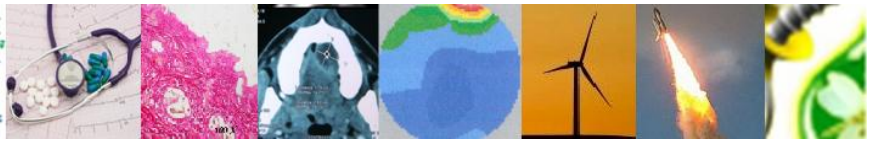

IIOAB
ISSN: 0976-3104

REVIEW: IMMUNOMICS AND VACCINOLOGY

THE REVERSE VACCINOLOGY – A CONTEXTUAL OVERVIEW

Anderson Santos[1], Amjad Ali [1], Eudes Barbosa[1], Artur Silva[2], Anderson Miyoshi[1], Debmalya Barh[3], and Vasco Azevedo[1*]

[1]*Biochemistry Departament, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, BRAZIL*

[2] *DNA Polimorfism Laboratory, Universidade Federal do Pará, Campus do Guamá - Belém, PA, BRAZIL*

[3] *Centre for Genomics and Applied Gene Technology, Institute Of Integrative Omics and Applied Biotechnology (IIOAB), Nonakuri, Purba Medinipur, West Bengal-721172, INDIA*

Received on: 14th-Nov-2010; **Revised on:** 26th-Dec -2010; **Accepted on:** 8th-Feb-2011; **Published on:** 3rd-May-2011.

***Corresponding author:** Email: vasco@icb.ufmg.br Tel: +55-3134092610; Fax: +55-3134092610

ABSTRACT

In recent years, the wide availability of complete genome sequences has changed the way we think about vaccine targets. From a few dozen potential targets we can now count on hundreds of targets per organism. This candidate vaccine is an extensively scrutinized plethora based on the concept of reverse vaccinology (RV) with special attention reserved for exported targets, generating promising results for various organisms. However it should be borne in mind that we still lack effective vaccines for organisms sequenced within a decade, a period much longer than expected for producing an effective vaccine by RV. This consideration leads to the reflection that, in the research on a vaccine, other variables may be as important as choosing a target exported. Attention is paid to the fact that the universe of possibilities for an effective vaccine can be exponential in the order of 2^n where n is the number of variables. This review compiles results of some key research using the concept of RV and raises some potential issues that may be hindering the efficient use of this technique to attain attractive and promising targets for vaccine research.

.

Keywords: *reverse vaccinology; vaccine variables chances; exported proteins; exponential function; vaccine candidates*

[I] INTRODUCTION

A decade has passed since the term *Reverse Vaccinology (RV)* was first introduced [1]. RV starts from the genomic sequence of a pathogen, which is an expected codified sequence for all the possible genes expressed in the life cycle of the pathogen. All Open Reading Frames (ORF's) derived from the genome sequence can be evaluated with a computer program in order to determine their ability to be vaccine candidates. Special attention is given to exported proteins because they are essential in host pathogen interaction. Examples of this interaction can be cited: (i) adherence to host cells, (ii) the invasion of the cell to which there was compliance, (iii) damage to host tissues, (iv) resistance to environmental stresses from machinery defense of the cell being infected and finally, (v) mechanisms for subversion of host immune response [2-5].

The word 'Reverse' from RV can be explained by the reverse genetics (RG) technique. Before the dawn of genomics, there have been attempts to discover the responsible genes from one phenotype. With Crick's Central Dogma (DNA → RNA → Protein) the research path was reversed. In possession of the likely gene sequence, several techniques were used to identify changes in the phenotype of an organism derived from sequence changes in genes. The principle of the Crick's dogma is also used by RV, in which possession of a gene sequence is searched for the possibility of a probable protein encoded by this sequence to be an antigen capable of stimulating an immune response in a host organism.

Long before the creation of the term RV, a number of approaches had been considered to meet the demand of exported proteins in order to move to the next step of the production of a subunit vaccine [6]. For example, the research using exported proteins was motivated as alternative to subunit vaccines based on polysaccharide capsule of meningococci. Vaccines produced with such antigens have low capacity to induce a satisfactory immune response. This research effort on exported proteins dates back to almost two decades of work searching for a

ISSN: 0976-3104

vaccine against meningococcal serogroup B, and now it produces good results. This vaccine currently is the best RV research results in the production of a subunit vaccine for *Neisseria meningitidis* serogroup B. Meningitis caused by serogroup B (Men B) is responsible for approximately half of the worldwide incidence of the disease [6] and this research result for targeted vaccination is commonly used as a reference card for the RV due to its excellent results. Currently a subunit vaccine against Men B created with antigens targeted by RV is selected in clinical trials of phase 2 [7, 8]. The advantages of RV are still attractive, enabling vaccine research for organisms whose cultivation in the laboratory remains difficult or impossible. However, reducing the time of selection by target proteins is feasibly usable in different species or strains at the same time and allows selecting vaccine candidates with possibility of eliciting adaptive immune responses. To achieve these benefits all we need is to have a sequenced genome, a personal computer and core software widely known by the scientific community. These conditions show another advantage of using RV, the low cost. What we agreeably call the core software is a set of tools for identifying well-known motifs such as, for example, SignalP, TMHMM, LipoP, and HMMSEARCH. In the use of core software there is still room for innovation when it determined that the choice can be directed to the identification of vaccine candidates specific to an organism such as in the case of gram-negative (bilayer) or gram positive (monolayer) or also placed according to the type of heuristic for selection of vaccine candidates with specific characteristics. For example, membrane or exported to the extracellular environment [9-12].

The concept of RV was adapted to fit a new reality of widespread availability of genomic data [13]. Instead of doing the research for vaccine targets in a single strain or subspecies of an organism, we can do it simultaneously in dozens of genomes, exploring potential joint antigens or exclusive to multiple genomes [14]. The possibility of having a large number of genomes available to implement RV leads to the emergence of the concept of Pan Genomics RV (PGRV) [8]. PGRV can also apply the concepts of core, extended, and character genomes. The core genome in PGRV is composed of exported genes (genes that transcribes for exported proteins) that are common to all strains, genes that could be candidates for a universal vaccine, while the extended genome consists of genes that are absent in at least one of the strains of the studied species and the character genome consists of genes that are specific to a strain [14]. From the standpoint of vaccine, the core and character genomes would be good candidates to compose a vaccine that is suitable for all strains studied, without losing sight of the particularities of specific genes in each strain.

[II] SYSTEMATICAL ANALYZES OF VARIABLES

Considering the motion that many studies using RV are yet to produce effective vaccines [15], an evidence that the limiting factors of RV still have considerable strength despite the enormous advances in genome sequencing has been created herein. Such limiting factors are insignificant amount of currently known antigens and the RV inability to detect non-protein antigens as polysaccharides and glycolipids [16]. These major drawbacks could be minimized with introduction of glycomics and lipidomics studies combined with genomics, proteomics, and peptidomics approaches in vaccine research that would culminate in knowledge and discovery of a wider range of antigens for *in silico* comparisons as new antigens from a survey of RV. More so, core software could also be created to identify patterns in polysaccharides and glycolipids, increasing the repertoire of antigens of an organism.

A limiting factor to the success of RV is the belief that identifying a set of exported proteins is the solution to the lack of production of an effective subunit vaccine against pathogen. Therefore, there are many possibilities of failure and only one chance of success; raising three hypothetical questions in planning a vaccine: **(A)** "Is the set of antigens suitable?" [17], **(E)** "Are antigens expressed in a critical stage of infection?" [10, 18] and (V) "What is the use of a DNA vaccine?" [19-20]. Supposing that initially, each of these three questions could have a TRUE or FALSE answer. In this case we can relate the questions A, E and V into a set of eight (2^3) possibilities, as shown in Table 1. It is the end result that matters, **(R)** "Will the vaccine be effective? The response is "YES" only if the three questions are answered with an assertive TRUE; otherwise the response will invariably be "NO".

Table-1 shows that there are possibilities, as earlier mentioned, of choosing a set of antigens sufficient to confer immunogenicity. In other words, there are chances of choosing a set of antigens effective in conferring immunogenicity, for example, for only one bacterium strain, or the selection of antigens not expressed in an important stage of infection or even the simple act of trying a subunit vaccine instead of DNA vaccine, though the set of selected antigens are adequate and expressed.

The planning of a hypothetical vaccine as shown in **Table-1** still leaves room for doubts by not taking the type of immune response most appropriate to a certain pathogen into consideration. Supposing, for example, a humoral response is not the most suitable for the pathogen of this hypothetical vaccine. Thus, even though **(A)**, **(E)**, and **(V)** are answered as TRUE, yet the vaccine could not induce protective immunity because the most appropriate response lies in the cellular immunity. So after including a fourth question being a variable in **Table-1**, **(C)** "Does vaccine generate immune response?" [21], a set of 16 possibilities was obtained (2^4) among which there are 15 possibilities of failure and only one possibility that matters the most.

This hypothetical example of planning a vaccine in **Table-1** may explain why only the selection of a set of suitable candidates still, leaving a lot of variables that can lead to failure of a vaccine approach. In planning a hypothetical vaccine for these four questions, even if the question **(A)** holds, there still remain seven other possibilities for failure to be adequately answered.

ISSN: 0976-3104

Table: 1. Possible vaccine results considering only three variables: The result shows seven failure possibilities and a record of just one success which matters the most.

No	(A) "Is the Set of antigens suitable?"	(E) "Are antigens expressed in a critical stage of infection?"	(V) "Use of a DNA vaccine?"	(R) "Will the vaccine be effective?"
1	FALSE	FALSE	FALSE	NOT
2	FALSE	FALSE	TRUE	NOT
3	FALSE	TRUE	FALSE	NOT
4	FALSE	TRUE	TRUE	NOT
5	TRUE	FALSE	FALSE	NOT
6	TRUE	FALSE	TRUE	NOT
7	TRUE	TRUE	FALSE	NOT
8	TRUE	TRUE	TRUE	YES

[III] DISCUSSION

The popularization of new technologies of genome sequencing has led to a substantial increase in the number of complete genomes for use in PGRV [14]. Given the particularities of the operating mode of each of various pathogen results and strategies, these can be used in the search for vaccine targets. Below are some of the pathogens for which RV has been used, starting with initial pathogens in the paper described the concept of RV [1] and as a result, we continue with other pathogens which do not necessarily affect humans.

3.1. Tuberculosis (TB)

Despite the prediction of decline in the world TB cases, its incidence continues to grow with more than 10 million cases reported only in 2010, keeping it among the diseases with the highest incidence worldwide [22]. Also, despite the vast amount of research for vaccine against TB, an efficient vaccine against this global scourge is still a promise. The first *Mycobacterium tuberculosis* genome sequence has been released over a decade [23, 24], but still insufficient to bring about a promising vaccine against TB. Considering the availability of complete M. tuberculosis genome sequences, the global urgency of a final solution against the scourge and the facility to conduct *in silico* research, it is inferable that in the search for vaccines understanding the wide range of research involving TB comes easier. A simple search for the term "tuberculosis" in the last three years using the PubMed database generated over 20 thousands published works that are directly or indirectly related to TB. RV was applied over *M. tuberculosis H37Rv* genome aimed at detecting secreted proteins, generating evidence of seven proteins as exoproteome properties that are possible targets for a vaccine [25]. Three secreted proteins belonging to the cutinase-like protein family (Culp) was tested and the Culp6 eliciting a strong cellular response was found [26]. It is the first cellular response recognized in patients affected by TB. These are examples of studies that fit the question **(A)** "Is the set of antigens suitable?" and the last example also characterizes the question **(C)** "Does vaccine generate immune response?" Although many of these studies did not explicitly cite the term RV, many fit the concept and try to get more information about the functional genome released by special attention to exported proteins. Questions of type **(C)** "Does vaccine generate immune response?" from our hypothetical vaccine shall be answered by researches for more effective antigens. The hypothetical protein Rv2626c was found capable of induction of adaptive and humoral immune responses [27]. However, using the concept of epitope density was to create a list of proteins with "hot spots" of the affinity of MHC class II molecules [28]. A hypothetical protein with high affinity to the promoter of genes fbp (Ag85 complex) was the result from search for over expressed factors in proteins from this antigen complex, a protein that belongs to the protein family of transcriptional regulators Mars [29].

Hypothetical membrane proteins were tested and evidenced that Rv0679c protein is expressed in only three strains of *M. tuberculosis*, although 26 strains of bacteria that possessed the gene for this protein were used [18]. Research like this show attempts to answer questions such as **(E)**, from **Table-1** for the planning of a vaccine, being crucially important as much as the question of identification of a secreted protein. Another variable that could be added to the planning table of vaccines would be **(D)** "How low is genetic diversity of selected antigens?" [30]. Included this variable, our universe of possibilities of failure would increase to 31 (2^5 - 1). For example, it was showed that classical vaccine candidates like genes such as esx, fbpB and Esat-6 would not be affected by genetic diversity in 88 strains of *M. tuberculosis* [31]. In this case the answer is "TRUE", increasing the chances of these candidates in our hypothetical screening of candidates. Under the aspect of PGRV, a set of character genes of *M. tuberculosis H37Rv* characterized as important for invasion and survival of the pathogen in the host was found by Al-Attiyah *et al*. 2010 [21]. Among these genes RD1504 was able to induce a strong

ISSN: 0976-3104

immune response T-helper (Th) type 1 and may be an important candidate for vaccine target.

3.2. Group B meningococcus

With rates of 16.9/100,000 for bacterial meningitis and 8.9/100,000 for *Neisseria meningitidis* and high number of fatalities in children, meningococcal disease remains a concern and compounded when considering the short period between infection and death, which can possibly be only one day [32]. The experience gained by *in silico* research for candidate vaccine against Men B [33] was a major factor that led to the creation of the term RV. In this work most of the antigens selected in silico and successfully expressed in *Escherichia coli* were exported proteins, including lipoproteins, OMP's, periplasm and membrane proteins. A list of five selected antigens of this study were tested with the adjuvant aluminum hydroxide, CpG oligonucleotides or MF59, achieving antibodies against more than 90% of 85 strains of meningococci representative of the global population diversity [7]. Research by adjuvants can also be included as a requirement to produce an effective vaccine that could be an additional variable in our planning of a hypothetical vaccine **[Table-1]**. It was showed that the amount of factor H, an important regulator of the complement pathway, is correlated with the level of expression of GNA1870, suggesting the inclusion of this protein in the set of antigens of Men B [34]. This research is useful in trying to answer questions of type **(E)** "Are antigens expressed in a critical stage of infection?" [35]. Although there are two vaccines in development for incorporating the Men B protein named Factor H-binding protein or fHbp [36]; it has not been possible to produce a comprehensive vaccine based on this antigen due to its wide antigenic variety [30, 37]. This variety motivated the establishment of a nomenclature to categorize this diversity [38], a study that answers questions such as **(D)** "How low is genetic diversity of selected antigens?". The discovery that convalescent patients develop long-term protective immunity against *N. meningitidis* motivated the search for antigens capable of eliciting such immune response [15]. Contrary to the RV concept, most of the antigens were found cytoplasmic and were not able to produce a satisfactory immune response in guinea pigs. There is also the protein RplY proven to belong to the cell surface of the pathogen. This result makes it a little more confusing to answer the question **(A)** "Is the set of antigens suitable? ", since most of candidates would not be exported. After a decade of the first results of RV on *N. meningitidis*, suggested antigens continue to be researched. It was discovered recently that GNA2132 known as a protein capable of inducing a bactericidal antibody in mice, is also capable of inducing protective immunity in humans. This protein is recognized by serum of convalescent patients, and has been renamed Neisserial Heparin Binding Antigen (NHBA), which is one of the most promising in the search for a vaccine against the pathogen [39] and helping to answer questions of type (C) "Does vaccine generate cellular immune response?".

3.3. Staphylococcus aureus

Staphylococcus aureus, a gram positive bacterium remains one of the major human pathogens and a major cause of nosocomial infections worldwide. Failure of antibiotic therapy to eradicate infection is frequently described in literatures and the rate of resistance to clinically relevant antibiotics, such as methicillin, is increasing. Furthermore, there has been an increase in the number of methicillin-resistant S. aureus community-acquired infections [40]. The high prevalence of infections is confounded by the ability of the pathogen to readily acquire genetic elements that confer resistance to antibiotics [41]. The first *S. aureus* complete genome was available on the Gene Bank databases and the Broad Institute since 2007 [42], followed by other 14 different strains at NCBI. There is need for decoding the sequences of complete genome of *S. aureus* that could offer the possibility for comprehensive screening to identify the targets for effective vaccine development [43]. So, it could be interesting to try the answer **(D)** "How low is genetic diversity of selected antigens?" when considering vaccine candidates. Clinical trials with monovalent traditional vaccines already failed to protect against the disease. Now the need is to shift from monovalent vaccine development towards the potential use of multivalent formulations, therapeutic antibodies, and more systematic and rapid identification of optimal antigens by applying *in silico* tools [44]. The RV concept is most suitable for the *S. aureus* to meet the research needs and there are case studies applying it. For example, there are at least 153 individual antigens characterized with the immunome of *S. aureus* [45], despite their subcelular location, which help to answer the question of type **(C)** "Does vaccine generate immune response?". Antibody responses produced against those antigens are accessible to B cells in vivo, most likely extracellular and cell wall-associated proteins, but also non-protein antigens, such as wall teichoic acids (WTA) lipoteichoic acids (LTA), and peptido glycans (PGN) [45]. Surface protein antigens IsdA, IsdB, SdrD, and SdrE were tested for its vaccine efficacy in a combination and individually, the serum IgG titers of immunized mice were almost the same [46]. Immunodominant antigen B (IsaB) is a surface protein believed to be a virulence factor, although its biological functions are not well defined. Its nucleic acid-binding activity is being observed. IsaB has greater affinity for dsDNA than it has for ssDNA or RNA, there is need to evaluate and understand the role of IsaB and its nucleic acid-binding activity which are important in establishment and/or progression of *S. aureus* infection [47] and to answer questions like **(A)** and **(C)** from our hypothetical vaccine planning. Immune dominance of extracellular and surface-exposed proteins has indeed been observed with an *S. aureus* genomic expression library, ribosome display, and 2D-IB. Also most surface associated genes in the core variable genome as well as a large amount of virulence and resistance factors, and many are encoded on mobile genetic elements [45], helping to answer questions like **(A)** and **(D)** from our hypothetical vaccine planning.

ISSN: 0976-3104

3.4. Corynebacterium diphtheriae

Corynebacterium diphtheriae bacteria are responsible for Diphtheria. The pathogen produces a toxin that can harm or destroy body tissues and organs, diphtheria toxin (DT). The disease primarily affects mucous membranes of the respiratory tract (Respiratory Diphtheria), although it can also affect the skin (Cutaneous Diphtheria). The bacteria were identified for the first time in the 1880's. In the 1890's, the first antitoxin were developed, and the first vaccine in 1921 [48]. The vaccine is made of inactive forms of the toxin. According to the World Health Organization (WHO), diphtheria affects people of all ages, but it is more frequent in non-immunized child. In 2004, 5000 deaths owing to the disease were reported worldwide [49]. The most effective treatment consists the administration of diphtheria antitoxin (DAT) associated with the elimination of the microorganism using appropriate antibiotics. Since the beginning of the 20th century, many countries produced their own antitoxin preparation from horses. However many factors led to fall of this traditional stocks. Among this factors it is possible to exemplify the lost of economic viability, once the incidence of the disease has fallen a lot, and public objections to the use of horses as donors [50]. In the 1990's, the lack of DAT and vaccines were the two major causes behind the outbreak of diphtheria in the former Union of Soviet Socialist Republics (USSR) states. Between 1990 and 1998, more than 157,000 cases and 5,000 deaths were registered, which represented 80% of diphtheria reports worldwide [51]. This was the major diphtheria epidemic since the 1950's, when the spread of immunization had began. The first genome annotation of *C. diphtheriae* was released in 2003 [52]. With the genome published, Hansmeier *et al.* mapped and analyzed the extracellular and membrane surface of the C7s(-)tox-lineage. This work identified unambiguously ~32% (85/263) of the protein previously described as being extracellular. There were 107 extracellular proteins and 53 of the cell surface, representing a total of 85 different proteins [53]. The importance of this study is to identify secreted proteins, once they can be involved in important interactions between bacteria and host, helping to plan for a vaccine according to the variables in table 1, more specifically question **(A)** "Is the set of antigens suitable?", characterizing a RV research.

Another possible question is **(P)** "Is the antigen in a Pathogenicity Island (PAI)?" [54]. In order to try answer such question for the model *C. diphtheriae*, a comparative genomic hybridization of different strains was made against the specie reference strain [54]. Now, our hypothetical vaccine planning could become more complicated, reaching a total of 63 (2^6 -1) failure possibilities. C7(-) strain was suggested to lack 11 PAIs among 13, while strain PW8, isolated earlier than the C7(-) strain, were lacking only 3 regions related to PAIs.

Additionally, a large genomic diversity among various *C. diphtheriae* strains and clinical isolates were observed. Although the difference between C7(-) strain was higher than the PW8, the adhesion of C7(-) were comparable to the reference strain, while PW8 showed reduce adherence compared to the other strains [54]. This is some odd result considering that adhesions of *C. diphtheriae* to human epithelial cells followed by internalization are signs of pathogenicity.

Searching the literature we can also find studies about the pilli, a very important structure in the adherence of the bacteria and the host [55, 56]. These structures are often involved in the initial adhesion of the bacteria to host tissues during colonization and so helps answer the question **(A)** "Is the set of antigens suitable?" Besides having a structural importance, there are some reports in the literature stating that it might be an important vaccine target, as it is critical in the invasion process [57, 58]. This also helps answer the question **(E)** "Are antigens expressed in a critical stage of infection?"

Compared to the research of vaccine targets of other pathogens in this review, the research on targets of *C. diphtheriae* is less intensive. This apparent calm may be associated with the impression that diphtheria is controlled and that existing vaccines, mostly based on only one antigen, is sufficient to control the disease. However, a feature that makes the development of a vaccine against the bacteria very interesting is the raise of the nontoxigenic strains. Although they don't release the DT in the organism, they are capable of causing morbidity and death [59]. They were isolated from injection drug users in Switzerland [60], homeless alcoholics in France [61], and from poor populations of Vancouver, Canada [59]. In addition, an increasing proportion of strains isolated in the United Kingdom are nontoxigenic [62]. For such non toxic *C. diphtheriae* strains it is possible that the search for adequate variables also can make the difference between a vaccine success or failure.

[V] CONCLUSION

Despite the extensive use of the initial concept of RV and advanced rated results in the search for vaccines against certain pathogens, in general its best and most practical results are still expected. This conclusion is based on fact that the genome sequences of some of the major human pathogens are known longer than the average time required for RV application and promising vaccine against these pathogens still seems far away. Also despite the limitations of RV, this is a low cost technique, fully feasible of use into the plethora of genomic data being generated. It is justifiable to use it in a broader range of pathogens. It is possible that for some of the major human and animal pathogens we can find an appropriate combination of antigens enabling the creation of effective vaccines capable of improving the people's quality of live directly through prevention of diseases or indirectly by improving the economic conditions dependent on breeding. However, as shown in the hypothetical example of planning a vaccine, the discovery of suitable antigens could be a small part of the problem of producing an effective vaccine, but never the less important.

ISSN: 0976-3104

FINANCIAL DISCLOSURE

This work was carried out with financial support from development agencies CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais), CAPES (Coordenação de aperfeiçoamento de Pessoal de Nível Superior)

ACKNOWLEDGEMENT

The authors thank Dr. Sergio Costa Oliveira for fruitful discussion and comments.

REFERENCES

[1] Rappuoli R. [2000] Reverse vaccinology. *Curr Opin Microbiol* 3:445-450.
[2] Sibbald MJJB, van Dij JML. [2009] Secretome Mapping in Gram-Positive Pathogens. In Karl Wooldridge (ed.), Bacterial Secreted Protein: Secretory Mechanisms and Role in Pathogenesis. *Caister Academic Press* :193-225.
[3] Simeone R, Bottai D, Brosch R, et al. [2009] ESX/type VII secretion systems and their role in host-pathogen interaction. *Curr Opin Microbiol* 12:4-10.
[4] Stavrinides J, McCann HC, Guttman DS, et al. [2008] Host-pathogen interplay and the evolution of bacterial effectors. *Cell Microbiol* 10:285-292.
[5] Bhavsar AP, Guttman JA, Finlay BB, et al. [2007] Manipulation of host-cell pathways by bacterial pathogens. *Nature* 449:827-834.
[6] Diaz Romero J, Outschoorn IM. [1994] Current status of meningococcal group B vaccine candidates: capsular or noncapsular? *Clin Microbiol Rev* 7:559-575.
[7] Giuliani MM, Adu-Bobie J, Comanducci M, et al. [2006] A universal vaccine for serogroup B meningococcus. *Proc Natl Acad Sci U S A* 103:10834-10839.
[8] Bambini S, Rappuoli R. [2009] The use of genomics in microbial vaccine development. *Drug Discov Today* 14:252-260.
[9] Barinov A, Loux V, Hammani A, et al. [2009] Prediction of surface exposed proteins in *Streptococcus pyogenes*, with a potential application to other Gram-positive bacteria. *Proteomics* 9:61-73.
[10] Yang H, Zhu Y, Qin J, et al. [2006] In silico and microarray-based genomic approaches to identifying potential vaccine candidates against *Leptospira interrogans*. *BMC Genomics* 7:293.
[11] Taylor PD, Attwood TK, Flower DR, et al. [2006] Combining algorithms to predict bacterial protein sub-cellular location: Parallel versus concurrent implementations. *Bioinformation* 1:285-289.
[12] Taylor PD, Toseland CP, Attwood TK, et al. [2006] TATPred: a Bayesian method for the identification of twin arginine translocation pathway signal sequences. *Bioinformation* 1:184-187.
[13] Rinaudo CD, Telford JL, Rappuoli R, et al. [2009] Vaccinology in the genome era. *J Clin Invest* 119:2515-2525.
[14] Lapierre P, Gogarten JP, [2009] Estimating the size of the bacterial pan-genome. *Trends Genet* 25:107-110.
[15] Mendum TA, Newcombe J, McNeilly CL, et al. [2009] Towards the Immunoproteome of *Neisseria meningitidis PLoS ONE* 4:.
[16] Rappuoli R. [2001] Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine* 19:2688-2691.
[17] Choi G, Eom S, Jung K, et al. [2010] CysA2: A candidate serodiagnostic marker for *Mycobacterium tuberculosis* infection. *Respirology* 15:636-642.
[18] Cifuentes DP, Ocampo M, Curtidor H, et al. [2010] Mycobacterium tuberculosis Rv0679c protein sequences involved in host-cell infection: potential TB vaccine candidate antigen. *BMC Microbiol* 10:109.
[19] Gat O, Grosfeld H, Ariel N, et al. [2006] Search for *Bacillus anthracis* Potential Vaccine Candidates by a Functional Genomic-Serologic Screen. *Infect Immun* 74:3987-4001.
[20] Dunachie SJ, Hill AVS, [2003] Prime-boost strategies for malaria vaccine development. *J Exp Biol* 206:3771-3779.
[21] Al-Attiyah R, Mustafa AS. [2010] Characterization of human cellular immune responses to *Mycobacterium tuberculosis* proteins encoded by genes predicted in RD15 genomic region that is absent in *Mycobacterium bovis* BCG. *FEMS Immunol Med Microbiol* 59:177-187.
[22] Dye C, Williams BG. [2010] The Population Dynamics and Control of Tuberculosis. *Science* 328:856-861.
[23] Cole ST, Brosch R, Parkhill J, et al. [1998] Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* 393:537-544.
[24] Cole ST, Barrell BG. [1998] Analysis of the genome of *Mycobacterium tuberculosis* H37Rv. *Novartis Found Symp* 217:160-72; discussion 172-7.
[25] Vizcaíno C, Restrepo-Montoya D, Rodríguez D, et al. [2010] Computational prediction and experimental assessment of secreted/surface proteins from *Mycobacterium tuberculosis H37Rv*. *PLoS Comput Biol* 6:e1000824.
[26] Shanahan ER, Pinto R, Triccas JA, et al. [2010] Cutinase-like protein-6 of *Mycobacterium tuberculosis* is recognised in tuberculosis patients and protects mice against pulmonary infection as a single and fusion protein vaccine. *Vaccine* 28:1341-1346.
[27] Bashir N, Kounsar F, Mukhopadhyay S, et al. [2010] *Mycobacterium tuberculosis* conserved hypothetical protein rRv2626c modulates macrophage effector functions. *Immunology* 130:34-45.
[28] Gaseitsiwe S, Valentini D, Mahdavifar S, et al. [2010] Peptide microarray-based identification of *Mycobacterium tuberculosis* epitope binding to HLA-DRB1*0101, DRB1*1501, and DRB1*0401. *Clin Vaccine Immunol* 17:168-175.
[29] Romero IC, Mehaffy C, Burchmore RJ, et al. [2010] Identification of promoter-binding proteins of the fbp A and C genes in *Mycobacterium tuberculosis Tuberculosis (Edinb)* 90:25-30.
[30] Nash JH, Findlay WA, Luebbert CC, et al. [2006] Comparative genomics profiling of clinical isolates of *Aeromonas salmonicida* using DNA microarrays. *BMC Genomics* 7:43.
[31] Davila J, Zhang L, Marrs CF, et al. [2010] Assessment of the genetic diversity of *Mycobacterium tuberculosis* esxA, esxH, and fbpB genes among clinical isolates and its implication for the future immunization by new tuberculosis subunit vaccines Ag85B-ESAT-6 and Ag85B-TB10.4. *J Biomed Biotechnol* 2010:208371.
[32] Theodoridou MN, Vasilopoulou VA, Atsali EE, et al. [2007] Meningitis registry of hospitalized cases in children: epidemiological patterns of acute bacterial meningitis throughout a 32-year period. *BMC Infect Dis* 7:101.

ISSN: 0976-3104

[33] Pizza M, Scarlato V, Masignani V, et al. [2000] Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 287:1816-1820.

[34] Madico G, Welsch JA, Lewis LA, et al. [2006] The Meningococcal Vaccine Candidate GNA1870 Binds the Complement Regulatory Protein Factor H and Enhances Serum Resistance. *J Immunol* 177:501-510.

[35] Montor WR, Huang J, Hu Y, et al. [2009] Genome-Wide Study of *Pseudomonas aeruginosa* Outer Membrane Protein Immunogenicity Using Self-Assembling Protein Microarrays▿. *Infect Immun* 77:4877-4886.

[36] Beernink PT, Welsch JA, Harrison LH, et al. [2007] Prevalence of Factor H–Binding Protein Variants and NadA among Meningococcal Group B Isolates from the United States: Implications for the Development of a Multicomponent Group B Vaccine. *J Infect Dis* 195:1472-1479.

[37] Lipsitch M, O'Hagan JJ. [2007] Patterns of antigenic diversity and the mechanisms that maintain them. *J R Soc Interface* 4:787-802.

[38] Brehony C, Wilson DJ, Maiden MC, et al. [2009] Variation of the factor H-binding protein of *Neisseria meningitidis*. *Microbiology* 155:4155-4169.

[39] Serruto D, Spadafina T, Ciucchi L, et al. [2010] Neisseria meningitidis GNA2132, a heparin-binding protein that induces protective immunity in humans. *Proc Natl Acad Sci U S A* 107:3770-3775.

[40] Kuklin NA, Clark DJ, Secore S, et al. [2006] A novel Staphylococcus aureus vaccine: iron surface determinant B induces rapid antibody responses in rhesus macaques and specific increased survival in a murine *S. aureus* sepsis model. *Infect Immun* 74:2215-2223.

[41] Ventura CL, Malachowa N, Hammer CH, et al. [2010] Identification of a novel *Staphylococcus aureus* two-component leukotoxin using cell surface proteomics. *PLoS ONE* 5:e11634.

[42] Herron-Olson L, Fitzgerald JR, Musser JM, et al. [2007] Molecular correlates of host specialization in *Staphylococcus aureus*. *PLoS ONE* 2:e1120.

[43] McCarthy AJ, Lindsay JA. [2010] Genetic variation in *Staphylococcus aureus* surface and immune evasion genes is lineage associated: implications for vaccine design and host-pathogen interactions. *BMC Microbiol* 10:173.

[44] Otto M. [2010] Novel targeted immunotherapy approaches for staphylococcal infection. *Expert Opin Biol Ther* 10:1049-1059.

[45] Holtfreter S, Kolata J, Bröker BM, et al. [2010] Towards the immune proteome of *Staphylococcus aureus* - The anti-S. aureus antibody response. *Int J Med Microbiol* 300:176-192.

[46] Stranger-Jones YK, Bae T, Schneewind O, et al. [2006] Vaccine assembly from surface proteins of *Staphylococcus aureus*. *Proc Natl Acad Sci U S A* 103:16942-16947.

[47] Mackey-Lawrence NM, Potter DE, Cerca N, et al. [2009] *Staphylococcus aureus* immunodominant surface antigen B is a cell-surface associated nucleic acid binding protein. *BMC Microbiol* 9:61.

[48] Agnew J. [2010] Medicine in the Old West: A History, 1850-1900. McFarland

[49] World Healt Organization [2010] Diphtheria [http://www.who.int/immunization/topics/diphtheria/en/index.html]

[50] Wagner K, Stickings P, White J, et al. [2009] A review of the international issues surrounding the availability and demand for diphtheria antitoxin for therapeutic use. *Vaccine* 28:14 - 20.

[51] Dittmann S, Wharton M, Vitek C, et al. [2000] Successful control of epidemic diphtheria in the states of the Former Union of Soviet Socialist Republics: lessons learned. *J Infect Dis* 181 Suppl 1:S10-22.

[52] Cerdeño-Tárraga AM, Efstratiou A, Dover LG, et al. [2003] The complete genome sequence and analysis of *Corynebacterium diphtheriae NCTC13129*. *Nucleic Acids Res* 31:6516-6523.

[53] Hansmeier N, Chao T, Kalinowski J, et al. [2006] Mapping and comprehensive analysis of the extracellular and cell surface proteome of the human pathogen *Corynebacterium diphtheriae*. Proteomics 6:2465-2476.

[54] Iwaki M, Komiya T, Yamamoto A, et al. [2010] Genome organization and pathogenicity of *Corynebacterium diphtheriae* C7(-) and PW8 strains. *Infect Immun* 78:3791-3800.

[55] Kang HJ, Paterson NG, Gaspar AH, et al. [2009] The *Corynebacterium diphtheriae* shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds. *Proc Natl Acad Sci U S A* 106:16967-16971.

[56] Mandlik A, Swierczynski A, Das A, et al. [2007] *Corynebacterium diphtheriae* employs specific minor pilins to target human pharyngeal epithelial cells. *Mol Microbiol* 64:111-124.

[57] Telford JL, Barocchi MA, Margarit I, et al. [2006] Pili in gram-positive pathogens. *Nat Rev Microbiol* 4:509-519.

[58] Proft T, Baker EN. [2009] Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease. *Cell Mol Life Sci* 66:613-635.

[59] Romney MG, Roscoe DL, Bernard K, et al. [2006] Emergence of an invasive clone of nontoxigenic *Corynebacterium diphtheriae* in the urban poor population of Vancouver, Canada. *J Clin Microbiol* 44:1625-1629.

[60] Gubler J, Huber-Schneider C, Gruner E, et al. [1998] An outbreak of nontoxigenic *Corynebacterium diphtheriae* infection: single bacterial clone causing invasive infection among Swiss drug users. *Clin Infect Dis* 27:1295-1298.

[61] Patey O, Bimet F, Riegel P, et al. [1997] Clinical and molecular study of *Corynebacterium diphtheriae* systemic infections in France. Coryne Study Group. *J Clin Microbiol* 35:441-445.

[62] Health Protection Agency. [2006] Diphtheria Notifications and Deaths: England and Wales 1986–2006. [http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/Diphtheria/EpidemiologicalData/]

ISSN: 0976-3104

ABOUT AUTHORS

Anderson Rodrigues Santos; *MSc, has a degree in Computer Science from Catholic University of Minas Gerais (1995) and MSc in Computer Science from Universidade Federal de Minas Gerais (1999). He has experience in computer science, with emphasis on Logic Programming. Possibly as a guest professor at several graduate courses as Electrical Engineering, C. Accounting, C. Administrative and Digital Games. Started a PhD in Bioinformatics in March 2008. Outstanding performance in the assembly and annotation of the first fully assembled and annotated genome in the state of Minas Gerais for Genomic Network of Minas Gerais on bacteria Corynebacterium pseudotuberculosis 1002, which infects goats. By LGCM and Genomic Network of Pará participated specifically in automatic and manual annotation of six other strains of the same bacteria (C231, I19, PAT10, 162, 258 and CIP5297) through the use of databases management systems and compilers*

Amjad Ali; *BS (Hons) MPhil in Biotechnology/Genetics (2008), PhD student in Genetics with special focus on Genetics of Microorganisms: assembly and annotation of genomes, pan genomics, comparative genomics and pathogenomics analysis of different Corynebacterium psuedotuberculosis strains, for the development of Vaccines against caseous Lymphadenitis (CLA) in sheep and Goats.*

Eudes Guilherme Vieira Barbosa; *is in the final year of Biology in Federal University of Minas Gerais and is an undergraduate research in the Genetic Department of the Instituto de Ciências Biológicas. Has experience in Genetics, focusing on Molecular Genetics and of Microorganisms. Currently is working in the assembly and annotation of Corynebacterium pseudotuberculosis strains.*

Prof. Artur Luiz da Costa da Silva*; MSc, PhD, has a master and PH.D decrees in Genetics and Molecular Biology from the Federal University of Pará (UFPA). Since 1996 he is a professor in UFPA and he one of the coordinators of the Genomic and Proteomic Web of Pará. He is a level 2 CNPq researcher and is affiliated to the Brazilian Academy of Science.*

Prof. Anderson Miyoshi; *M.Sc, PhD, is adjunt professor of the General Biology Departmant in Federal University of Minas Gerais. His research interests are in Molecular Genetics and of Microorganisms. Currently is working in the development of new genomic expression systems. His total research publications are 35 and 5 book chapters.*

Prof. Vasco Azevedo; *DVM, M.Sc, PhD, FESC, is full Professor of Federal University of Minas Gerais. Professor Azevedo is a pioneer of genetics of Lactic Acid Bacteria and Corynebacterium pseudotuberculosis in Brazil. He has specialized in bacterial genetics, genome, transcriptome, proteome, developement of new vaccines and diagnostic against infectious diseases. He is the Associate Editor of Genetics and Molecular Research, and member of editorial board of Open Veterinary Science Journal and Internacional Journal of Microbiology. His total number of research publications is 103 and has authored 11 book chapters.*