemplified by RNA-sequencing (RNA-Seq) and transposon sequencing (Tn-Seq). However, it will not be considered in this review.

The use of genomics for veterinary infectious agents

As more tools are developed for genome sequencing, in conjunction with further technological developments, we are likely to see the continuation of a move from genomics studies with a single or a few isolates to those targeting tens or hundreds of isolates. Indeed, some centres, such as the Wellcome Trust Sanger Institute in the United Kingdom or the Broad Institute in the United States, focus on higher-throughput analyses and routinely conduct large-scale studies, which are frequently at the forefront of genomic developments. As mentioned above, few studies of purely animal infectious agents have achieved this high volume, which is becoming typical of human, zoonotic or biodefence infectious agents. It should be stressed that single isolate or small-scale genomics studies remain of value in foundational or discovery investigations, although careful interpretation is necessary since any similarities or differences observed may be unique among the small number of isolates being compared, rather than being features which distinguish particular 'strains' or 'types'.

Several infectious agents listed by the World Organisation for Animal Health (OIE) (5) have been subjected to comparative genomics, ranging from a few isolates to tens of isolates. Larger-scale studies, in particular, provide examples of the possibilities of detailed and extensive genomics investigations. An indication is provided in Table I, which summarises OIE-listed infectious agents that have been the subject of comparative genomics analyses. These, and the many excellent examples focusing on human infectious agents, serve as models for outcomes from both small-scale and more extensive genomics studies. In brief, assiduous genome comparisons can be used to define:

– phylogeny and evolution

– virulence factors

– determinants of host selectivity/specificity

– the basis of niche adaptation and adaptability

– targets for intervention (e.g. vaccines; see below).

Defining novel strains

Genome sequencing has proved to be an invaluable resource for the identification and analysis of novel bacterial strains associated with disease. Such investigations may focus on one or a few isolates which are implicated in infection and involve a detailed characterisation of their genomes, confirming putative virulence factors and facilitating an understanding of the disease process. They may also take a broader metagenomic approach and seek to identify the population structure associated with a disease state when the aetiological agent is uncertain.

In the first instance, genomics has been used to investigate the genomes of novel *Escherichia coli* from a range of animal infections, including avian pathogenic *E. coli* (6, 7), bovine metritis (8), bovine mastitis (9), and porcine enterotoxigenic strains (10), amongst others. Such analyses have often found that strains isolated from animal infections and from similar disease states in humans (6, 10) regularly share genomic features, even when these strains are fairly distantly genetically related (8). These investigations of *E. coli* serve to highlight the 'compartmentalisation' of genome content into 'core' and 'accessory' content. The latter typically contains virulence factors. Conversely, sequencing and the functional annotation of genomes of different *E. coli* strains from the same clinical syndrome – specifically, avian colisepticaemia (11) – have also revealed disparate pathways to common virulence capacity. Together, these analyses re-emphasise that pathogenicity is conferred by combination(s) of factors and, consequently, that careful, detailed characterisation of pathogens and other infectious agents is an important step towards disease monitoring and control.

Genome-wide/population characterisation

Concurrent with advances in technology and the reduced cost of genome sequencing, studies have now moved towards employing tens or even hundreds of bacterial isolates from particular diseases in order to gain population-level characterisation of the infectious agent's genomes. The use of large numbers of genome sequences provides statistical power to detect the association between genes and phenotypes, which may be missed in studies that focus on a few isolates. One of the foundational methods for studying genome-wide data at the population level is the genome-wide association (GWA) study design. Originally used in 2005 to test the association between alleles in human genes and age-related macular degeneration (12), this design has been used to investigate the association between alleles and a range of veterinary diseases, such as the susceptibility and resistance of cattle to *Mycobacterium tuberculosis* (13, 14), paratuberculosis (Johne's disease) (15, 16, 17, 18, 19, 20, 21) and mastitis (22), as well as the susceptibility of chickens to *Salmonella* colonisation (23). Such methods are also beginning to be adapted to identify genomic loci correlating with infectivity and pathogenicity from the infectious agent's perspective. For example, in *Haemophilus parasuis,* the causative agent of Glasser's disease in pigs (24), a study using more than 200 genome sequences, alongside metadata concerning the virulence of each isolate, provided a level of resolution which led to the identification of genes associated with pathogenicity. Many of these had been missed in previous studies, which used only a handful of isolates, although any role in disease remains putative until proven functionally. As greater effort is put into obtaining more sequences from livestock and other species, their increased availability – along with their associated metadata – will undoubtedly aid many investigators to thoroughly characterise organisms of interest and exploit their findings for future progress.

Typing, tracking and tracing

Genome sequencing approaches are able to provide exquisite resolution for tracking and tracing outbreaks, particularly when the causative agent is highly clonal and unsuited for traditional typing methods, such as multilocus sequence typing (MLST). For example, by using a gene-by-gene approach, Antwerpen *et al.* were able to identify three independent clones of *Francisella tularensis* as being responsible for outbreaks of lethal tularemia among non-human primates at two German animal facilities (25). Whole-genome sequencing has also been used to attempt to address the controversy surrounding the association between *Mycobacterium bovis* – which causes bovine tuberculosis – and its carriage in badgers. In a study employing the resolution of whole-genome sequencing, Biek *et al.* showed that *M. bovis* persisted on or near farms for several years, even in the absence of detected cattle infections, and that isolates recovered from badgers had between zero and four SNPs from the nearest cattle isolate, providing evidence for recent transmission between the two hosts (26), although not intimating the direction of transmission. More recently, these techniques have also been used to infer the transmission of *M. bovis* from cattle to pet cats (27), which could have implications for onward transmission to companion animal owners.

The high-resolution differentiating capacity of whole-genome sequencing-based typing through SNPs is increasingly being applied in short-term and long-term studies of transmission, as well as in tracking the local and global spread of major infectious agents of humans and of antibiotic-resistant bacterial strains. It is interesting to consider which animal infectious agents might be the object of similar scales of investigation as broad-scope sequencing becomes more routine.

Infectious agent evolution

Read *et al.* performed a genomic analysis of 20 *Chlamydia psittaci* isolates from 14 host species (28), which provided a comprehensive and detailed phylogenetic comparison. It identified, among other features, a subgroup of *C. psittaci*, which were originally isolated from eight distinct host species yet were highly closely related. This provides support for the hypothesis that particular clones of *C. psittaci* are predominant and/or have greater potential for cross-species transmission and epidemic spread. Whilst the vast majority of genome content (911 genes) was highly similar across all *C. psittaci* isolates examined, showing small inter-strain variations, the regionalisation of genomic variation to particular genomic loci was highlighted. These 'plasticity zones' represent 'hotspots' of variability and analysis showed that the widely distributed subgroup of *C. psittaci* possessed several additional genes when compared to related strains. These genes are associated with metabolism and attachment as well as undefined (hypothetical) function, again highlighting the multifactorial nature of pathogenicity, concomitant with the profound influence that small inter-strain differences may have on host specificity and disease.

Wattam *et al.* investigated the evolution of facultatively intracellular infectious agents in the *Brucellae*, including the major pathogenic species *Brucella abortus*, *B. melitensis*, *B. suis*, *B. ovis* and *B. canis*, from the closely related species, *Ochrobactrum* (29). A typical genome of *Brucella* spp. encodes approximately 3,100 genes, many of which are common across the genus. This study revealed that the adoption of a pathogenic lifestyle in the *Brucellae* was associated with

both the loss and acquisition of genes. Alterations within metabolic capabilities were shown to play an important role in the adoption of host niche, with many metabolic systems being discarded or diminished. Significantly, many of these genomic changes occurred through the acquisition of regions of genes with linked functions, which have been termed 'genome islands'. This highlights the significance of the exchange of gene content through horizontal gene transfer within and between bacterial species and genera and its influence on niche occupation and host preference as well as pathogenicity.

Characterisation of unculturable agents

Many organisms listed in Table I can be readily cultured in the laboratory, a feature which undoubtedly aids in their identification and further analysis. However, it has been estimated that only around 1% of bacterial species can be cultured (30, 31, 32). Table I also lists several highly fastidious organisms which require specialised culture capabilities, which are host dependent and require co-culture with host cells, or which are (as yet) unculturable. Hence, sampling of organisms such as these directly from their habitat (i.e. their infected host) becomes a necessity. Notably, in samples of infected tissues taken directly from host animals, bacterial genomic material will be very low in abundance in comparison to DNA from the host, hence approaches to enrich or purify target microorganisms or their DNA from such samples may be necessary.

The group of prokaryotes recently termed 'haemotropic mycoplasmas' (haemoplasmas) serves as an example of genomics providing a highly effective means for detailed characterisation of novel, host-dependent infectious agents in their host habitat. These organisms are pathogens/parasites on many animal species. Recently and controversially reclassified as mycoplasmas, these organisms were found to be closely associated with erythrocytes as a cause of infectious anaemia. Among the increasing number of recognised members of this group, several representative isolates have been genome sequenced directly from infected animal cells (33, 34, 35, 36, 37, 38, 39, 40, 41). These haemoplasmas possess very small genomes, the composition of which confirms their distinct taxonomic position in relation to mycoplasmas. Like other mycoplasmas, the haemoplasmas lack conventional virulence factors and show remarkable genomic diversity despite the genome's minimal content. Prediction and analysis of protein composition – in some cases, complemented with transcriptomics or proteomics (42) – has begun to illuminate the determinants of host adaptation and dependency. Many of the predicted proteins – as much as 80% in some strains – are hypothetical, although some of these are postulated to be involved in the evasion of immune response. As well as providing the 'prototype' for these organisms, it is expected that genomics will continue to offer an avenue for identifying and targeting diagnosis, treatment and prevention strategies, all of which present challenges at present.

In some respects, haemoplasma genomics represents a 'model' for similar developments with other host-dependent organisms, although specific sites of infection will dictate the methods required to isolate organisms and their genomic materials. Several of the bacteria of the less well-studied organisms from the OIE list could be appropriate targets for direct sequencing approaches. For instance, the controversial aetiology of necrotising hepatopancreatitis could be fully defined at the genome level on the proviso that sufficient breadth of sampling were incorporated. This approach to characterisation of the agent(s) would not only provide unparalleled detailed insights into the biology of this listed disease but would also aid in its detection and, perhaps, its control.

Vaccine characterisation and candidate identification

Genomics has shown success in characterising vaccines and vaccine strains in a manner similar to the characterisation of novel strains outlined above. For instance, its application to investigate the underlying mechanisms behind spontaneously attenuated strains of *B. abortus* and *B. melitensis* (which are used for both animal and human vaccines against brucellosis) (42, 43, 44, 45) identified multiple genomic features (including gene variants, genes, regions and reorganisations distinguishing vaccine and pathogenic strains). Arguably, the increasing use of genomics in a pipeline often termed 'reverse vaccinology' offers a more systematic and rational approach towards vaccine candidate identification, which has resulted in at least one instance of a vaccine for *Neisseria meningitides* being licensed for use in humans (46). The usefulness of reverse vaccinology has yet to be fully realised, although its potential is evident from its application to bacterial strains, effective vaccines for which have been difficult to define by other means. Similarly, the application to animal infectious agents is not widespread as yet, although several examples have been published (see Table II). Put simply, this approach combines genomics with other biological disciplines, including bioinformatics and immunology along with infection model(s), as a crucial component in candidate evaluation. Complementing these approaches with proteomics/immunoproteomics (e.g. to identify protein targets of antibody and/or cellular immune responses), transcriptomics (e.g. to determine bacterial genes expressed during infection), or functional

Table II
Selected examples of reverse vaccinology applied to animal infectious agents

Infectious agent	Evidence	Reference
Anaplasma marginale	Candidate prediction	47, 48
Corynebacterium pseudotuberculosis	Candidate prediction	49
Escherichia coli (avian pathogenic – APEC)	Candidate prediction	50
Leptospira spp.	Candidate prediction	51, 52
Mycoplasma hyopneumoniae	Candidate prediction	53, 54
Haemophilus parasuis	Immunogenicity in model species	55
Brucella melitensis	Immunogenicity in target species (goats)	56
Dichelobacter nodosus	Immunogenicity in target species (sheep)	57
Ehrlichia ruminantium	Immunogenicity in target species (sheep)	58
Leptospira interrogans	Efficacy in model species	59
Leptospira borgpetersenii	Efficacy in model species	60
Streptococcus suis	Efficacy in model species	61
Brachyspira hyodysenteriae	Efficacy in target species (swine)	62
Pasteurella multocida	Efficacy in target species (chickens)	63

screening (e.g. Tn-Seq to define gene products involved in colonisation, survival and pathogenicity *in vivo*) offers a means to refine candidate selection. Furthermore, reverse vaccinology has proved scalable from individual through to multiple genomes, although a crucial step is to verify candidate distribution and conservation across a broad range of disease isolates.

Developments such as reverse vaccinology ably demonstrate the great potential of combining 'omics', informatics and other biological disciplines, and the progress made to date may be considered a model for what could be achieved by applying such approaches to the infectious agents of animals more systematically.

Concluding comments

Whole-genome sequencing and associated genomics and functional approaches have multiple applications, and this paper has briefly summarised those many uses. Many impressive studies that apply a range of genomics approaches to bacterial infectious agents have been reported in the literature, principally targeting human infectious agents. Given the increasing availability and relative cost effectiveness of genome sequencing, it can be predicted that genomics will become a mainstay in analyses of infectious agents for both routine (e.g. diagnosis) and high-resolution (e.g. outbreak tracking) purposes. Significantly, assiduous analysis and interpretation of output data are essential and these processes have not yet become easy or routine. Indeed, it is arguably preferable that reliance should not be placed on automated analytical protocols but, rather, the biologist should remain actively involved throughout the processes from sample isolation and preparation through to data management, curation and analysis and, crucially, biological interpretation.

Nouvelles perspectives offertes par l'analyse des génomes des bactéries responsables de maladies infectieuses

R.J. Goldstone & D.G.E. Smith

Résumé

Les récentes avancées technologiques réalisées dans le domaine du séquençage du génome et la mise au point de systèmes permettant de manipuler et de traiter les données de séquençage ont transformé la génomique bactérienne en une méthode utilisée quasiment au quotidien dans le cadre d'études à grande ou à petite échelle sur les agents infectieux. Néanmoins, s'agissant d'agents pathogènes affectant les animaux, les applications de la génomique sont bien moins avancées que dans le domaine des agents pathogènes humains, en particulier dans le cadre d'études à grande échelle et ce, malgré l'utilisation croissante de nombreuses espèces animales dans l'alimentation. Les études génomiques réalisées en continu offrent des avantages considérables, non seulement parce qu'elles apportent des informations précises pour mieux comprendre les agents de maladies infectieuses mais aussi par leurs applications concrètes, qui permettent d'obtenir une meilleure justesse de diagnostic, de procéder à un typage de haute résolution et de concevoir des interventions plus efficaces. De telles applications et d'autres encore à venir vont probablement se développer considérablement dans un futur proche et nous pouvons nous attendre à ce qu'elles soient enfin utilisées pour étudier et caractériser les principaux agents pathogènes affectant les animaux. Les auteurs résument les objectifs de l'analyse des génomes bactériens et ses accomplissements les plus récents, en illustrant leur propos d'exemples portant essentiellement sur des pathogènes affectant les animaux, dont plusieurs figurent sur la liste de l'Organisation mondiale de la santé animale.

Mots-clés

Agents pathogènes listés par l'Organisation mondiale de la santé animale – Bactérie infectieuse – Études d'association pan-génomique – Métagénomique – Polymorphisme nucléotidique simple – Séquençage du génome entier – Vaccinologie – Vaccinologie inverse.

Nuevas perspectivas que abren los análisis genómicos de agentes infecciosos bacterianos

R.J. Goldstone & D.G.E. Smith

Resumen

Gracias a los últimos adelantos de las técnicas de secuenciación genómica y de los sistemas para procesar y explotar los datos resultantes, la genómica bacteriana se utiliza ahora de modo casi sistemático en el estudio (a pequeña o a gran escala) de agentes infecciosos. Sin embargo, pese a la creciente importancia que están cobrando muchas especies animales como fuente de productos alimentarios, la aplicación de la genómica a los agentes infecciosos animales (especialmente en los estudios a gran escala) aún va rezagada con respecto a su aplicación a los patógenos humanos. La realización asidua de estudios genómicos depara grandes beneficios, no solo porque procura un detallado conocimiento de los agentes infecciosos, sino también porque este saber sirve después para

efectuar diagnósticos más exactos, hacer tipificaciones de alta resolución y definir intervenciones terapéuticas más eficaces. Es muy probable que el uso de la genómica con estos y otros fines vaya en aumento en los próximos años, y cabe augurar que también se extenderá al estudio y la caracterización de importantes agentes infecciosos animales. Tomando básicamente como ejemplo una serie de patógenos animales (entre ellos varios agentes infecciosos incluidos en la lista de la Organización Mundial de Sanidad Animal), los autores resumen con concisión los objetivos y avances de una serie de recientes trabajos de análisis del genoma bacteriano.

Palabras clave
Agente infeccioso bacteriano – Agentes infecciosos de la lista de la OIE – Agentes de infecciones incluidas en la lista de la Organización Mundial de Sanidad Animal – Estudios de asociación del genoma completo – Metagenómica – Polimorfismos de un solo nucleótido – Secuenciación del genoma completo – Vacunología – Vacunología inversa.

References

1. Mojica S., Huot Creasy H., Daugherty S., Read T.D., Kim T., Kaltenboeck B., Bavoil P. & Myers G.S.A. (2011). – Genome sequence of the obligate intracellular animal pathogen *Chlamydia pecorum* E58. *J. Bacteriol.*, **193** (14), 3690. doi:10.1128/JB.00454-11.

2. Vannucci F.A., Kelley M.R. & Gebhart C.J. (2013). – Comparative genome sequencing identifies a prophage-associated genomic island linked to host adaptation of *Lawsonia intracellularis* infections. *Vet. Res.*, **44**, 49. doi:10.1186/1297-9716-44-49.

3. Macaulay I.C. & Voet T. (2014). – Single cell genomics: advances and future perspectives. *PLoS Genet.*, **10** (1), e1004126. doi:10.1371/journal.pgen.1004126.

4. Bhatt V.D., Ahir V.B., Koringa P.G., Jakhesara S.J., Rank D.N., Nauriyal D.S., Kunjadia A.P. & Joshi C.G. (2012). – Milk microbiome signatures of subclinical mastitis-affected cattle analysed by shotgun sequencing. *J. Appl. Microbiol.*, **112** (4), 639–650. doi:10.1111/j.1365-2672.2012.05244.x.

5. World Organisation for Animal Health (OIE) (2016). – OIE-listed diseases, infections and infestations in force in 2016. OIE, Paris. Available at: www.oie.int/en/animal-health-in-the-world/oie-listed-diseases-2016/ (accessed on 10 January 2016).

6. Zhu Ge X., Jiang J., Pan Z., Hu L., Wang S., Wang H., Leung F.C., Dai J. & Fan H. (2014). – Comparative genomic analysis shows that avian pathogenic *Escherichia coli* isolate IMT5155 (O2:K1:H5; ST Complex 95, ST140) shares close relationship with ST95 APEC O1:K1 and human ExPEC O18:K1 strains. *PLoS ONE*, **9** (11), e112048. doi:10.1371/journal.pone.0112048.

7. Mora A., Lopez C., Dabhi G., Blanco M., Blanco J.E., Alonso M., Herrera A., Mamani R., Bonacorsi S., Moulin-Schouleur M. & Blanco J. (2009). – Extraintestinal pathogenic *Escherichia coli* O1:K1:H7/NM from human and avian origin: detection of clonal groups B2 ST95 and D ST59 with different host distribution. *BMC Microbiol.*, **9** (1), 132. doi:10.1186/1471-2180-9-132.

8. Goldstone R.J., Popat R., Schuberth H.-J., Sandra O., Sheldon I.M. & Smith D.G.E. (2014). – Genomic characterisation of an endometrial pathogenic *Escherichia coli* strain reveals the acquisition of genetic elements associated with extra-intestinal pathogenicity. *BMC Genomics*, **15** (1), 1075. doi:10.1186/1471-2164-15-1075.

9. Blum S., Sela N., Heller E.D., Sela S. & Leitner G. (2012). – Genome analysis of bovine-mastitis-associated *Escherichia coli* O32:H37 strain P4. *J. Bacteriol.*, **194** (14), 3732. doi:10.1128/JB.00535-12.

10. Shepard S.M., Danzeisen J.L., Isaacson R.E., Seemann T., Achtman M. & Johnson T.J. (2012). – Genome sequences and phylogenetic analysis of K88- and F18-positive porcine enterotoxigenic *Escherichia coli*. *J. Bacteriol.*, **194** (2), 395–405. doi:10.1128/JB.06225-11.

11. Dziva F., Hauser H., Connor T.R., van Diemen P.M., Prescott G., Langridge G.C., Eckert S., Chaudhuri R.R., Ewers C., Mellata M., Mukhopadhyay S., Curtiss R. III, Dougan G., Wieler L.H., Thomson N.R., Pickard D.J. & Stevens M.P. (2013). – Sequencing and functional annotation of avian pathogenic *Escherichia coli* serogroup O78 strains reveal the evolution of *E. coli* lineages pathogenic for poultry via distinct mechanisms. *Infect. Immun.*, **81** (3), 838–849. doi:10.1128/IAI.00585-12.

12. Haines J.L., Hauser M.A., Schmidt S., Scott W.K., Olson L.M., Gallins P., Spencer K.L., Kwan S.Y., Noureddine M., Gilbert J.R., Schnetz-Boutaud N., Agarwal A., Postel E.A. & Pericak-Vance M.A. (2005). – Complement factor H variant increases the risk of age-related macular degeneration. *Science*, **308** (5720), 419–421. doi:10.1126/science.1110359.

13. Bermingham M.L., Bishop S.C., Woolliams J.A., Pong-Wong R., Allen A.R., McBride S.H., Ryder J.J., Wright D.M., Skuce R.A., McDowell S.W. & Glass E.J. (2014). – Genome-wide association study identifies novel loci associated with resistance to bovine tuberculosis. *Heredity*, **112** (5), 543–551. doi:10.1038/hdy.2013.137.

14. Finlay E.K., Berry D.P., Wickham B., Gormley E.P. & Bradley D.G. (2012). – A genome wide association scan of bovine tuberculosis susceptibility in Holstein-Friesian dairy cattle. *PLoS ONE*, **7** (2), e30545. doi:10.1371/journal.pone.0030545.

15. Settles M., Zanella R., McKay S.D., Schnabel R.D., Taylor J.F., Whitlock R., Schukken Y., Van Kessel J.S., Smith J.M. & Neibergs H. (2009). – A whole genome association analysis identifies loci associated with *Mycobacterium avium* subsp. *paratuberculosis* infection status in US Holstein cattle. *Anim. Genet.*, **40** (5), 655–662. doi:10.1111/j.1365-2052.2009.01896.x.

16. Zanella R., Settles M.L., McKay S.D., Schnabel R., Taylor J., Whitlock R.H., Schukken Y., Van Kessel J.S., Smith J.M. & Neibergs H.L. (2011). – Identification of loci associated with tolerance to Johne's disease in Holstein cattle. *Anim. Genet.*, **42** (1), 28–38. doi:10.1111/j.1365-2052.2010.02076.x.

17. Alpay F., Zare Y., Kamalludin M.H., Huang X., Shi X., Shook G.E., Collins M.T. & Kirkpatrick B.W. (2014). – Genome-wide association study of susceptibility to infection by *Mycobacterium avium* subspecies paratuberculosis in Holstein cattle. *PLoS ONE*, **9** (2), e111704. doi:10.1371/journal.pone.0111704.

18. Kirkpatrick B.W., Shi X., Shook G.E. & Collins M.T. (2011). – Whole-genome association analysis of susceptibility to paratuberculosis in Holstein cattle. *Anim. Genet.*, **42** (2), 149–160. doi:10.1111/j.1365-2052.2010.02097.x.

19. Minozzi G., Buggiotti L., Stella A., Strozzi F., Luini M. & Williams J.L. (2010). – Genetic loci involved in antibody response to *Mycobacterium avium* ssp. *paratuberculosis* in cattle. *PLoS ONE*, **5** (6), e11117. doi:10.1371/journal.pone.0011117.

20. Pant S.D., Schenkel F.S., Verschoor C.P., You Q., Kelton D.F., Moore S.S. & Karrow N.A. (2010). – A principal component regression based genome wide analysis approach reveals the presence of a novel QTL on BTA7 for MAP resistance in Holstein cattle. *Genomics*, **95** (3), 176–182. doi:10.1016/j.ygeno.2010.01.001.

21. Van Hulzen K.J.E., Schopen G.C.B., van Arendonk J.A.M., Nielen M., Koets A.P., Schrooten C. & Heuven H.C.M. (2012). – Genome-wide association study to identify chromosomal regions associated with antibody response to *Mycobacterium avium* subspecies *paratuberculosis* in milk of Dutch Holstein-Friesians. *J. Dairy Sci.*, **95** (5), 2740–2748. doi:10.3168/jds.2011-5005.

22. Tiezzi F., Parker-Gaddis K.L., Cole J.B., Clay J.S. & Maltecca C. (2015). – A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. *PLoS ONE*, **10** (2), e0114919. doi:10.1371/journal.pone.0114919.

23. Fife M.S., Howell J.S., Salmon N., Hocking P.M., van Diemen P.M., Jones M.A., Stevens M.P. & Kaiser P. (2011). – Genome-wide SNP analysis identifies major QTL for *Salmonella* colonization in the chicken. *Anim. Genet.*, **42** (2), 134–140. doi:10.1111/j.1365-2052.2010.02090.x.

24. Howell K., Weinert L., Chaudhuri R., Luan S.-L., Peters S., Corander J., Harris D., Angen O., Aragon V., Bensaid A., Williamson S.M., Parkhill J., Langford P.R., Rycroft A.N., Wren B.W., Holden M.T.G., Tucker A.W. & Maskell D.J. on behalf of the BRaDP1T Consortium (2014). – The use of genome wide association methods to investigate pathogenicity, population structure and serovar in *Haemophilus parasuis*. *BMC Genomics*, **15** (1), 1179. doi:10.1186/1471-2164-15-1179.

25. Antwerpen M.H., Prior K., Mellmann A., Höppner S., Splettstoesser W.D. & Harmsen D. (2015). – Rapid high resolution genotyping of *Francisella tularensis* by whole genome sequence comparison of annotated genes ('MLST+'). *PLoS ONE*, **10** (4), e0123298. doi:10.1371/journal.pone.0123298.

26. Biek R., O'Hare A., Wright D., Mallon T., McCormick C., Orton R.J., McDowell S., Trewby H., Skuce R.A. & Kao R.R. (2012). – Whole genome sequencing reveals local transmission patterns of *Mycobacterium bovis* in sympatric cattle and badger populations. *PLoS Pathog.*, **8** (11), e1003008. doi:10.1371/journal.ppat.1003008.

27. Roberts T., O'Connor C., Nuñez-Garcia J., de la Rua-Domenech R. & Smith N.H. (2014). – Unusual cluster of *Mycobacterium bovis* infection in cats. *Vet. Rec.*, **174** (13), 326. doi:10.1136/vr.102457.

28. Read T.D., Joseph S.J., Didelot X., Liang B., Patel L. & Dean D. (2013). – Comparative analysis of *Chlamydia psittaci* genomes reveals the recent emergence of a pathogenic lineage with a broad host range. *mBio*, **4** (2), e00604–e00612. doi:10.1128/mBio.00604-12.

29. Wattam A.R., Foster J.T., Mane S.P., Beckstrom-Sternberg S.M., Beckstrom-Sternberg J.M., Dickerman A.W., Keim P., Pearson T., Shukla M., Ward D.V., Williams K.P., Sobral B.W., Tsolis R.M., Whatmore A.M. & O'Callaghan D. (2014). – Comparative phylogenomics and evolution of the *Brucellae* reveal a path to virulence. *J. Bacteriol.*, **196** (5), 920–930. doi:10.1128/JB.01091-13.

30. Staley J.T. & Konopka A. (1985). – Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu. Rev. Microbiol.*, **39**, 321–346. doi:10.1146/annurev.mi.39.100185.001541.

31. Amann R.I., Ludwig W. & Schleifer K.H. (1995). – Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.*, **59** (1), 143–169.

32. Hugenholtz P., Goebel B.M. & Pace N.R. (1998). – Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.*, **180** (18), 4765–4774.

33. Guimaraes A.M.S., Santos A.P., do Nascimento N.C., Timenetsky J. & Messick J.B. (2014). – Comparative genomics and phylogenomics of hemotrophic mycoplasmas. *PLoS ONE*, **9** (3), e91445. doi:10.1371/journal.pone.0091445.

34. Messick J.B., Santos A.P. & Guimaraes A.M.S. (2011). – Complete genome sequences of two hemotropic mycoplasmas, *Mycoplasma haemofelis* strain Ohio2 and *Mycoplasma suis* strain Illinois. *J. Bacteriol.*, **193** (8), 2068–2069. doi:10.1128/JB.00133-11.

35. Oehlerking J., Kube M., Felder K.M., Matter D., Wittenbrink M.M., Schwarzenbach S., Kramer M.M., Hoelzle K. & Hoelzle L.E. (2011). – Complete genome sequence of the hemotrophic *Mycoplasma suis* strain KI3806. *J. Bacteriol.*, **193** (9), 2369–2370. doi:10.1128/JB.00187-11.

36. Dos Santos A.P., Guimaraes A.M.S., do Nascimento N.C., SanMiguel P.J. & Messick J.B. (2012). – Complete genome sequence of *Mycoplasma wenyonii* strain Massachusetts. *J. Bacteriol.*, **194** (19), 5458–5459. doi:10.1128/JB.01240-12.

37. Do Nascimento N.C., Guimaraes A.M.S., Santos A.P., SanMiguel P.J. & Messick J.B. (2012). – Complete genome sequence of *Mycoplasma haemocanis* strain Illinois. *J. Bacteriol.*, **194** (6), 1605–1606. doi:10.1128/JB.06781-11.

38. Barker E.N., Helps C.R., Peters I.R., Darby A.C., Radford A.D. & Tasker S. (2011). – Complete genome sequence of *Mycoplasma haemofelis*, a hemotropic mycoplasma. *J. Bacteriol.*, **193** (8), 2060–2061. doi:10.1128/JB.00076-11.

39. Barker E.N., Darby A.C., Helps C.R., Peters I.R., Heesom K.J., Arthur C.J., Crossett B., Hughes M.A., Radford A.D. & Tasker S. (2011). – Molecular characterization of the uncultivatable hemotropic bacterium *Mycoplasma haemofelis*. *Vet. Res.*, **42** (1), 83–83. doi:10.1186/1297-9716-42-83.

40. Do Nascimento N.C., Santos A.P., Guimaraes A.M., SanMiguel P.J. & Messick J.B. (2012). – *Mycoplasma haemocanis* – the canine hemoplasma and its feline counterpart in the genomic era. *Vet. Res.*, **43** (1), 66. doi:10.1186/1297-9716-43-66.

41. Do Nascimento N.C., dos Santos A.P., Chu Y., Guimaraes A.M., Baird A.N., Weil A.B. & Messick J.B. (2014). – Microscopy and genomic analysis of *Mycoplasma parvum* strain Indiana. *Vet. Res.*, **45**, 86. doi:10.1186/s13567-014-0086-7.

42. Felder K.M., Carranza P.M., Gehrig P.M., Roschitzki B., Barkow-Oesterreicher S., Hoelzle K., Riedel K., Kube M. & Hoelzle L.E. (2012). – Insights into the gene expression profile of uncultivable hemotrophic *Mycoplasma suis* during acute infection, obtained using proteome analysis. *J. Bacteriol.*, **194** (6), 1505–1514. doi:10.1128/JB.00002-12.

43. Crasta O.R., Folkerts O., Fei Z., Mane S.P., Evans C., Martino-Catt S., Bricker B., Yu G., Du L. & Sobral B.W. (2008). – Genome sequence of *Brucella abortus* vaccine strain S19 compared to virulent strains yields candidate virulence genes. *PLoS ONE*, **3** (5), e2193. doi:10.1371/journal.pone.0002193.

44. Jiang H., Du P., Zhang W., Wang H., Zhao H., Piao D., Tian G., Chen C. & Cui B. (2013). – Comparative genomic analysis of *Brucella melitensis* vaccine strain M5 provides insights into virulence attenuation. *PLoS ONE*, **8** (8), e70852. doi:10.1371/journal.pone.0070852.

45. Yu D., Hui Y., Zai X., Xu J., Liang L., Wang B., Yue J. & Li S. (2015). – Comparative genomic analysis of *Brucella abortus* vaccine strain 104M reveals a set of candidate genes associated with its virulence attenuation. *Virulence,* **6** (8), 745–754. doi:10.1080/21505594.2015.1038015.

46. Finco O. & Rappuoli R. (2014). – Designing vaccines for the twenty-first century society. *Front. Immunol.*, **5**, 12. doi:10.3389/fimmu.2014.00012.

47. Palmer G.H., Brown W.C., Noh S.M. & Brayton K.A. (2012). – Genome-wide screening and identification of antigens for rickettsial vaccine development. *FEMS Immunol. Med. Microbiol.*, **64** (1), 115–119. doi:10.1111/j.1574-695X.2011.00878.x.

48. Dark M.J., Al-Khedery B. & Barbet A.F. (2011). – Multistrain genome analysis identifies candidate vaccine antigens of *Anaplasma marginale*. *Vaccine*, **29** (31), 4923–4932. doi:10.1016/j.vaccine.2011.04.131.

49. Soares S.C., Trost E., Ramos R.T., Carneiro A.R., Santos A.R., Pinto A.C., Barbosa E., Aburjaile F., Ali A., Diniz C.A., Hassan S.S., Fiaux K., Guimarães L.C., Bakhtiar S.M., Pereira U., Almeida S.S., Abreu V.A., Rocha F.S., Dorella F.A., Miyoshi A., Silva A., Azevedo V. & Tauch A. (2013). – Genome sequence of *Corynebacterium pseudotuberculosis* biovar *equi* strain 258 and prediction of antigenic targets to improve biotechnological vaccine production. *J. Biotechnol.*, **167** (2), 135–141. doi:10.1016/j.jbiotec.2012.11.003.

50. Olsen R.H., Christensen H. & Bisgaard M. (2012). – Comparative genomics of multiple plasmids from APEC associated with clonal outbreaks demonstrates major similarities and identifies several potential vaccine-targets. *Vet. Microbiol.*, **158** (3–4), 384–393. doi:10.1016/j.vetmic.2012.03.008.

51. De Alvarenga Mudadu M., Carvalho V. & Leclercq S.Y. (2015). – Nonclassically secreted proteins as possible antigens for vaccine development: a reverse vaccinology approach. *Appl. Biochem. Biotechnol.*, **175** (7), 3360–3370. doi:10.1007/s12010-015-1507-4.

52. Yang H.L., Zhu Y.Z., Qin J.H., He P., Jiang X.C., Zhao G.P. & Guo X.K. (2006). – *In silico* and microarray-based genomic approaches to identifying potential vaccine candidates against *Leptospira interrogans*. *BMC Genomics*, **7**, 293. doi:10.1186/1471-2164-7-293.

53. Simionatto S., Marchioro S.B., Maes D. & Dellagostin O.A. (2013). – *Mycoplasma hyopneumoniae*: from disease to vaccine development. *Vet. Microbiol.*, **165** (3–4), 234–242. doi:10.1016/j.vetmic.2013.04.019.

54. Damte D., Suh J.W., Lee S.J., Yohannes S.B., Hossain M.A. & Park S.C. (2013). – Putative drug and vaccine target protein identification using comparative genomic analysis of KEGG annotated metabolic pathways of *Mycoplasma hyopneumoniae*. *Genomics*, **102** (1), 47–56. doi:10.1016/j.ygeno.2013.04.011.

55. Hong M., Ahn J., Yoo S., Hong J., Lee E., Yoon I., Jung J.K. & Lee H. (2011). – Identification of novel immunogenic proteins in pathogenic *Haemophilus parasuis* based on genome sequence analysis. *Vet. Microbiol.*, **148** (1), 89–92. doi:10.1016/j.vetmic.2010.07.023.

56. Gomez G., Pei J., Mwangi W., Adams L.G., Rice-Ficht A. & Ficht T.A. (2013). – Immunogenic and invasive properties of *Brucella melitensis* 16M outer membrane protein vaccine candidates identified via a reverse vaccinology approach. *PLoS ONE*, **8** (3), e59751. doi:10.1371/journal.pone.0059751.

57. Myers G.S., Parker D., Al-Hasani K., Kennan R.M., Seemann T., Ren Q., Badger J.H., Selengut J.D., Deboy R.T., Tettelin H., Boyce J.D., McCarl V.P., Han X., Nelson W.C., Madupu R., Mohamoud Y., Holley T., Fedorova N., Khouri H., Bottomley S.P., Whittington R.J., Adler B., Songer J.G., Rood J.I. & Paulsen I.T. (2007). – Genome sequence and identification of candidate vaccine antigens from the animal pathogen *Dichelobacter nodosus*. *Nat. Biotechnol.*, **25** (5), 569–575. doi:10.1038/nbt1302.

58. Liebenberg J., Pretorius A., Faber F.E., Collins N.E., Allsopp B.A. & van Kleef M. (2012). – Identification of *Ehrlichia ruminantium* proteins that activate cellular immune responses using a reverse vaccinology strategy. *Vet. Immunol. Immunopathol.*, **145** (1–2), 340–349. doi:10.1016/j.vetimm.2011.12.003.

59. Atzingen M.V., Gonçales A.P., de Morais Z.M., Araujo E.R., De Brito T., Vasconcellos S.A. & Nascimento A.L. (2010). – Characterization of leptospiral proteins that afford partial protection in hamsters against lethal challenge with *Leptospira interrogans*. *J. Med. Microbiol.*, **59** (Pt 9), 1005–1015. doi:10.1099/jmm.0.021485-0.

60. Murray G.L., Lo M., Bulach D.M., Srikram A., Seemann T., Quinsey N.S., Sermswan R.W., Allen A. & Adler B. (2013). – Evaluation of 238 antigens of *Leptospira borgpetersenii* serovar Hardjo for protection against kidney colonisation. *Vaccine*, **31** (3), 495–499. doi:10.1016/j.vaccine.2012.11.028.

61. Liu L., Cheng G., Wang C., Pan X., Cong Y., Pan Q., Wang J., Zheng F., Hu F. & Tang J. (2009). – Identification and experimental verification of protective antigens against *Streptococcus suis* serotype 2 based on genome sequence analysis. *Curr. Microbiol.*, **58** (1), 11–17. doi:10.1007/s00284-008-9258-x.

62. Song Y., La T., Phillips N.D., Bellgard M.I. & Hampson D.J. (2009). – A reverse vaccinology approach to swine dysentery vaccine development. *Vet. Microbiol.* **137** (1–2), 111–119. doi:10.1016/j.vetmic.2008.12.018.

63. Hatfaludi T., Al-Hasani K., Gong L., Boyce J.D., Ford M., Wilkie I.W., Quinsey N., Dunstone M.A., Hoke D.E. & Adler B. (2012). – Screening of 71 *P. multocida* proteins for protective efficacy in a fowl cholera infection model and characterization of the protective antigen PlpE. *PLoS ONE*, **7** (7), e39973. doi:10.1371/journal.pone.0039973.