# Genetic diversity and immune evasion of bacterial pathogens

## INAUGURALDISSERTATION

zur

Erlangung der Würde einer Doktorin der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

**Charlotte Huber**

aus Zürich und Biel (BE)

Basel, 2011

creative commons

**Attribution-Noncommercial-No Derivative Works 2.5 Switzerland**

---

## You are free:

to Share — to copy, distribute and transmit the work

## Under the following conditions:

**Attribution.** You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

**Noncommercial.** You may not use this work for commercial purposes.

**No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- Nothing in this license impairs or restricts the author's moral rights.

**Your fair dealing and other rights are in no way affected by the above.**

This is a human-readable summary of the Legal Code (the full license) available in German:
http://creativecommons.org/licenses/by-nc-nd/2.5/ch/legalcode.de

**Disclaimer:**
The Commons Deed is not a license. It is simply a handy reference for understanding the Legal Code (the full license) — it is a human-readable expression of some of its key terms. Think of it as the user-friendly interface to the Legal Code beneath. This Deed itself has no legal value, and its contents do not appear in the actual license. Creative Commons is not a law firm and does not provide legal services. Distributing of, displaying of, or linking to this Commons Deed does not create an attorney-client relationship.

Quelle: http://creativecommons.org/licenses/by-nc-nd/2.5/ch/deed.en Datum: 3.4.2009

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel auf Antrag von

Prof. Dr. Gerd Pluschke und Prof. Dr. Ulrich Certa

Basel, 14. Dezember 2010

Prof. Dr. Martin Spiess
Dekan

Dedicated to Ruth Ruprecht
who has become a dear friend
and mother to me

# ACKNOWLEDGMENTS

My first thanks go to my supervisor Prof. Gerd Pluschke who granted me the opportunity to do this thesis in his lab. It was a happy time and a time of personal growth that I would never want to miss. Also I want to express my gratitude towards Prof. Ulrich Certa for agreeing to be my co-referee and reading the following pages.

Sincere thanks also go to Prof. Marcel Tanner for having welcomed me at the Swiss TPH, as well as for his positive spirit and outstanding leadership.

My gratitude is also addressed to PD Dr. Claudia Daubenberger. Thanks for the friendship, the excellent supervision and the joint adventure of launching the collaboration with Prof. Revathi Gunturu and Dr. Francis McOdimba from the Aga Khan Hospital in Nairobi. Thanks also to Revathi and Francis, meeting you was a gift from God.

Many thanks to Jean-Pierre Dangy, Julia Hauser, Miriam Bolz, Anita Dreyer, Martin Bratschi, Katharina Röltgen, Theresa Ruf, Angèle Bénard, Nicole Scherr, Mark Bauer and also the former members of my lab group, Denise Vogel, Martin Nägeli, Shinji Okitsu, Elisabetta Peduzzi, Vanessa Racloz and Weihong Qi. Thank you for your support, your acceptance and the fun times we shared. Thanks also to Dr. Michael Käser for the good work we did together. Special thanks go to Valentin Pflüger. Thank you for your support, your loyalty and your irreplaceable sense of humour. My thanks are also addressed to Kerensa McElroy, one of the most sincere people I know. Talking about Australian friends, I also want to thank Paul Johnson; it was good having you with us. Many thanks also to the Ghanaian ladies, my friend Ernestina Mensa-Quanioo and Dorothy Yeboah.

I also want to express my gratitude towards Jürg Utzinger, Jennifer Keiser, Reto Brun, Ingrid Felger, Hans-Peter Beck, Sebastien Gagneux, Pascal Mäser, Yvette Endriss, Till Voss, Christian Scheurer, Kathrin Witmer, Christian Flück, Igor Niederwieser, Caroline Kulangara, Dania Müller, Sophie Oehring, Claudia List,

Esther Pachlatko, and all the members of the MPI. Thank you for all your help, the fun times and the “chaleur humaine” you brought into my life here at the Swiss TPH. Thanks to Lukas Camenzind, Dominique Forster, Simon Schlumpf and the entire IT team for solving my computer problems. And many thanks to Marcel Stöckle, Beatrice Wäckerlin, Joachim Pelikan and all collaborators at the Institute. You make the Swiss TPH a place to be.

Finally I would like to thank my family and all my friends far and near. Thanks for being by my side. A special thanks goes to my boyfriend, David Preston. Thank you for patiently waiting for me till we can hit the road.

# SUMMARY

Within the framework of this thesis the genetic diversity of three bacterial pathogens, *Mycobacterium ulcerans, Neisseria meningitidis* and *Streptococcus agalactiae* was investigated. The aim of these analyses was to contribute to the understanding of how genetic properties of the pathogens contribute to immune evasion. Implications of the findings for vaccine design are discussed.

## *Mycobacterium ulcerans*

Buruli ulcer is a disease of skin and soft tissue caused by the bacterial pathogen *Mycobacterium ulcerans*. *M. ulcerans* has recently diverged from an *M. marinum* progenitor through the acquisition of a virulence plasmid, lateral gene transfer and reductive evolution. Isolates of *M. ulcerans* deriving from different regions of the globe can be associated with two distinct lineages, either the ancestral or the classical lineage. Here, we show that the two copies of the *esx*B-*esx*A gene cluster present in the genome of *M. marinum* are both deleted from the genome of *M. ulcerans* strains belonging to the highly virulent classical lineage. Members of the ancestral *M. ulcerans* lineage instead retained copies of the *esx*B-*esx*A gene cluster. Additionally, the *hsp*X gene was present in the strains of the ancestral lineage and absent in the classical lineage. Our results indicate that *M. ulcerans* is adapting to an environment that is screened by immune recognition mechanisms by loss of highly immunogenic proteins.

## *Neisseria meningitidis*

Certain hypervirulent lineages of *Neisseria meningitidis*, a commensal of the human nasopharynx, are a major cause of meningitis and septicaemia. Here we have investigated subcapsular antigens of serogroup A *Neisseria meningitidis* strains isolated in the course of longitudinal colonization and disease surveys in the African meningitis belt. In the course of clonal waves of colonization and disease we observed no sequence diversification of the outer membrane proteins PorA, PorB

and FetA. In contrast, high variability in the expression of Opa proteins was observed due to changing numbers of pentamer repeats within the open reading frames of the four *opa* genes *opa*A, *opa*B, *opa*D and *opa*J. Furthermore, we found some exchange of alleles of the *opa* genes OpaA and OpaJ by horizontal gene transfer. Herd immunity may thus be a stronger driving force for diversification of Opa proteins than for other outer membrane proteins.

## *Streptococcus agalactiae*

While *Streptococcus agalactiae*, the group B streptococcus (GBS), is traditionally considered a neonatal pathogen, it is also emerging as a significant cause of morbidity in adults. Here we have analysed the population structure of GBS isolates, collected from carriers and clinical cases in Kenya. Multi-locus sequence typing differentiated the 173 strains analyzed into 22 sequence types (STs), including 5 novel STs. A close correlation between STs and distinct capsular serotypes was found with the disease isolates being more diversified with respect to both STs and capsular serotypes than carrier isolates. The STs and capsular serotypes most prevalent in Kenya were also commonly found in many other regions of the world.

In this investigation, the highest genetic variablility was found in our GBS collection. In the *N. meningitidis* isolates collected during clonal waves of meningococcal colonization and disease, there was a striking lack of diversification, with the exception of the *opa* genes. The most conserved bacterial pathogen in this study was *M. ulcerans,* where no genetic variability could be found within a geographic region.

No vaccines exist to date against *M. ulcerans* or *S. agalactiae,* and an affordable universal vaccine against *N. meningitidis* is urgently needed. The technological advances in whole genome sequencing are likely to facilitate efforts towards finding suitable candidate antigens for subunit vaccines.

# CHAPTER1: INTRODUCTION

## General Introduction

Bacteria are unicellular prokaryotes usually surrounded by a complex cell wall and often a thick capsule. The bacterial chromosome consists of a double-stranded DNA molecule which is not contained within a nuclear membrane.

Gram-staining and the Ziehl-Neelsen stain allow the differentiation of bacterial species into broad groups, and bacteria are either Gram-positive or Gram-negative, based on the properties of their cell walls. The main component of the cell wall is the peptidoglycan murein, a polymer of *N*-acetylglucosamine and *N*-acetylmuramic acid as well as amino acids. In gram positive bacteria, murein forms a relatively thick outer layer (20-80 nm), whereas the murein layer in Gram-negative bacteria is relatively thin (5-10 nm) and covered by an outer membrane.

The cell wall of Mycobacteria consists of a hydrophobic, waxy outer layer containing a variety of mycolic acids and a peptidoglycan layer linked by the polysaccharide arabinogalactan. Lacking an outer cell membrane, Mycobacteria are considered Gram-positive. They do not readily take up the Gram-stain however, but can be identified by the Ziehl-Neelsen stain and are known as acid-fast bacteria.

Species identification of bacteria can be performed either by non-cultural techniques such as microscopy and the detection of bacterial antigens in specimens, or by cultivation. The following characteristics are traditionally looked at in cultivated bacteria: Gram reaction, cell morphology and arrangement, growth requirements, the ability to produce certain enzymes as well as distinct metabolic properties.

Gene Sequencing of 16S rRNA has been established as an identification method of bacterial species, and matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) mass spectroscopy is likely to be a front-line identification method of the future.

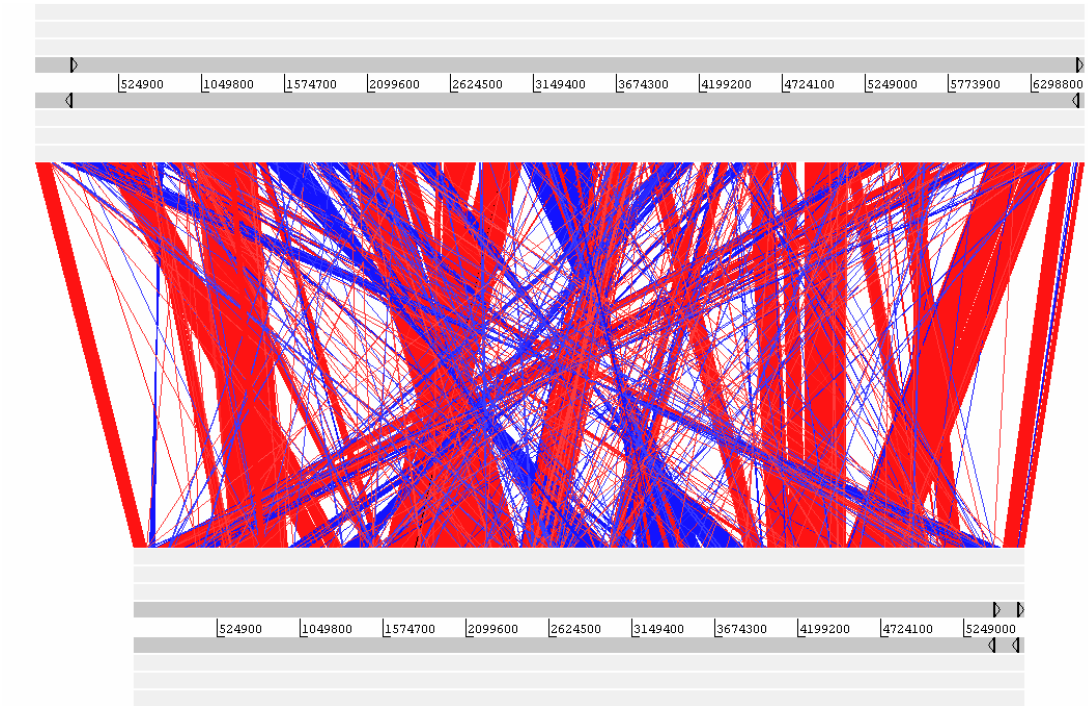
At the start of the 20th century, a vast number of people still died from infectious diseases that today are easily curable with the help of antibiotics. However, the genetic variability of bacteria by either mutation or recombination and the widespread use of antibiotics have been leading to the evolution of resistant bacterial strains. The bacterial genetic variability may also impair the efficacy and development of vaccines and can allow bacteria to escape herd immunity.

Mutations may either be induced by chemicals and other agents or spontaneously occur as a result of faulty DNA replication. Point mutations are changes in single nucleotides. When located in protein-encoding sequences they are resulting either in silent mutations, missense mutations or nonsense mutations by changing the triplet code. While silent mutations do not alter the amino acid sequence of a protein encoded by its gene, missense mutations confer an amino acid replacement and nonsense mutations form a premature stop codon in a gene. Other mutational changes in the DNA may involve insertion, deletion, inversion or replacement of a number of bases.

Transposable elements are sequences of DNA that can change the position within the genome of a single cell and may promote a variety of genetic rearrangements. Insertion sequences are the smallest transposable elements, only encoding functions that are required for the relocation within the bacterial DNA. Larger transposable elements may contain other genes, such as virulence genes and genes encoding antibiotic resistance. Bacterial recombination can take place through three different mechanisms: transformation, transduction and conjugation. Naked DNA can be taken up by certain bacterial species through transformation. New genetic material can also be taken up into a bacterium through transduction by a bacteriophage, making the DNA less vulnerable to deterioration by environmental agents. Bacteriophages are host-specific however, and can usually only move DNA between bacteria of the same or related species. Conjugation is a mechanism of horizontal gene transfer which involves physical contact between donor and recipient cell, mediating the transfer of DNA with high efficency.

## Mycobacterium ulcerans

Buruli ulcer is a disease of skin and soft tissue with the potential to leave sufferers scarred and disabled. *M. ulcerans*, the etiologic agent of Buruli ulcer was discovered by a team of Australian researchers in 1948 [1]. The disease typically occurs in poor rural communities of West and Central Africa. *M. ulcerans* infection is also found in several countries outside Africa, including rural areas of Papua New Guinea, Malaysia, French Guiana, Mexico, as well as Australia [2]. Buruli ulcer is considered to be the third most common mycobacterial disease (http://www.who.int/buruli/information/antibiotics/en/). Partly attributable to the lack of genetic diversity, the exact mode of transmission has remained elusive [3]. *M. ulcerans* has been indicated to have recently evolved via lateral gene transfer and reductive evolution from the environmental species *Mycobacterium marinum*, an ubiquitous pathogen of fish and amphibia [4], to become a niche adapted specialist [5].




Schematic view of an alignment of *M. marinum* M and *M. ulcerans* Agy99 whole genome sequences displayed by the Artemis comparison tool [6]. Regions of conformity are shown in parallel red plains. Inverted DNA segments are depicted in blue.

In aquatic hosts *M. marinum* causes a disseminated granulomatous disease, and in humans *M. marinum* provokes relatively minor granulomatous skin lesions, usually on the cooler extremities of the body [7]. Comparison between the 5.8 Mb genome of the *M. ulcerans* Ghanaian strain Agy99 and the 6.6 Mb genome of the *M. marinum* strain "M" showed that *M. ulcerans* has recently diverged from an *M. marinum* progenitor [8].

*M. ulcerans* has evolved through lateral gene transfer and reductive evolution, the acquisition of a virulence plasmid of the size of 174 kb, required for the production of mycolactone, pseudogene formation, gene rearrangements and gene deletion [5],[9]. Many of these changes have been mediated by some of the 213 copies of *IS 2404* and 91 copies of *IS 2606* [5], neither of which are present in *M. marinum* [8]. Standard molecular typing methods such as multi-locus sequence typing (MLST) and typing of variable numbers of tandem repeats (VNTR) have shown an apparent lack of genetic diversity of *M. ulcerans* within geographic regions [10].

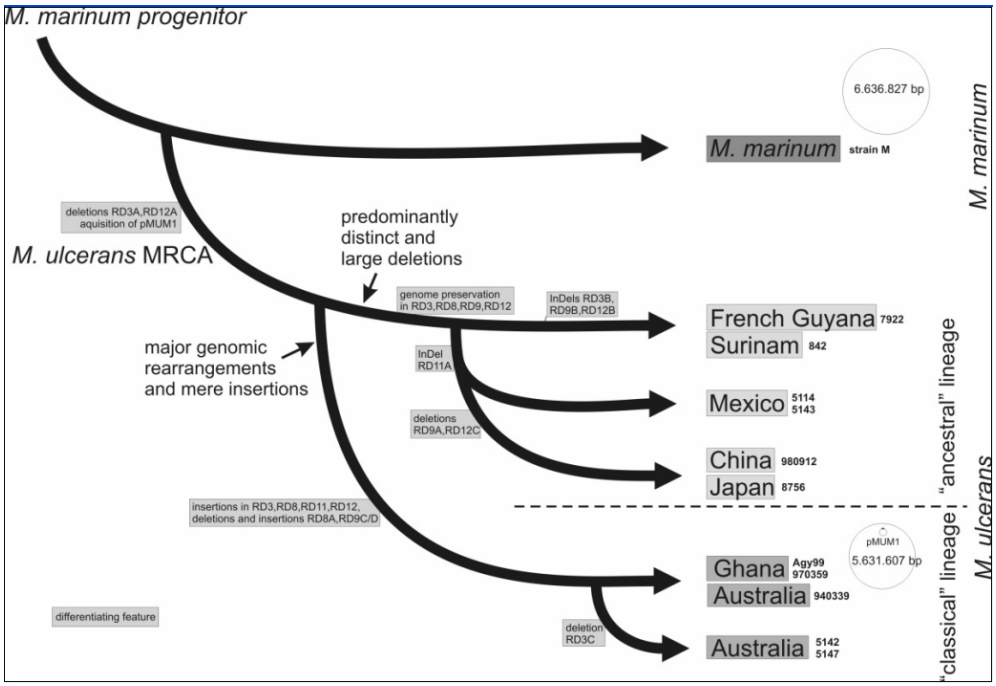



Image was taken from Käser et al. BMC Evol Biol. 2007 Sep 27;7:177. Evolutionary scenario for *M.ulcerans*, basically distinguishing two major lineages. Both the *M. marinum* progenitor and the *M. ulcerans* MRCA (most recent common ancestor) are hypothetical strains.

However, comparative genomic hybridization studies allowed the differentiation of *M. ulcerans* isolates deriving from different regions of the globe. Two distinct *M. ulcerans* lineages could be defined, the ancestral lineage of strains from Asia, South America and Mexico, which are genetically closer to *M. marinum*, and the classical lineage of strains from Africa, Australia and South East Asia [11], [12]. Although strains of *M. ulcerans* from different continents could be well differentiated, the typing of strains within a geographic region has remained a challenge. However, VNTR typing has provided some resolution among clinical isolates of *M. ulcerans* from Africa, confirming the existence of genotypic diversity among African strains [13].

To systemically and comprehensively study the genetic diversity and evolution of *M. ulcerans* strains, two Ghanaian patient isolates from different residential districts and of different VNTR types [13] were selected and their genomes were sequenced using 454 and Solexa technologies, respectively. A Japanese patient isolate was also included as a representative of the ancestral lineage [14]. The genomes were compared with the previously sequenced genome of strain Agy99 that had also been isolated in Ghana [5]. Comparison with the Agy99 reference genome revealed 26,564 SNPs in the Japanese strain. Only 173 SNPs were found when comparing Agy99 with the two other Ghanaian strains. The results of this study indicated that the divergence of the Ghanaian clade of *M. ulcerans* from the Japanese strain may have taken place 394 to 529 thousand years ago, and that the Ghanaian subtypes may have diverged about 1000 to 3000 years ago [14]. A collection of 54 Ghanaian strains was analyzed using the SNPs discovered, and 13 distinct SNP haplotypes could be differentiated [14]. In a follow up study, 74 strains isolated from patients living in the BU endemic Densu river basin in the Ga District of Ghana were analyzed, and 10 different haplotypes could be identified. When 15 strains collected in African countries other than Ghana were typed using the Ghanaian set of SNPs, 13 strains clustered together and differred from all the Ghanian strains, indicating the prevalence of a different SNP pattern. The other 2 strains had SNP patterns similar

to the ones found in Ghana and could be distinguished from each other as well as the other 13 non Ghanaian strains [15].

## Neisseria meningitidis

*Neisseria meningitidis*, a Gram-negative diplococcus, is an obligate human commensal. Although usually carried asymptomatically in the upper airways of healthy individuals, the meningococcus is also a major cause of meningitis and septicaemia. The overall incidence of meningococcal disease in Europe and North America is 1-3 per 100 000 population per year. In the so-called “meningitis belt” in sub-Saharan Africa, extending from Ethiopia to Senegal, annual incidence rates may be as high as 1000 per 100 000 per year during the most severe epidemics [16].

The meningococcal genome has the size of approximately 2.2 Megabases encoding around 2000 genes [17]. Meningococcal populations, especially those isolated from asymptomatic carriers in Europe and North America, have been found to be highly diverse with extensive genetic exchange generating novel combinations of existing genes [18]

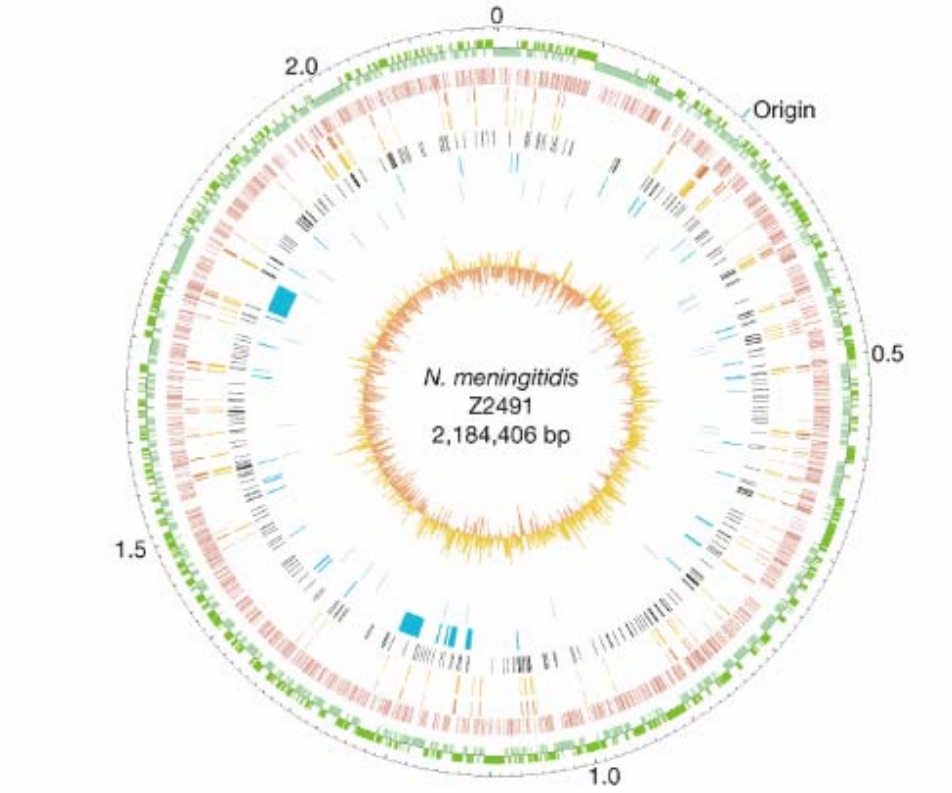
A striking characteristic of the meningococcal genome is the abundance and diversity of repetitive DNA contributing to genome fluidity. About 20% of the meningococcal chromosome consists of repeated sequences of different kinds with the most obvious example being the neisserial DNA uptake sequence (DUS). Nearly 2000 copies of the 12bp uptake sequence could be found in sequenced meningococcal genomes. The so called dRS3 elements, a family of 20 bp repeats with conserved 6 bp terminal inverted repeats occur almost 700 times in the meningococcal genome. Together with the families of 30-160 bp RS elements they make up the “neisserial intergenic mosaic elements” (NIMEs) [19], [20]. It has been shown that the most abundant member of the dRS3 repeat family serves as a target site for chromosomal integration of a filamentous phage [21], and it was suggested that the phage integrase might also catalyze the recombination between dRS3 elements, resulting in permanent genomic changes, such as gene insertions and chromosomal rearrangements [22]. Correia elements (CEs) represent about 2% of the *N. meningitidis* genome. Correia elements are apparently mobile elements

comparable to small insertion sequences (IS) of the size of 100-155bp, but in contrast to conventional IS elements they do not encode a transposase. Insertion sequences and IS remnants are also spread throughout the meningococcal genome promoting genomic variability of *N. meningitidis* [20].

*N. meningitidis* has a large repertoire of phase-variable genes, accounting for almost 4% of all CDs. So called simple sequence repeats or contingency loci comprise short tandem sequence repeats either within or upstream to a coding region. The number of these repeated motifs can be modified during replication through slipped strand mispairing influencing transcription or translation [20]. When tandem repeats occur in the coding sequence, the promoter region or close to the promoter region, they can change the transcriptional and translational state of the gene resulting in phase variation. Slipped-strand mispairing on the synthesis strand during replication generates addition events, whereas slipped strand mispairing on the the template strand induces deletion events [23]. Phase variable genes in meningococci may be involved in biosynthesis and modification of pili, capsular polysaccharide, lipopolysaccharide, opacity proteins, haemoglobin receptors, PorA outer membrane protein, Opc outer membrane protein, ferric receptor, and the putative adhesin NadA [24]. Antigenic variation is a mechanism of immune evasion where only some variants of certain surface components may be expressed. In *N. meningitidis*, antigenic variation occurs in several surface components, including type IV pili, lipooligosaccharides and opa proteins [23]

*N. meningitidis* bacteria are naturally transformable, and DNA may be taken up through transformation and incorporated into the meningococcal chromosome, most likely by homologous recombination [25], [26], [27]. Transformation in *Neisseria* spp requires the presence of a specific DNA uptake sequence (DUS) or uptake signal sequence (USS), respectively, in the incoming DNA, allowing discrimination between DNA from closely related strains or species and foreign DNA. Competent bacteria possess complex machineries to facilitate transformation. *Neisseria* spp express type IV pili that are required for transformation, and also, a type IV system exporting

DNA into the environment has been described in most gonococci and some strains of meningococci. Transformation in the pathogenic *Neisseria* has fuelled high rates of recombination, and it has been estimated that an allele of the *N. meningitidis* genome is ten times more likely to change by recombination than by point mutation [28]. Despite being closely related, *N. meningitidis, N. gonorrhoea*, and *N. lactamica* are highly diverse. For example, at the time of writing (September 2010) the PubMLST database for Neisseria, which catalogues genetically distinct members of the three species as sequence types (STs), listed a total of 8508 unique STs [18]. (http://pubmlst.org/perl/bigsdb/bigsdb.pl?page=downloadProfiles&db=pubmlst_neisseria_seqdef&scheme_id=1).




The image was taken from Parkhill et al., Nature. 2000 Mar 30;404(6777):502-6 Circular representation of the *N. meningitidis* Z2471 genome. The isolate had been sampled in the Gambia in 1983 from an invasive strain. The isolate had the serogroup A and was of the sequence type 4 [29].

The *N. meningitidis* genome is characterized by the horizontal acquisition of multiple genetic islands, acquired from other *N. meningitids* strains, as well as from *N. gonorrhoea* or *N. lactamica.* Genetic islands may also be transferred from other respiratory colonizers [29]. Although it had been thought that transformation is the

major vehicle of lateral gene transfer in Neisseria, recent data show that extensive genetic variation originates from phages and other mobile elements [28], [29].

*N. meningitidis* can be encapsulated or unencapsulated, and there are 13 serogroups based on different capsular polysaccharide structures, but only six serogroups (A, B, C, W-135, X and Y) are currently associated with significant pathogenic potential [30]. Serogroup A strains are responsible for major epidemics and pandemics [19], and the large epidemics in Africa are mainly associated with serogroup A [16].

## Streptococcus agalactiae

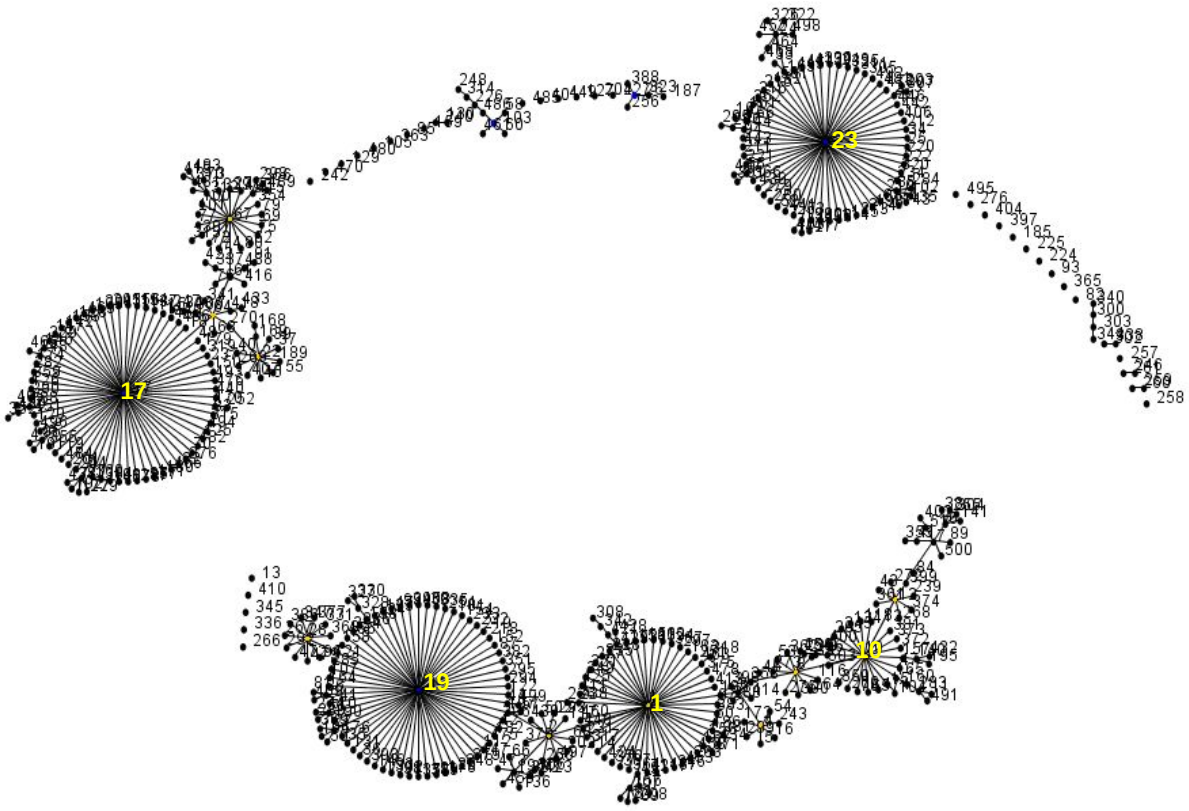
The Group B Streptococcus (GBS), or *Streptococcus agalactiae*, a Gram-positive, β-haemolytic, chain-forming coccus is a commensal of the lower gastrointestinal and genitourinary tracts of 30-50% of healthy adults [31], and an estimated 20 - 30 % of all pregnant women are GBS carriers [32]. However, *S. agalactiae* is also a leading cause of life-threatening bacterial infection in neonates, a cause of invasive infections in the mother, as well as an emerging pathogen of nonpregnant adults, especially the elderly, and persons with underlying conditions such as diabetes and cancer [33], [32].

In newborns, GBS can cause sepsis, pneumonia, meningitis, and less frequently GBS may lead to focal infections such as osteomyelitis, septic arthritis or cellulitis. In pregnant women, GBS may be the cause of urinary tract infection, chorio-amnionitis, endometritis, bacteraemia, and most likely stillbirth [32]. GBS disease in adults includes skin and soft-tissue infection, bacteraemia, urinary tract infections, pneumonia, and osteomyelitis. Meningitis, endocarditits and the streptococcal toxic shock syndrome are rare but serious clinical syndromes of GBS infection [34], [35], [36]. *S. agalactiae* infection may be treated with penicillins and cephalosporins, and GBS remains largely susceptible to β-lactam antibiotics. However, in case of β-lactam allergy, the emergence of widespread resistance to clindamycin and erythromycin poses a serious clinical problem [37].

The genome of the bacterial species *S. agalactiae* has the size of approximately 2.1 to 2.2 Megabases which are encoding around 2100 to 2200 genes.[38] [39] [40]. A bacterial species can be described by its “pan-genome” which includes a core genome containing genes present in all strains and a dispensable genome composed of genes absent from other strains of the same species. It has been proposed that the core genome of *S. agalactiae* consists of 1,806 genes [40], and that the pan-genome is relatively large, exceeding 2,800 genes. *S. agalactiae* has its habitat in both humans and animals, and this broad habitat range may provide a

great available gene pool for lateral gene transfer [41], [42]. It has been demonstrated by experimental and in silico approaches that DNA segments of up to 334 kb can be transferred through conjugation, and that large DNA exchanges may have contributed to the genome dynamics in the natural population [43].

*S. agalactiae* has been described to possess 10 different capsular types [44], and capsular switching may take place by either switching of capsule specific genes, or more often by the exchange of the entire capsular locus. However, capsular switching in *S. agalactiae* has been proposed to be rare [45]. Five serotypes (Ia, Ib, II, III and V) have been described to be primarily prevalent in the US [46], as well as other areas of the world, including the Central African Republic, Senegal [47], England [48], Norway [49], Israel [50] and Korea [51].




eBurst image of 503 sequence types of Streptococcus agalactiae. The founders of the clonal complexes (CCs) CC1, CC10, CC17, CC19 and CC23 are marked.

The development of a multi-locus sequence based typing (MLST) scheme for GBS allows unambiguous comparison of the population structures of GBS strains among different geographical areas. This MLST uncovers sequence variation among seven conserved housekeeping genes, classifying strains into numerous clones, or sequence types (STs) [52]. More than 500 STs are known to date (http://pubmlst.org/perl/mlstdb-net/mlstdbnet.pl?page=download_profiles&file=gbs_profiles.xml) and STs could be grouped together into clusters or clonal complexes (CCs) following phylogenetic analyses [52]. 5 CCs (CC23, CC19, CC17, CC10 and CC1) were highly prevalent in different regions of the globe, such as Sweden [53], Italy [54], England [48], the US [55], the Central African Republic, Senegal [47] as well as Israel.

## References

[1] MacCULLUM P. J pathol bacteriol 1948;60:93-102
.

[2] Johnson PDR, Stinear T, Small PLC, Pluschke G, Merritt RW, Portaels F, Huygen K, Hayman JA, Asiedu K. Plos med 2005;2:e108.

[3] Stragier P, Ablordey A, Bayonne LM, Lugor YL, Sindani IS, Suykerbuyk P, Wabinga H, Meyers WM, Portaels F. Emerging infect. dis 2006;12:844-847.

[4] Stinear TP, Seemann T, Harrison PF, Jenkin GA, Davies JK, Johnson PDR, Abdellah Z, Arrowsmith C, Chillingworth T, Churcher C, Clarke K, Cronin A, Davis P, Goodhead I, Holroyd N, Jagels K, Lord A, Moule S, Mungall K, Norbertczak H, Quail MA, Rabbinowitsch E, Walker D, White B, Whitehead S, Small PLC, Brosch R, Ramakrishnan L, Fischbach MA, Parkhill J, Cole ST. Genome res 2008;18:729-741.

[5] Stinear TP, Seemann T, Pidot S, Frigui W, Reysset G, Garnier T, Meurice G, Simon D, Bouchier C, Ma L, Tichit M, Porter JL, Ryan J, Johnson PDR, Davies JK, Jenkin GA, Small PLC, Jones LM, Tekaia F, Laval F, Daffé M, Parkhill J, Cole ST. Genome res 2007;17:192-200.

[6] Carver TJ, Rutherford KM, Berriman M, Rajandream M, Barrell BG, Parkhill J. Bioinformatics 2005;21:3422-3423.

[7] Pozos TC, Ramakrishan L. Current opinion in immunology 2004;16:499-505.

[8] Stinear TP, Jenkin GA, Johnson PD, Davies JK. J. bacteriol 2000;182:6322-6330.

[9] Demangel C, Stinear TP, Cole ST. Nat. rev. microbiol 2009;7:50-60.

[10] Ablordey A, Swings J, Hubans C, Chemlal K, Locht C, Portaels F, Supply P. J. clin. microbiol 2005;43:1546-1551.

[11] Rondini S, Käser M, Stinear T, Tessier M, Mangold C, Dernick G, Naegeli M, Portaels F, Certa U, Pluschke G. Emerging infect. dis 2007;13:1008-1015.

[12] Käser M, Rondini S, Naegeli M, Stinear T, Portaels F, Certa U, Pluschke G. Bmc evol. biol 2007;7:177.

[13] Hilty M, Yeboah-Manu D, Boakye D, Mensah-Quainoo E, Rondini S, Schelling E, Ofori-Adjei D, Portaels F, Zinsstag J, Pluschke G. J. bacteriol 2006;188:1462-1465.

[14] Qi W, Käser M, Röltgen K, Yeboah-Manu D, Pluschke G. Plos pathog 2009;5:e1000580.

[15] Röltgen K, Qi W, Ruf M, Mensah-Quainoo E, Pidot SJ, Seemann T, Stinear TP, Käser M, Yeboah-Manu D, Pluschke G. Plos negl trop dis 2010;4:e751.

[16] Caugant DA. Infect. genet. evol 2008;8:558-565.

[17] Bentley SD, Vernikos GS, Snyder LAS, Churcher C, Arrowsmith C, Chillingworth T, Cronin A, Davis PH, Holroyd NE, Jagels K, Maddison M, Moule S, Rabbinowitsch E, Sharp S, Unwin L, Whitehead S, Quail MA, Achtman M, Barrell B, Saunders NJ, Parkhill J. Plos genet 2007;3:e23.

[18] Maiden MC. Curr. opin. microbiol 2008;11:467-471.

[19] Parkhill J, Achtman M, James KD, Bentley SD, Churcher C, Klee SR, Morelli G, Basham D, Brown D, Chillingworth T, Davies RM, Davis P, Devlin K, Feltwell T, Hamlin N, Holroyd S, Jagels K, Leather S, Moule S, Mungall K, Quail MA, Rajandream MA, Rutherford KM, Simmonds M, Skelton J, Whitehead S, Spratt BG, Barrell BG. Nature 2000;404:502-506.

[20] Schoen C, Tettelin H, Parkhill J, Frosch M. Vaccine 2009;27 Suppl 2:B103-111.

[21] Kawai M, Uchiyama I, Kobayashi I. Dna research 2006;12:389 -401.

[22] Schoen C, Blom J, Claus H, Schramm-Glück A, Brandt P, Müller T, Goesmann A, Joseph B, Konietzny S, Kurzai O, Schmitt C, Friedrich T, Linke B, Vogel U, Frosch M. Proceedings of the national academy of sciences 2008;105:3473 -3478.

[23] Davidsen T, Tønjum T. Nat. rev. microbiol 2006;4:11-22.

[24] Schoen C, Joseph B, Claus H, Vogel U, Frosch M. Int. j. med. microbiol 2007;297:601-613.

[25] CATLIN BW. Science 1960;131:608-610.

[26] Sun Y, Exley R, Li Y, Goulding D, Tang C. J. bacteriol 2005;187:3273-3276.

[27] Maiden MC. Clin. infect. dis 1998;27 Suppl 1:S12-20.

[28] Treangen TJ, Ambur OH, Tonjum T, Rocha EP. Genome biol 2008;9:R60-R60.

[29] Dunning Hotopp JC, Grifantini R, Kumar N, Tzeng YL, Fouts D, Frigimelica E, Draghi M, Giuliani MM, Rappuoli R, Stephens DS, Grandi G, Tettelin H. Microbiology (reading, engl.) 2006;152:3733-3749.

[30] Stephens DS, Greenwood B, Brandtzaeg P. Lancet 2007;369:2196-2210.

[31] van der Mee-Marquet N, Fourny L, Arnault L, Domelier A, Salloum M, Lartigue M, Quentin R. J clin microbiol 2008;46:2906-2911.

[32] Gibbs RS, Schrag S, Schuchat A. Obstet gynecol 2004;104:1062-1076.

[33] Rajagopal L. Future microbiology 2009;4:201-221.

[34] Skoff TH, Farley MM, Petit S, Craig AS, Schaffner W, Gershman K, Harrison LH, Lynfield R, Mohle Boetani J, Zansky S, Albanese BA, Stefonek K, Zell ER, Jackson D, Thompson T, Schrag SJ. Clinical infectious diseases 2009;49:85-92.

[35] Farley MM. Clin. infect. dis 2001;33:556-561.

[36] Ulett KB, Benjamin WH, Zhuo F, Xiao M, Kong F, Gilbert GL, Schembri MA, Ulett GC. J clin microbiol 2009;47:2055-2060.

[37] Castor ML, Whitney CG, Como-Sabetti K, Facklam RR, Ferrieri P, Bartkus JM, Juni BA, Cieslak PR, Farley MM, Dumas NB, Schrag SJ, Lynfield R. Infect dis obstet gynecol 2008;2008.

[38] Tettelin H, Masignani V, Cieslewicz MJ, Eisen JA, Peterson S, Wessels MR, Paulsen IT, Nelson KE, Margarit I, Read TD, Madoff LC, Wolf AM, Beanan MJ, Brinkac LM, Daugherty SC, DeBoy RT, Durkin AS, Kolonay JF, Madupu R, Lewis MR, Radune D, Fedorova NB, Scanlan D, Khouri H, Mulligan S, Carty HA, Cline RT, Van Aken SE, Gill J, Scarselli M, Mora M, Iacobini ET, Brettoni C, Galli G, Mariani M, Vegni F, Maione D, Rinaudo D, Rappuoli R, Telford JL, Kasper DL, Grandi G, Fraser CM. Proceedings of the national academy of sciences of the united states of america 2002;99:12391 -12396.

[39] Glaser P, Rusniok C, Buchrieser C, Chevalier F, Frangeul L, Msadek T, Zouine M, Couvé E, Lalioui L, Poyart C, Trieu-Cuot P, Kunst F. Mol. microbiol 2002;45:1499-1513.

[40] Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJB, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. Proc. natl. acad. sci. u.s.a 2005;102:13950-13955.

[41] Lefébure T, Stanhope MJ. Genome biol 2007;8:R71.

[42] Marri PR, Hao W, Golding GB. Mol. biol. evol 2006;23:2379-2391.

[43] Brochet M, Rusniok C, Couvé E, Dramsi S, Poyart C, Trieu-Cuot P, Kunst F, Glaser P. Proc. natl. acad. sci. u.s.a 2008;105:15961-15966.

[44] Slotved H, Kong F, Lambertsen L, Sauer S, Gilbert GL. J. clin. microbiol 2007;45:2929-2936.

[45] Martins ER, Melo-Cristino J, Ramirez M. J bacteriol 2010;192:1361-1369.

[46] Edwards MS. Hum vaccin 2008;4:444-448.
[47] Brochet M, Couvé E, Bercion R, Sire J, Glaser P. J clin microbiol 2009;47:800-803.

[48] Jones N, Oliver KA, Barry J, Harding RM, Bisharat N, Spratt BG, Peto T, Crook DW. Clin. infect. dis 2006;42:915-924.

[49] Bergseng H, Afset JE, Radtke A, Loeseth K, Lyng RV, Rygg M, Bergh K. Clin. microbiol. infect 2009;15:1182-1185.

[50] Bisharat N, Jones N, Marchaim D, Block C, Harding RM, Yagupsky P, Peto T, Crook DW. Microbiology (reading, engl.) 2005;151:1875-1881.

[51] Seo YS, Srinivasan U, Oh K, Shin J, Chae JD, Kim MY, Yang JH, Yoon H, Miller B, DeBusscher J, Foxman B, Ki M. J. korean med. sci 2010;25:817-823.

[52] Jones N, Bohnsack JF, Takahashi S, Oliver KA, Chan M, Kunst F, Glaser P, Rusniok C, Crook DWM, Harding RM, Bisharat N, Spratt BG. J clin microbiol 2003;41:2530-2536.
[53] Luan S, Granlund M, Sellin M, Lagergård T, Spratt BG, Norgren M. J. clin. microbiol 2005;43:3727-3733.

[54] Gherardi G, Imperi M, Baldassarri L, Pataracchia M, Alfarone G, Recchia S, Orefici G, Dicuonzo G, Creti R. J. clin. microbiol 2007;45:2909-2916.

[55] Bohnsack JF, Whiting A, Gottschalk M, Dunn DM, Weiss R, Azimi PH, Philips JB, Weisman LE, Rhoads GG, Lin FC. J. clin. microbiol 2008;46:1285-1291.

# CHAPTER 2: Genetic diversity of *Mycobacterium ulcerans*

# Multiple Loss of Immunogenic Proteins in *Mycobacterium ulcerans* Suggests Immune Evasion

Charlotte A. Huber, Marie-Thérèse Ruf, Gerd Pluschke, and Michael Käser*

Swiss Tropical Institute, Socinstr. 57, 4002 Basel, Switzerland

**Running Head**
Loss of immunogenic proteins in *Mycobacterium ulcerans*

*) Corresponding author:
Michael Käser, Swiss Tropical Institute, Molecular Immunology, Socinstr. 57, 4002 Basel, Switzerland Phone: +41-61-2848-171. Fax: +41-61-2848-101.
E-mail: m.kaeser@unibas.ch

Published in the journal of Clinical and Vaccine immunology

## Abstract

The highly immunogenic mycobacterial proteins ESAT-6, CFP-10 and HspX represent potential target antigens for the development of subunit vaccines and immunodiagnostic tests. Recently, the complete genome sequence revealed the absence of these coding sequences in *Mycobacterium ulcerans*, causative agent of the emerging human disease Buruli ulcer. Genome reduction and the acquisition of a cytopathic and immunosuppressive macrolide toxin plasmid are regarded crucial for the emergence of this pathogen from its environmental progenitor, *Mycobacterium marinum*. Earlier, we have shown the evolution of *M. ulcerans* into two distinct lineages. Here we show that while the genome of *M. marinum* M contains two copies of the *esxB-esxA* gene cluster at different loci, both copies are deleted from the genome of *M. ulcerans* strains belonging to the classical lineage. Members of the ancestral lineage instead have lost these gene clusters either by newly identified genomic insertional-deletional events or by conversions of functional genes to pseudogenes via point mutations. Thus, the *esxA* (ESAT-6), *esxB* (CFP-10) and *hspX* genes are located in hot spot regions for genomic variation where functional disruption seems to be favored by selection pressure. Our detailed genomic analyses have identified a variety of independent genomic changes that have led to loss of expression of functional ESAT-6, CFP-10, and HspX proteins. Loss of these immunodominant proteins may help bypassing the host's immunological response and represents part of an ongoing adaptation of *M. ulcerans* to survival in host environments that are screened by immunological defense mechanisms.

## Introduction

The emerging pathogen *Mycobacterium ulcerans* is the causative agent of Buruli ulcer, a mycobacterial disease of skin and soft tissue with the potential to leave sufferers scarred and disabled. While it is endemic in more than 30 countries (26), the major disease burden lies on children living in poor rural communities of West-Africa. Buruli ulcer is prevalent in riverine, slow-flowing and swampy areas, but the exact mode of transmission has remained elusive. This is partly attributable to a clonal population structure and an associated lack of high-resolution genetic fingerprinting methods for micro-epidemiologic studies.

*M. ulcerans* seems to have recently evolved via lateral gene transfer and reductive evolution from the fish disease causing environmental species *M. marinum* (40,43). Particularly, it has acquired the virulence plasmid, pMUM001, encoding the genes for the synthesis of the macrolide toxin, mycolactone. This toxin has cytopathic and immunomodulatory properties and plays a decisive role in producing an extracellular infection after an initial phase within macrophages (4,41,42,47). In addition, *M. ulcerans* has undergone extensive gene loss due to DNA deletions, DNA rearrangements, and pseudogene formation which apparently drives its evolution towards a niche adapted specialist (27,34,39). Previous findings suggest that *M. ulcerans* lineages from different geographic areas reveal variations in virulence (27,32), and F. Portaels, submitted).

The ESX-1 secretion system is required for the virulence of *M. tuberculosis* and related pathogenic mycobacteria. It comprises the 6 kDa early secretory antigenic protein (ESAT-6) and the 10 kDa culture filtrate protein (CFP-10) which are among the strongest T-cell response elicitors in tuberculosis patients (7,8). The genes encoding these proteins are localized on the region of difference 1 (RD1) locus which is intact in virulent members of the *M. tuberculosis* complex, but absent from the attenuated vaccine strain *M. bovis* BCG (□$RD1^{BCG}$) (21,29). Similarly, the vole bacillus, *M. microti*, was found to have a natural deletion (□$RD1^{microti}$) overlapping

with deletion □RD1$^{BCG}$ (6,18). The so called extended RD1 encompasses most of the genes that form the ESX-1 secretion apparatus (7,16,17) or are crucial for both

ESAT-6/CFP-10 secretion and virulence (7,17,19,31). This secretion apparatus enhances virulence in *M. tuberculosis* and *M. marinum* infection by secretion of effector proteins into the cytosol of infected macrophages (37), prevention of phagolysosomal maturation (28,45) and cytolytic activity (24). On the other hand, infected individuals develop strong T-cell responses against these proteins, which seem to be relevant for immune protection (8). The 16-kDa heat shock protein HspX or □-crystallin-like protein (Acr), a dominant protein expressed during static growth in *M. tuberculosis*, is required for mycobacterial persistence within the macrophage. It is yet another potent immune response elicitor and suitable for detection of *M. tuberculosis* infection (14,15,20,25,35,49).

In mycobacterial disease control, highly antigenic proteins serve both as targets for diagnostic tests and as candidate proteins for vaccine development (1,8,30). While being present in the sequenced *M. marinum* strain M (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/m_marinum), genes encoding ESAT-6, CFP-10 and HspX are absent from the genome of the sequenced Ghanaian *M. ulcerans* strain Agy99 (http://genopole.pasteur.fr/Mulc/BuruList.html). However, earlier data showed that some *M. ulcerans* isolates and other related mycolactone producing mycobacteria harbor at least segments of these genes (32,48). Recently, we have identified two distinct genetic lineages of *M. ulcerans*, with representatives of the ancestral lineage being phylogenetically closer to its progenitor, *M. marinum*, than members of the *M. ulcerans* classical lineage (27). Here, we have analyzed a world wide collection of *M. ulcerans* strains belonging to these two lineages for the presence of *esxA*, *esxB* and *hspX* and their surrounding genomic regions.

## Materials and Methods

### Mycobacterial strains and genomic DNA extraction

*M. marinum* strain M was used for interspecies comparison. A world wide strain collection of *M. ulcerans* was used earlier for investigation of genomic strain variations (34). Although several attempts to differentiate these strains revealed low resolution (2,3,11,22,23,38,44), this collection of patient isolates was shown to be divided in two lineages displaying major genomic differences (27). In this study, we used *M. ulcerans* clinical isolates of both lineages as follows. For the classical lineage: Ghana Agy99, Ghana ITM 970321, Ghana ITM 970359, Ghana ITM 970483, Ivory Coast ITM 940662, Ivory Coast ITM 940815, Ivory Coast ITM 940511, Benin ITM 970111, Benin ITM 940886, Benin ITM 940512, Benin ITM 970104, Democratic Republic of Congo (DRC) ITM 5150, DRC ITM 5151, DRC ITM 5155, Togo ITM 970680, Angola ITM 960657, Angola ITM 960658, Papua New Guinea (PNG) ITM 941331, PNG ITM 9537, Malaysia ITM 941328, Australia ITM 941324, Australia ITM 941325, Australia ITM 941327, Australia ITM 9549, Australia ITM 9550, Australia ITM 8849, Australia ITM 940339, Australia ITM 5142, and Australia ITM 5147. For the ancestral lineage: China ITM 980912, Japan ITM 8756, French Guiana ITM 7922, Surinam ITM 842, and Mexico ITM 5143. Presence of the specific PCR products obtained with primer pairs CH1/CH4 and CH3/CH4 (that exclude each other by design, see Fig. 1) occurred concomitantly in the strains ITM 5151 DRC and ITM 941331 PNG. Since also VNTR typing analysis indicated that these strains are impure, we excluded these strains from further analysis.

Bacterial pellets of about 60 mg (wet weight) were heat inactivated for 1 hour at 95°C in 500 µl extraction buffer (50 mM Tris-HCl, 25 mM EDTA, 5% monosodium glutamate), and sequentially treated with lysozyme (2 h, 37°C, 17 M lysozyme) and proteinase K (overnight, 45°C, 0.3 M proteinase K in proteinase K buffer: 1 mM Tris-HCl, 5 mM EDTA, 0.05% SDS, pH7.8). After digestion, the samples were subjected to bead beater treatment (7 min, 3000 rpm, Mikro-Dismembrator, B. Braun Biotech International, Melsungen, Germany) with 300 µl of 0.1 mm zirconia beads (BioSpec

Products, Bartlesville, OK, USA). DNA was extracted from the supernatants by phenol-chloroform (Fluka, Buchs, Switzerland) extraction and subjected to ethanol

precipitation. DNA concentration was measured by optical density at 260 nm (GeneQuant spectrophotometer, Pharmacia Biotech, Cambridge, UK).

**DNA methods**

PCR was performed using FirePol 10x BD buffer and 0.5 µl FirePolTaq-Polymerase (Solis BioDyne, Tartu, Estonia), 2.5 ng genomic DNA or the according volume of RNAse free water as a negative control, 0.6 µM forward and reverse primers each, 1.7 mM $MgCl_2$ and 0.3 mM of each dNTP in a total volume of 30 µl. PCR reactions were run in a GeneAmp PCR System 9700 PCR machine. The thermal profile for PCR amplification of *M. ulcerans* genomic DNA included an initial denaturation step of 95-98$^{o}$C for 3 min, followed by 32 cycles of 95$^{o}$C for 20 sec, annealing at 58-65$^{o}$C for 20 sec, and elongation at 72$^{o}$C for 30 sec up to 4min. The PCR reactions were finalized by an extension step at 72$^{o}$C for 10 min. For experiments with more than 30 samples Hot Star Taq® (QIAGEN AG, Hombrechtikon, Switzerland) was used according to the manufacturer's protocol. In order to retrieve PCR products that were subsequently subjected to sequencing, iProof™ High fidelity DNA Polymerase (Bio-Rad Laboratories, Hercules, CA) was used. PCR products were analyzed on 1-2% agarose gels by gel electrophoresis using ethidium bromide staining and the AlphaImager illuminator and AlphaImager software (Alpha Innotech, San Leandro, CA, USA). Primers as summarized in table 1 were designed using the Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). PCRs fragments produced for analysis of unknown genomic sequences were purified using the NucleoSpin purification kit (Machery-Nagel GmbH & Co. Ko, Düren, Germany) and subjected to direct sequencing or cloned using the TOPO TA Cloning® Kit (Corporate Headquarters, Invitrogen Corporation, Carlsbad, CA, USA), transformed into JM109 (Sigma Aldrich, Buchs, Switzerland) bacterial cells, and sequenced after DNA preparation (Miniprep-Kit, Sigma Aldrich, Buchs, Switzerland). VNTR undertaken for confirmation of strain identities was performed according to (44).

Sequencing was performed using the Big Dye kit and the AbiPrism310 genetic sequence analyzer (Perkin-Elmer, Waltham, MA, USA). All gene sequences were reproduced and subjected to alignment and comparison with the AbiPrism Autoassembler version 1.4.0 (Perkin-Elmer, Waltham, MA, USA).

**Data analyses and bioinformatics**

Retrieved sequences were compared to the BuruList (http://genopole.pasteur.fr/Mulc/BuruList.html) and the *M. marinum* (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/m_marinum) blast servers and analyzed using the sequence manipulation suite (http://bioinformatics.org/sms/index.html), the sequence alignment tool blast 2 sequences (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi), the multiple sequence alignment website Multalin (http://bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html) and the Artemis software release 9 (The Wellcome Trust Sanger Institute, Hinxton, UK; (36)). The sequences for *M. tuberculosis* were retrieved from the following web page: (http://www.sanger.ac.uk/Projects/M_tuberculosis). Linear genomic comparison was performed using the Artemis Comparison Tool software release 6 (9).

**Accession Numbers**

The GenBank (http://www.ncbi.nlm.nih.gov/Genbank/index.html) accession numbers for sequences from the following *M. ulcerans* strains are: Japan 8756: *CFP-10*, EU257146; *ESAT-6*, EU257151; *HspX/Acr*, EU257156; China 980912: *CFP-10*, EU257147; *ESAT-6*, EU257152; *HspX/Acr*, EU257157; Surinam 842 *CFP-10*, EU257148; *ESAT-6*, EU257153; *HspX/Acr*, EU257158; French Guiana 9722: *CFP-10*, EU257149; *ESAT-6*, EU257154; *HspX/Acr*, EU257159; Mexico 5143: *CFP-10*, EU257150; *ESAT-6*, EU257155; *HspX/Acr*, EU257160

## Results

### Presence of *esxB/esxA* in *M. ulcerans* strains of the ancestral lineage

Blast searches of the partially annotated genome of *M. marinum* M (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/m_marinum) showed that this strain contains two copies of the *esxB (CFP-10)-esxA (ESAT-6)* gene cluster. Both copies are deleted in the genome of the African *M. ulcerans* isolate Agy99 (43). The corresponding two regions of difference (RDs) between the genome sequences of the two mycobacterial species have been designated MURD152 (*M. marinum* genome position 6489253-6592034) and MURD4 (*M. marinum* genome position 218302-230285) (43).

Compared to *M. marinum* M, the *M. ulcerans* Agy99 genome has a 2.8 kb deletion in MURD152, which is associated with a large inversion at the 5' end of the deletion (Fig. 1A). To test whether all *M. ulcerans* lineages share this genome constellation in MURD152, we screened a comprehensive *M. ulcerans* strain collection of world-wide origin by PCR analysis using a primer pair (CH3 and CH4) that yields a PCR product of 162 bp only when MURD152 is deleted and flanked by the inverted sequence (Fig. 1A and B). Whereas members of the ancestral lineage (strains from Asia, South America and Mexico) were negative, members of the classical lineage (strains from Africa, Papua New Guinea, Malaysia and Australia) were positive, except for strain Australia 9549 which has a larger deletion in this region (see below). Likewise, a PCR using a primer pair (CH8 and CH9) specific for the sequence constellation of strain Agy99 in MURD4 revealed a PCR product of 1712 bp only for representatives of the classical but not for members of the ancestral lineage (Fig. 1B), demonstrating genomic diversity between the two *M. ulcerans* lineages in this locus.

A PCR with primers (CH1 and CH2) corresponding to the 5' end of the *esxB* coding sequence and the 3' end of *esxA* coding sequence (Fig. 1A) yielded a PCR product of the expected size of 610 bp with genomic DNA from the *M. marinum* control as well as in all *M. ulcerans* strains belonging to the ancestral lineage (Fig. 2). Primers

corresponding to the flanking regions of either the MURD4 or the MURD152 associated *esxB-esxA* gene cassette were used to analyze for the localization of this cluster in the genome of these *M. ulcerans* strains (Fig. 2). Results indicated that *esxB-esxA* of the Asian and South American strains is located in MURD152, whereas in the Mexican strain the gene cluster is located in MURD4 (Fig. 2). These localizations were verified by PCR analyses extending several kilobases further into the flanking regions. While in the Asian and South American haplotypes the respective *M. marinum* MURD152 genome constellations were found, the cluster was flanked in the case of the Mexican haplotype by the MURD4 associated sequences of *M. marinum*.

**Unique deletions in MURD152 in strains 5143 from Mexico and 9549 from Australia**

While the MURD152 *esxB-esxA* is deleted in the Mexican strain 5143 (Fig. 2), no PCR product specific for the MURD152 constellation of the strains belonging to the classical lineage was obtained with primers CH3 and CH4 (Fig. 1B), giving evidence for a larger deletion. A PCR analysis with primers corresponding to different positions of the genomic sequences flanking MURD152 demonstrated that strain Mexico 5143 has a deletion, (designated □RD13A; Fig. 3), that is substituted by an IS*2404* element. This InDel event can have occurred either from an *M. marinum* M like genome constellation or from an *M. ulcerans* Agy99 like constellation (loss of 41.8 kb or of 8 kb, respectively). The DNA sequences flanking □RD13A in the Mexican strain have a slightly higher identity to the corresponding sequence stretches of *M. ulcerans* Agy99 than to those of *M. marinum* M (98% vs. 94% over 986 bp).

Failure to obtain a PCR product with both the CH1/CH2 and the CH3/CH4 PCR primers for the Australian strain 9549 (Fig. 1) provided evidence for yet another deletion type within the MURD152 region. PCR analysis using primers located in the sequences flanking the corresponding region in the *M. ulcerans* Agy99 genome led to the characterization of a deletion of 13662 bp (□RD13B; Fig. 3) including an

IS*2404* element on each of the ends of the deleted DNA segment. The deleted DNA stretch was substituted by an IS*2404* element that, upon sequence analysis, clustered to neither of the deleted versions of IS*2404*.

**Sequence variation in ESAT-6 and CFP-10**

PCR products obtained with primers corresponding to MURD locus-specific flanking regions and comprising the respective *esxB-esxA* clusters (Fig. 2) were sequenced. Deduced amino acid sequences of all versions of *M. ulcerans* ESAT-6 and CFP-10 encoded in MURD4 (Mexico 5143) or MURD152 (South American and Asian strains) were compared with the *M. marinum* M sequences in the two loci (Fig. 4 and supplementary material). As expected, the translated ESAT-6 amino acid sequence of the Mexican strain clustered to and was identical with the MURD4-associated *M. marinum* M sequence (Fig. 4B). While the four MURD152-associated *M. ulcerans* ESAT-6 sequences of the Asian and the South American strains were identical among each other, their amino acid sequences differed at six positions from the MURD152-associated *M. marinum* sequence, but only at two positions from the MURD4-associated *M. marinum* sequence (Fig. 4B). At nucleotide level, the *esxA* gene of the Asian and South American strains appear as hybrids composed of an *M. marinum* MURD4 sequence stretch at the 5' and a MURD152 stretch at the 3' end.

The two *M. marinum esxB* genes differ only at three nucleotide positions at the 5' end (Fig. 4A), encoding CFP-10 proteins with identical deduced amino acid sequences (Fig. 4B). The *esxB* gene of the Mexican strain differed at four positions from the *M. marinum* M MURD4 locus, but only at one position from the MURD152 locus. While the *esxB* gene sequences of the South American *M. ulcerans* strains were identical to the MURD152 associated sequence, a frameshift mutation has converted *esxB* of the Asian strains to a pseudogene (Fig 4B).

**Lack of the immunodominant HspX/Acr protein in the classical lineage of *M. ulcerans***

Next we screened the world-wide *M. ulcerans* strain collection for the presence of the CDS encoding the immunogenic protein HspX (Acr) located in MURD92 (*M. marinum* genome position 4271366-4313737; (43)). Using primers (CH14 and

CH15) corresponding to the *hspX* flanking regions, a PCR product of 791 bp comprising the complete *hspX* gene was obtained for all members of the ancestral lineage, but for none of the strains belonging to the classical lineage (not shown). Instead, amplification of a 469 bp PCR product (primers CH16 and CH17) obtained with a complementary PCR again demonstrated the presence of the Agy99 genome constellation (related to the MURD92 deletion) in all members of the classical lineage. While strains coming from the same geographical area had identical gene sequences, Asian and South American sequences differed slightly from each other and from the *M. marinum* sequence (Fig. 5A and supplementary material). In the case of the Mexican strain, nucleotide insertions resulted in a frameshift mutation leading to a truncated translation product (Fig. 5B).

## Discussion

The *M. tuberculosis* proteins ESAT-6, CFP-10 and HspX are strong T- and B-cell immunogens. This makes them to suitable targets for immunodiagnostic tests (7,8,14,15,20) and potentially also for subunit vaccine development (1,30,35). These approaches cannot be duplicated for Buruli ulcer, since these proteins are not expressed by *M. ulcerans* strains belonging to the classical lineage that are found in the endemic areas of Africa and Australia and are responsible for the vast majority of clinical cases world wide.

The genome of the *M. marinum* strain M harbors two *esxB/esxA* gene clusters at distant chromosomal locations, one in MURD4 and the other in MURD152. Such duplications are common for proteins of the esx protein family (46). In this report we demonstrate that all analyzed *M. ulcerans* strains belonging to the ancestral lineage have lost only one copy of the *esxB/esxA* cassette, the Asian and South American strains the MURD4 copy and the Mexican strain the MURD152 copy, respectively. Furthermore, a frameshift mutation has converted the remaining *esxB* gene of the Asian strains to a pseudogene. The basis for the high identity of the N-terminal *esxA* nucleotide sequence located in the MURD152 locus in the South American and Asian haplotypes with the *M. marinum* MURD4 sequence is unclear, but implies a history of homologous recombination between the two copies of *esxB* and *esxA* genes before loss of the MURD4 region. Members of the classical lineage have lost both copies, probably in a bottleneck situation that forged this lineage.

Since MURD152, 92 and 4 do not only show genomic differences between *M. marinum* and *M. ulcerans*, but also within *M. ulcerans* strains, we designated these regions of difference RD13, RD14 and RD15, respectively, in continuation of the previously assigned RDs within the species *M. ulcerans* (34). A detailed alignment of the chromosomal organization in RD13, which corresponds to RD1 in *M. tuberculosis*, is shown in Fig. 3. These RDs represent hot spots of genetic variation potentially suitable to perform genetic fingerprinting of *M. ulcerans*.

In addition to the previously identified five *M. ulcerans* InDel haplotypes (27,34) strain Australia 9549 was identified to represent a sixth InDel haplotype which is defined by □RD13B.

Alone in MURD152 at least three different deletion events are responsible for the InDel diversity within *M. ulcerans* (Table 2). When this region was analyzed for variations among a collection of mycolactone producing mycobacteria, an unclear situation was suggested for a Mexican strain (48). Here, we show that the InDel of 8kb substituted by an IS*2404* element (□RD13A) in the Mexican strain (or 41.8 kb with respect to the *M. marinum* backbone) differs from the MURD152 deletion in Agy99. This deletion is independent of yet another extended deletion of 13.7 kb (□RD13B) in this genomic region in strain Australia 9549. The latter deletion is also substituted by an IS*2404* element and displays a second large sequence polymorphism within Australian isolates, after the earlier described RD3 (27,34). It will be worth investigating the distribution of this InDel polymorphism within a collection of Australian *M. ulcerans* isolates using primer pair combinations CH10/CH11 demonstrating the presence of the □RD13B deletion and both CH10/12 and CH13/CH11 displaying positive for strains with the sequence configuration of Agy99.

The described deletions encompass also CDSs surrounding the *esxA*, *esxB* and *hspX* genes, indicating loss or modification of molecular apparatuses or pathways. First, PE35, essential for secretion (7), was lost in both MURD152 and □RD13A and is also commonly deleted in □RD1$^{BCG}$ and □RD1$^{microti}$ (Fig. 3). Second, many of the genes of the ESX-1 secretion system (genes Rv3866/MMAR_5441 through Rv3881/MMAR_5457/espB, corresponding to extRD1) are equally affected by deletions □RD13A and/or MURD151 through MURD153, namely the AAA protein family member Rv3868/MMAR_5443, Rv3871/MMAR_5446, and Rv3877/MMAR_5452 (7,17,19). Members of the classical lineage omit an MMAR_5457 orthologue in MURD153 which was recently described a secreted

product and renamed espB (31). Also in MURD92, HspX was jointly deleted with the co-regulated Rv2032/nitroreductase gene (33).

As for ESAT-6 and CFP-10, we also found for HspX different genetic mechanisms that have led to loss of expression, comprising both deletions of genomic sequences and single base differences (Table 2). Many of the sequence variations across the *M. ulcerans* haplotypes leading to loss (of function) of these highly immunogenic proteins appear to have emerged independently of each other. This may indicate a counterselection for expression of these proteins. HspX seems to be a negative growth regulator involved in hypoxic shiftdown to promote non-replicating persistence of *M. tuberculosis* (15,20,25). Both ESAT-6 and CFP-10 were shown to be virulence factors of *M. tuberculosis*, and their loss reduces infectivity due to dysfunction of the ESX-1 secretion apparatus (5,10,12,13). The mycolactone producing and largely extracellular *M. ulcerans* has a profoundly different survival strategy in mammalian hosts than the intracellular *M. tuberculosis* has, therefore it is most likely that its pathogenicity for mammalian hosts is due to other virulence factors. Thus, our data suggest that functional disruption or complete loss of major targets of the immune response may confer a selective advantage to this emerging pathogen. Still, it is currently not clear, whether pathogenicity for mammalian hosts, i.e. shedding into the environment from chronic wounds, contributes significantly to survival of the species *M. ulcerans*. However, the observed loss of expression of highly immunogenic proteins caused by a variety of genomic changes may represent an indication that immune selection plays a role in the adaptation of *M. ulcerans* to a more stable environment.

**Abbreviations:**

RD – region of difference (including a sequence locus in which several genomic events may have led to various configurations)

InDel – Insertion-deletion (an event that includes an insertion substituting a deleted sequence in contrast to an insertion or a deletion only)

CDS – coding sequence

ISE – insertion sequence element (for *M. ulcerans*, two transposable elements are known as: IS*2404* and IS*2606*)

**Acknowledgments**

We gratefully acknowledge F. Portaels for provision of most of the *M. ulcerans* strains included in this study, P. C. Small for provision of the *M. marinum* strain M, T. Stinear and J. Parkhill for providing the *M. marinum* gene annotation ahead of publication, and C. Daubenberger for stimulating discussion.

**Funding**

This work was partially supported by the Stanley Thomas Johnson Foundation. M. Käser was supported by a research grant from the Deutsche Forschungs-gemeinschaft, KA 1842/1-1.

**Figure Legends**

**Fig. 1**: Confirmation of the MURD specific deletions affecting *esxB (CFP10)* and *esxA (ESAT6)* in an *M. ulcerans* world-wide strain collection. **A**: Schematic view of an alignment of *M. marinum* M (upper bar) and *M. ulcerans* Agy99 (lower bar) genomic sequences displayed by the Artemis Comparison Tool (9). Regions of conformity are shown in parallel grey plains, an inverted DNA segment is depicted as an inverted surface, and white areas represent unique sequences like MURD152 which is present only in *M. marinum* M but deleted from *M. ulcerans* Agy99. Indicated are the genes *esxB* and *esxA* and the PCR primers (CH1 through CH4) used for this experiment. **B**: PCR products of 162 bp or 1712 bp proofed the MURD152 deletion of 2.8 kb and the MURD4 deletion of 12 kb, respectively.

**Fig. 2**: **Localization of the two *esxB-esxA* clusters in the genomes of strains of the *M. ulcerans* ancestral lineage**. Positions of the corresponding primers are indicated for the PCR product of the *esxB-esxA* cluster where CH1 and CH2 correspond to sequences within the CDSs of both locations, and of the slightly larger PCR products amplified with flanking primers specific for either MURD152 or MURD4 (primers see Table 1).

**Fig. 3: Chromosomal organization of CDSs in RD13 including deletional variations between *M. ulcerans* and other mycobacteria**. Gene names are indicated for *M. tuberculosis* (http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj&cmd=Retrieve&dopt=Overview&list_uids=224), *M. marinum* (http://www.sanger.ac.uk/Projects/M_marinum/), and *M. ulcerans* (http://www.ncbi.nlm.nih.gov/sites/entrez?Db=genomeprj&cmd=ShowDetailView&TermToSearch=16230) and orthologous genes are aligned. RD13 of *M. ulcerans* corresponds to RD1 in *M. tuberculosis*. Deletions in *M. bovis* BCG, *M. microti* and various *M. ulcerans* strains are indicated by solid bars as marked.

**Fig. 4: Nucleotide variations (A) and amino acid sequence alignments (B) in CFP-10 and ESAT-6 CDSs and their gene products.** Position 1 of the nucleotide alignment reflects the start of gene *esxB*. For the DNA sequences, only differing nucleotides are shown (positions as indicated). For whole sequence alignments see supplementary

material. Orthologous sequences of *M. tuberculosis* H37Rv and *M. bovis* AF2122/97 are included in the amino acid alignments.

**Fig. 5: Nucleotide variations (A) and amino acid sequence alignments (B) in the HspX CDS and its gene product.** Position 1 of the nucleotide alignment reflects the start of gene *hspX*. For the DNA sequences, only differing nucleotides are shown (positions as indicated). For whole sequence alignments see supplementary material. Orthologous sequences of *M. tuberculosis* H37Rv and *M. bovis* AF2122/97 are included in the amino acid alignments.

**Supplementary material**

**Figure S1**: Nucleotide sequence alignment (using Multalin) of the *esxB-esxA* cluster in the two *M. marinum* loci and the *M. ulcerans* strains Surinam 842, French Guiana 7922, Japan 8756, China 980912 and Mexico 5143 in comparison with the respective *M. tuberculosis* and *M. bovis* CDSs. Nucleotides of CDSs are at the positions 9-309 (*esxB*) and 348-633 (*esxA*).

**Figure S2**: Nucleotide sequence alignment (using Multalin) of *hspX* in the two *M. marinum* loci and the *M. ulcerans* strains Surinam 842, French Guiana 7922, Japan 8756, China 980912 and Mexico 5143 in comparison with the respective *M. tuberculosis* and *M. bovis* CDSs. Nucleotides of the CDS are at position 192-623 (*hspX*).

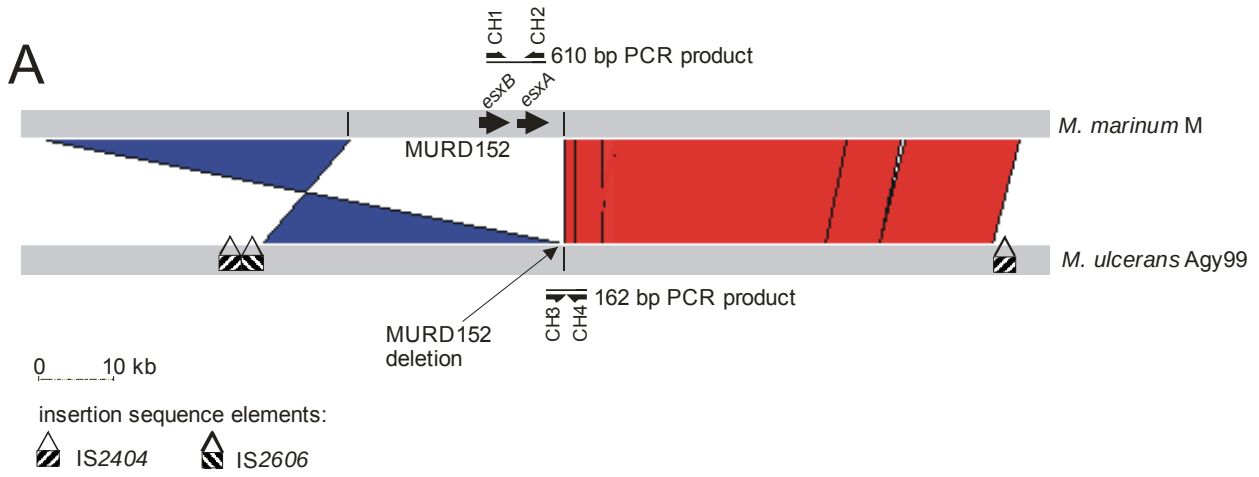
## Tables

| RD | MURD | Description of PCR product | expected product size [bp] | Primer1 | Primer2 |
|---|---|---|---|---|---|
| 13/14 | 4/152 | presence of *esxB-esxA* cluster in MURD4 and/or MURD152 | 610 | CH1-tgaagaccgatgccgctac | CH2-aacatccccgtgacgttg |
| 13 | 152 | MURD152 deletion as in Agy99 | 162 | CH3-cgttggggtgaatttctttg | CH4-agtctgacggcgactcatct |
| 13 | 152 | presence of *esxB-esxA* cluster in MURD 152 | 968 | CH5-ttggcgaggaaagaaagaga | CH4-agtctgacggcgactcatct |
| 14 | 4 | presence of esxB-esxA cluster in MURD4 | 810 | CH6-gacccaaagagatagagagtcca | CH7-tcatcggtgtcggtgtagtg |
| 14 | 4 | MURD4 deletion as in Agy99 | 1712 | CH8-gacccagacgatgtgaattg | CH9-ggagcatgttcacgatgttg |
| 13 | 152 | deletion □RD13A | 2354 | CH18-cagttatcgtgcgggaattt | CH19-atcgggagaaagaccgaagt |
| 13 | 152 | deletion □RD13B | 1650 | CH10-ctggcggaaacaacaacc | CH11-tcctggtcaagttggagacc |
| 13 | 152 | MURD152 deletion as in Agy99 | 3198 | CH10-ctggcggaaacaacaacc | CH12-gccgctaacttgaagaatcg |
| 13 | 152 | MURD152 deletion as in Agy99 | 1662 | CH13-ttctcgctcaatctcccta | CH11-tcctggtcaagttggagacc |
| 15 | 92 | presence of *hspX* in MURD92 | 791 | CH14-ggcgcttaaaccggtcgttg | CH15-cgccaaacccaggacaatca |
| 15 | 92 | MURD92 deletion as in Agy99 | 469 | CH16-agctggctagcgtcgtacc | CH17-cccaaagctcgtagatcagc |

**Table 1: Primers used in this study and description of respective PCR products**
All primers are listed in 5'-3' orientation.

| Species | Lineage | Strain/ InDel haplotype | RD13 (MURD152) | | RD14 (MURD4) | | RD15 (MURD92) |
|---|---|---|---|---|---|---|---|
| | | | *esxA (ESAT-6)* | *esxB (CFP-10)* | *esxA_1 (ESAT-6)* | *esxB_1 (CFP-10)* | *hspX* |
| *M. marinum* | | M | + | + | + | + | + |
| *M. ulcerans* | ancestral | South America | + [1] | + | deleted unlike MURD4[2] | deleted unlike MURD4[2] | + [5] |
| | | Asia | + [1] | pseudogenized due to frameshift mutation | deleted < 2.8 kb[3] | deleted < 2.8 kb[3] | + [5] |
| | | Mexico | deleted ΔRD13 A | deleted ΔRD13 A | + [4] | + [4] | pseudogenized due to frameshift mutation [5] |
| *M. ulcerans* | classical | Agy99 Africa/Australia | deleted MURD152 | deleted MURD152 | deleted MURD4 | deleted MURD4 | deleted MURD92 |
| | | Australia 9549 | deleted ΔRD13 B | deleted ΔRD13 B | deleted MURD4 | deleted MURD4 | deleted MURD92 |

**Table 2: Genomic deletions and amino acid changes in CDSs of immunogenic proteins**. (+) presence of CDS.
1) The following amino acid exchanges as compared to the *M. marinum* protein sequence were observed: A17S, Q19G, T23G, R51Q, N57K, S68A. 2) Lack of PCR products as shown in Fig. 1B suggests a deletion differing from the one in the classical lineage. 3) The deletion in the Asian strains is smaller than 2.8 kb and hence differs from both the *M. ulcerans* Agy99 MURD4 and South American haplotype specific deletions. 4) a screen using outward directed primers within each the *esxB-esxA* cluster, IS*2404* and IS*2606* and subsequent tests with nested PCR gave evidence for the presence of both ISEs in the vicinity in strain Mexico 5143 only, indicating yet further genomic changes in this region. 5) The following amino acid exchanges as compared to the *M. marinum* M protein sequence were observed: A105S (South American haplotype), V149F (Asian haplotype), D63A and L73R (Mexican haplotype, in addition to the frameshift mutati


A
CH1
CH2
610 bp PCR product
esxB
esxA
M. marinum M
MURD152
M. ulcerans Agy99
CH3
CH4
162 bp PCR product
MURD152
deletion
0
10 kb
insertion sequence elements:
IS2404
IS2606

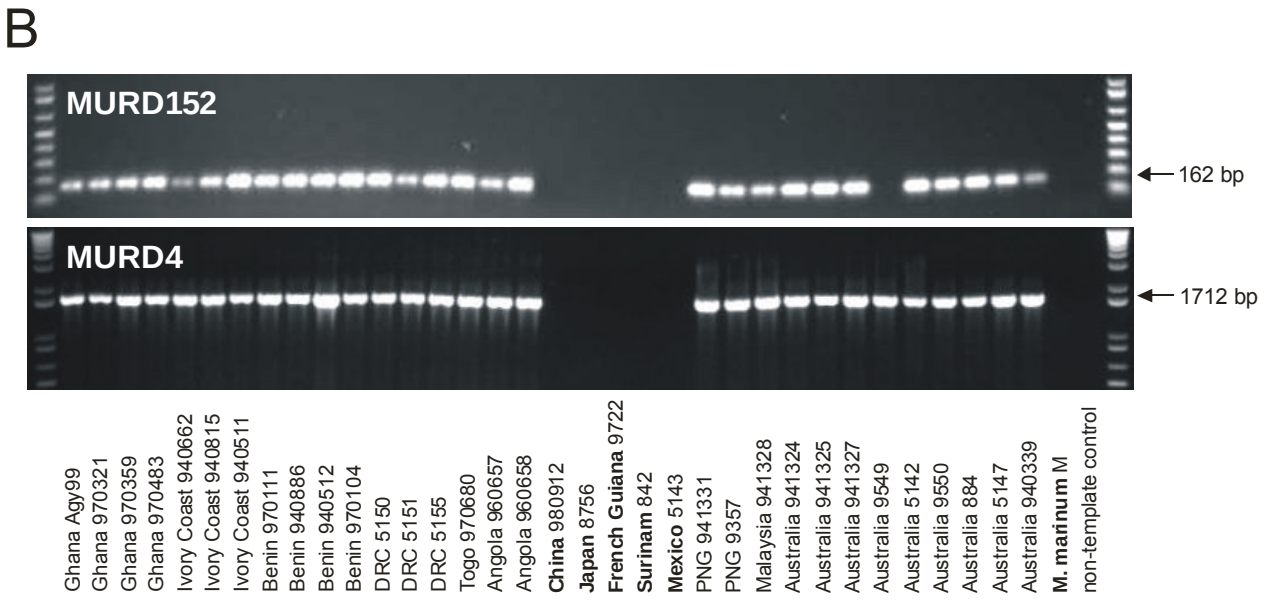

B
MURD152
162 bp
MURD4
1712 bp
Ghana Agy99
Ghana 970321
Ghana 970359
Ghana 970483
Ivory Coast 940662
Ivory Coast 940815
Ivory Coast 940511
Benin 970111
Benin 940886
Benin 940512
Benin 970104
DRC 5150
DRC 5151
DRC 5155
Togo 970680
Angola 960657
Angola 960658
China 980912
Japan 8756
French Guiana 9722
Surinam 842
Mexico 5143
PNG 941331
PNG 9357
Malaysia 941328
Australia 941324
Australia 941325
Australia 941327
Australia 9549
Australia 5142
Australia 9550
Australia 884
Australia 5147
Australia 940339
M. marinum M
non-template control

Fig. 1

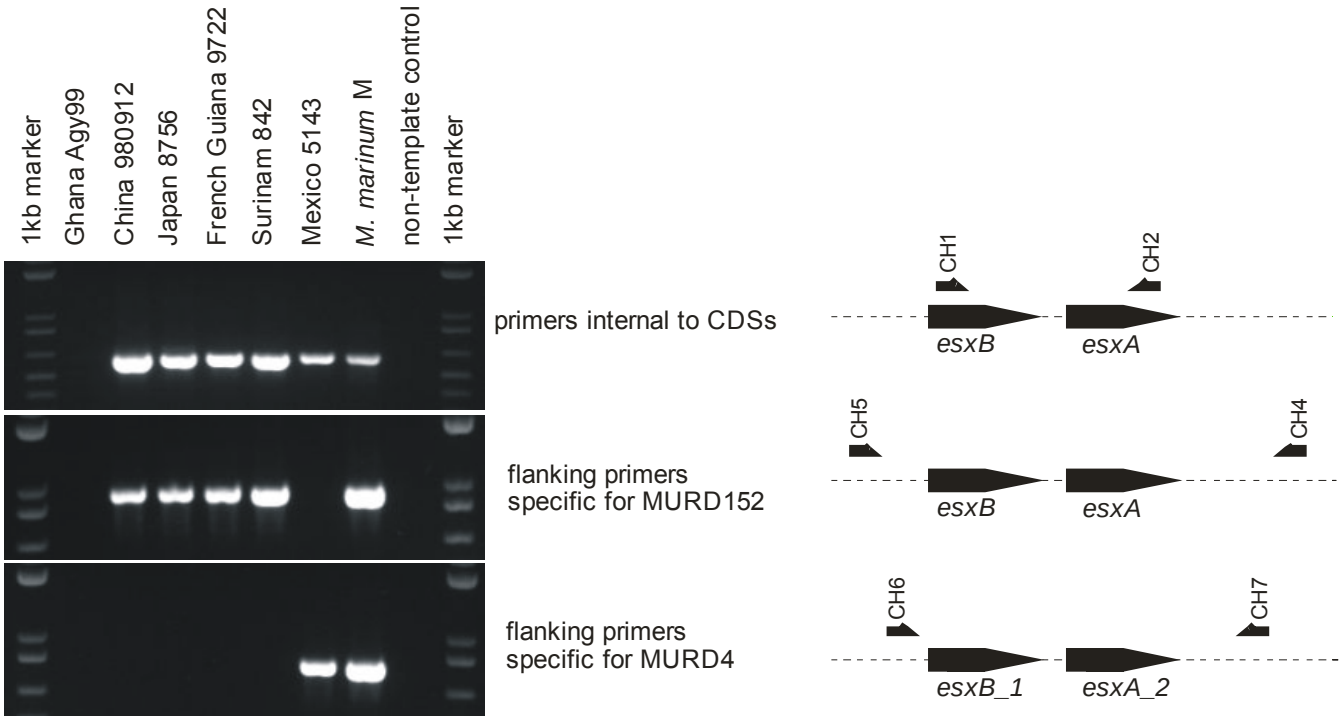

1kb marker
Ghana Agy99
China 980912
Japan 8756
French Guiana 9722
Surinam 842
Mexico 5143
M. marinum M
non-template control
1kb marker
primers internal to CDSs
flanking primers specific for MURD152
flanking primers specific for MURD4
CH1
CH2
esxB
esxA
CH5
CH4
esxB
esxA
CH6
CH7
esxB_1
esxA_2

Fig. 2

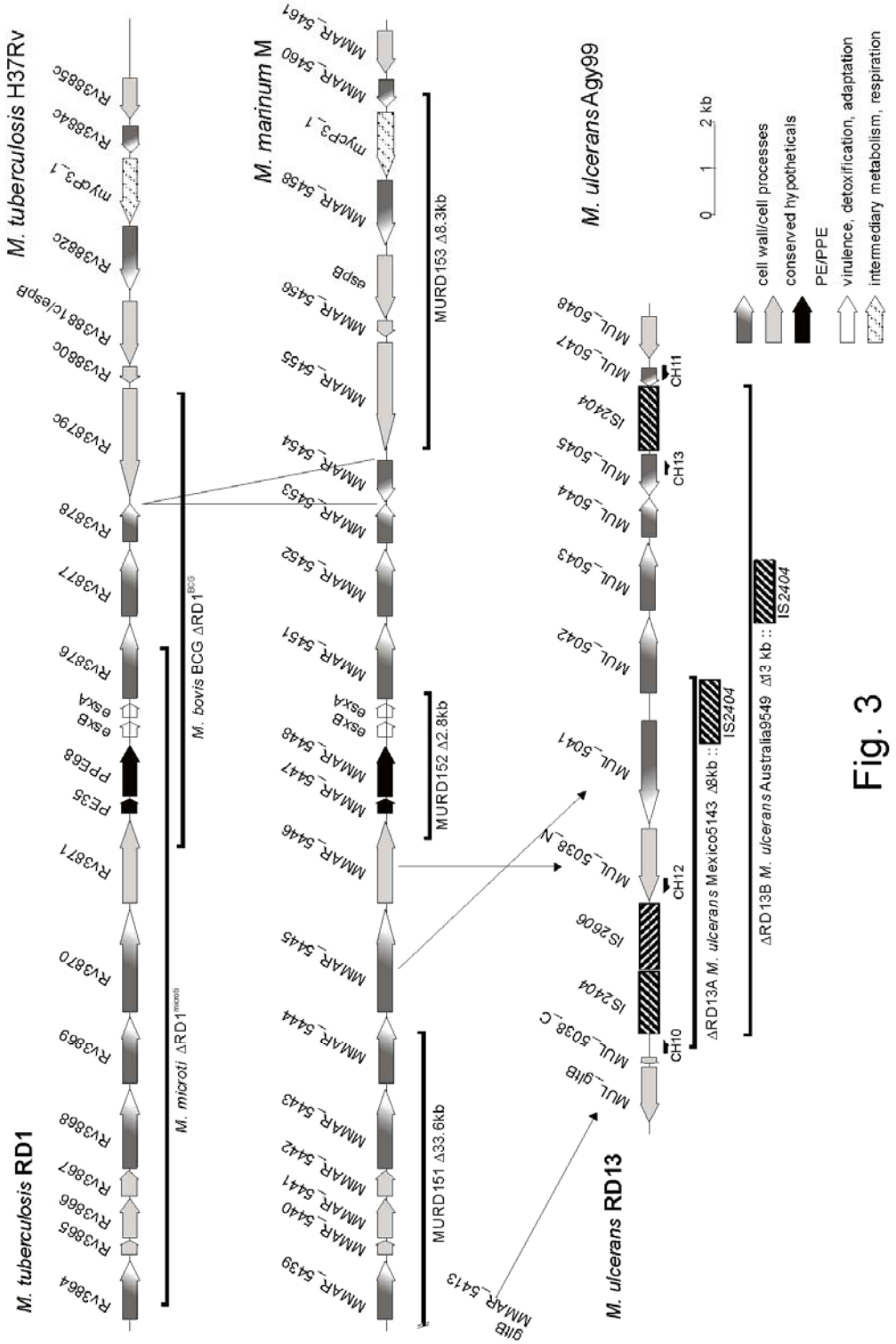

M. tuberculosis H37Rv
M. tuberculosis RD1
M. microti ΔRD1mouse
M. bovis BCG ΔRD1BCG
M. marinum M
MURD153 Δ8.3kb
MURD152 Δ2.8kb
MURD151 Δ33.6kb
M. ulcerans Agy99
M. ulcerans RD13
ΔRD13A M. ulcerans Mexico5143 Δ8kb :: IS2404
ΔRD13B M. ulcerans Australia9549 Δ13 kb :: IS2404
0 1 2 kb
cell wall/cell processes
conserved hypotheticals
PE/PPE
virulence, detoxification, adaptation
intermediary metabolism, respiration

Fig. 3

| Position | 48 | 51 | 57 | 75 | 156 | 181 | 196 | 197 | 312 | 313 | 318 | 378 | 388 | 393 | 394 | 395 | 406 | 407 | 471 | 494 | 510 | 525 | 541 | 610 | 621 | 623 | 624 | 627 | 628 | 630 | 632 | 634 | 635 | 636 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *M. marinum* MURD152 | T | T | G | G | G | G | A | A | A | A | T | A | G | T | C | A | A | C | T | G | T | G | T | A | T | C | T | A | T | C | C | C | T | C |
| *M. u.* Surinam 842 | T | T | G | G | G | G | A | A | G | A | G | C | T | C | G | G | G | G | T | A | G | C | G | A | T | C | T | A | T | C | C | - | T | C |
| *M. u.* French Guiana 792 | T | T | G | G | G | G | A | A | G | A | G | C | T | C | G | G | G | G | T | A | G | C | G | A | T | C | T | A | T | C | C | - | T | C |
| *M. u.* China 980912 | T | T | G | G | G | A | - | C | G | T | C | G | C | T | C | G | G | G | T | A | G | C | G | A | T | C | T | A | T | C | C | C | T | C |
| *M. u.* Japan 8756 | T | T | G | G | A | - | C | G | T | C | G | C | T | C | G | G | G | G | T | A | G | C | G | A | T | C | T | A | T | C | C | C | T | C |
| *M. marinum* MURD4 | C | C | A | G | G | G | A | A | A | A | C | C | T | C | G | G | G | G | G | C | A | G | C | G | G | G | C | G | C | G | G | A | T | G |
| *M. u.* Mexico 5143 | T | T | G | C | G | G | A | A | G | A | G | C | T | C | G | G | G | G | G | T | A | G | C | G | G | C | G | A | C | G | G | A | T | G |
| | *esxB (CFP-10)* | | | | | | | | | | | *esxA (ESAT-6)* | | | | | | | | | | | | | | | | | | | | | | |

**Fig. 4A**

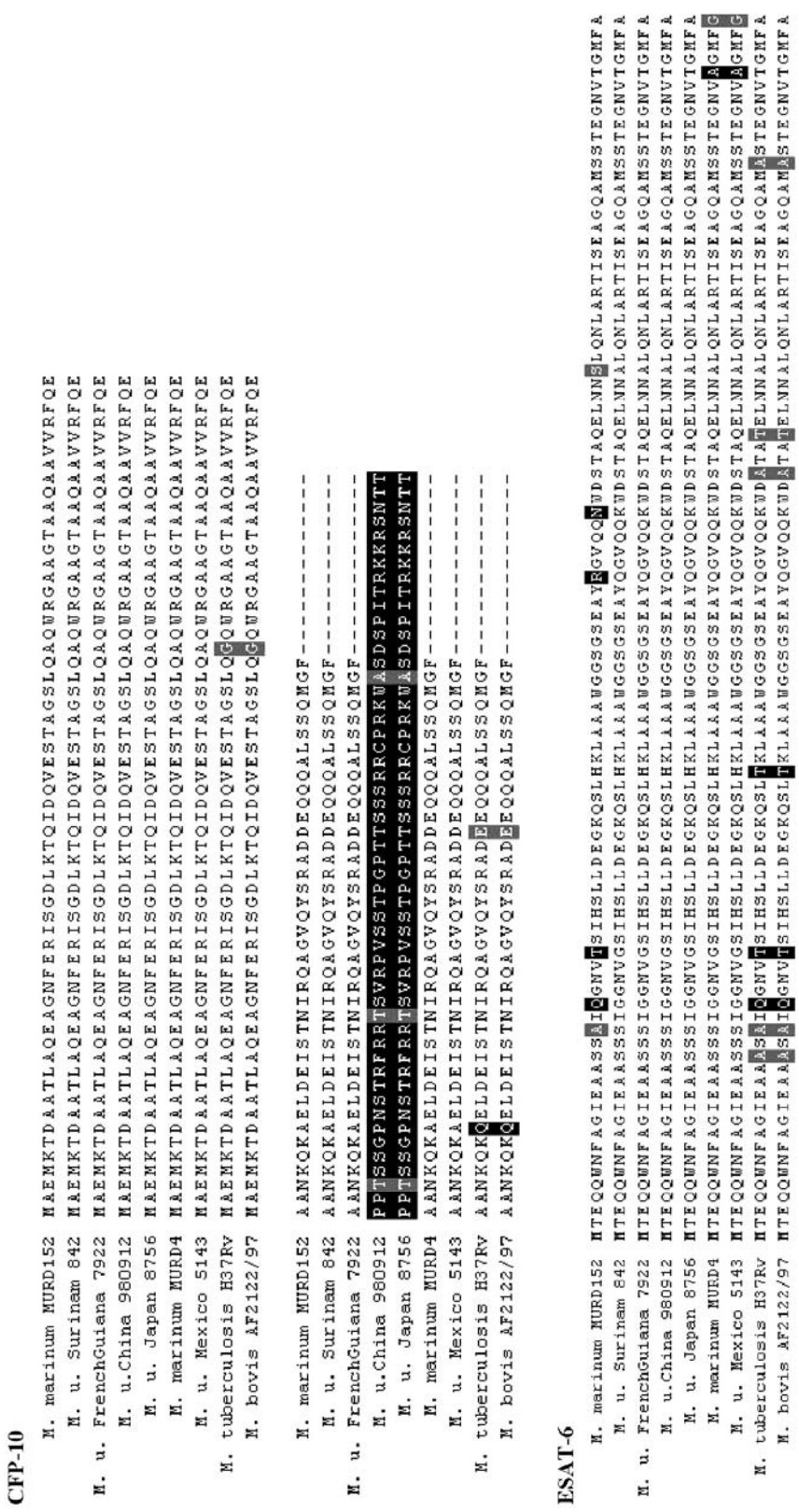

CFP-10
M. marinum MURD152
M. u. Surinam 842
M. u. FrenchGuiana 7922
M. u.China 980912
M. u. Japan 8756
M. marinum MURD4
M. u. Mexico 5143
M. tuberculosis H37Rv
M. bovis AF2122/97
ESAT-6
M. marinum MURD152
M. u. Surinam 842
M. u. FrenchGuiana 7922
M. u.China 980912
M. u. Japan 8756
M. marinum MURD4
M. u. Mexico 5143
M. tuberculosis H37Rv
M. bovis AF2122/97

Fig. 4B

**A**

| Position | 191 | 221 | 222 | 246 | 281 | 315 | 323 | 417 | 476 | 518 |
|---|---|---|---|---|---|---|---|---|---|---|
| *M. marinum* MURD92 | A | - | - | G | C | G | C | G | T | C |
| *M. u.* China 980912 | A | - | - | G | C | G | T | T | G | T |
| *M. u.* Japan 8756 | A | - | - | G | C | G | T | T | G | T |
| M. u. Surinam 842 | A | - | - | G | T | T | C | G | T | C |
| *M. u.* French Guiana 7922 | A | - | - | G | T | T | C | G | T | C |
| *M. u.* Mexico 5143 | C | G | C | C | C | G | C | G | T | C |

*hspX*

**B**

**HspX**

```
      M. marinum MURD92  M-GTLSVQRPARSLFPEFSELFAAFPSFAGLRPAFDSRLMKLEDEMKGDRYELRAEIPGVDPEKDIEITVLDGQLTIKAE
      M. u. Surinam 842  M-GTLSVQRPARSLFPEFSELFAAFPSFAGLRPAFDSRLMKLEDEMKGDRYELRAEIPGVDPEKDIEITVLDGQLTIKAE
M. u. FrenchGuiana 7922  M-GTLSVQRPARSLFPEFSELFAAFPSFAGLRPAFDSRLMKLEDEMKGDRYELRAEIPGVDPEKDIEITVLDGQLTIKAE
      M. u.China 980912  M-GTLSVQRPARSLFPEFSELFAAFPSFAGLRPAFDSRLMKLEDEMKGDRYELRAEIPGVDPEKDIEITVLDGQLTIKAE
       M. u. Japan 8756  M-GTLSVQRPARSLFPEFSELFAAFPSFAGLRPAFDSRLMKLEDEMKGDRYELRAEIPGVDPEKDIEITVLDGQLTIKAE
      M. u. Mexico 5143  M-GTLSVQRPARSLFPEFSELFAAFPSFAGLRPAFDSRLMKLEDEMKGDRYELRAEIPGVDPEKAIEITVLDGQR-----
  M. tuberculosis H37Rv  MATTLPVQRHPRSLFPEFSELFAAFPSFAGLRPTFDTRLMRLEDEMKEGRYEVRAELPGVDPDKDVDINVRDGQLTIKAE
     M. bovis AF2122/97  MATTLPVQRHPRSLFPEFSELFAAFPSFAGLRPTFDTRLMRLEDEMKEGRYEVRAELPGVDPDKDVDINVRDGQLTIKAE

      M. marinum MURD92  RSEKKEFNGRSEFCYGSFIRTVPLPAGVTEDGIEASYDKGILTVSVPVAEAKPAERRVRVQSVN
      M. u. Surinam 842  RSEKKEFNGRSEFCYGSFIRTVPLPSGVTEDGIEASYDKGILTVSVPVAEAKPAERRVRVQSVN
M. u. FrenchGuiana 7922  RSEKKEFNGRSEFCYGSFIRTVPLPSGVTEDGIEASYDKGILTVSVPVAEAKPAERRVRVQSVN
      M. u.China 980912  RSEKKEFNGRSEFCYGSFIRTVPLPAGVTEDGIEASYDKGILTVSVPVAEAKPAERRVRFQSVN
       M. u. Japan 8756  RSEKKEFNGRSEFCYGSFIRTVPLPAGVTEDGIEASYDKGILTVSVPVAEAKPAERRVRFQSVN
      M. u. Mexico 5143  ------------------------------------------------------------
  M. tuberculosis H37Rv  RTEQKDFDGRSEFAYGSFVRTVSLPVGADEDDIKATYDKGILTVSVAVSEGKPTEKHIQIRSTN
     M. bovis AF2122/97  RTEQKDFDGRSEFAYGSFVRTVSLPVGADEDDIKATYDKGILTVSVAVSEGKPTEKHIQIRSTN
```

**Fig. 5**

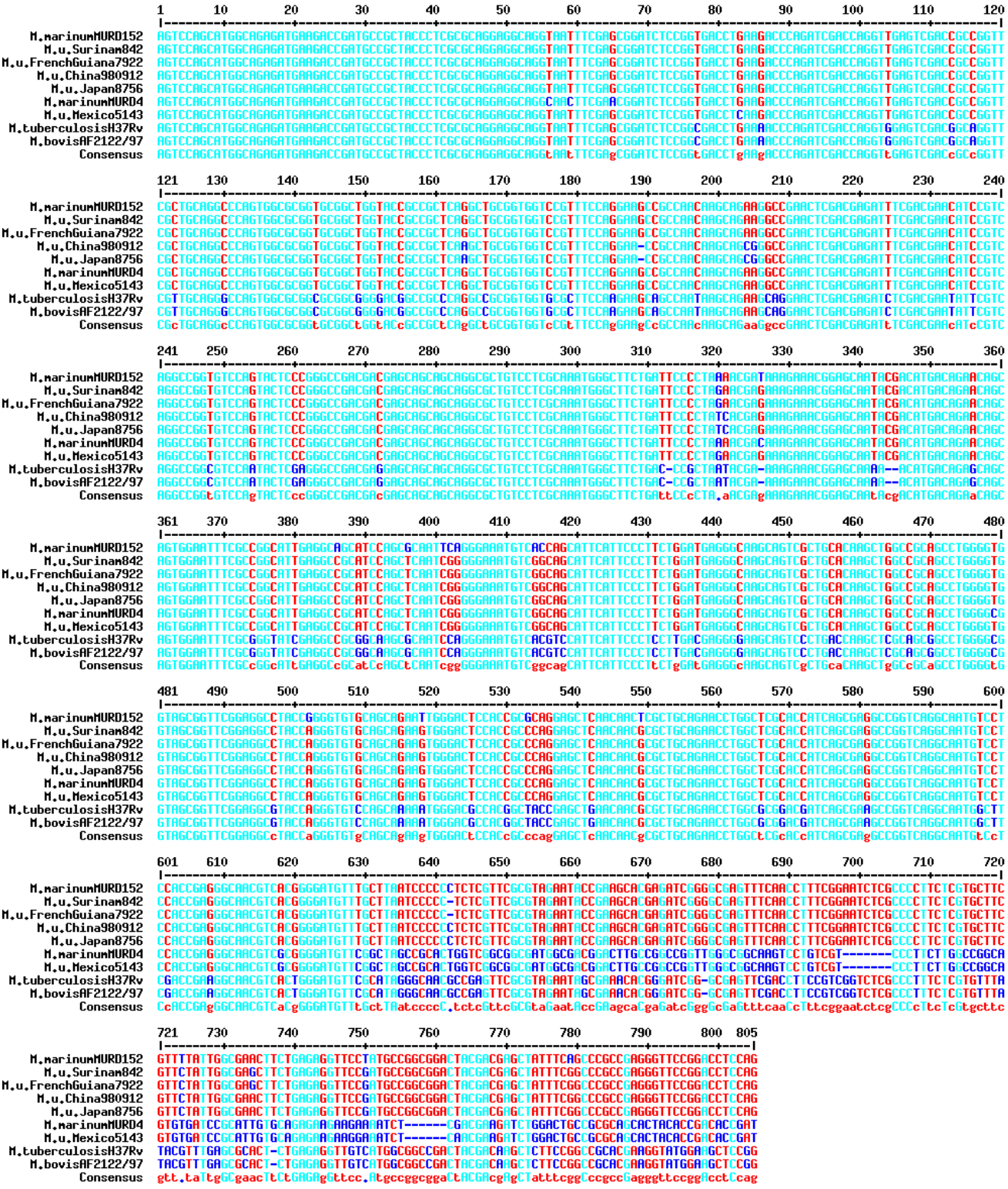



**Figure S1:** Nucleotide sequence alignment (using Multalin) of the esxB-esxA cluster in the two *M. marinum* loci and the *M. ulcerans* strains Surinam 842, French Guiana 7922, Japan 8756, China 980912 and Mexico 5143 in comparison with the respective *M. tuberculosis* and *M. bovis* CDSs. Nucleotides of CDSs are at the positions 9-309 (esxB) and 348-633 (esxA).

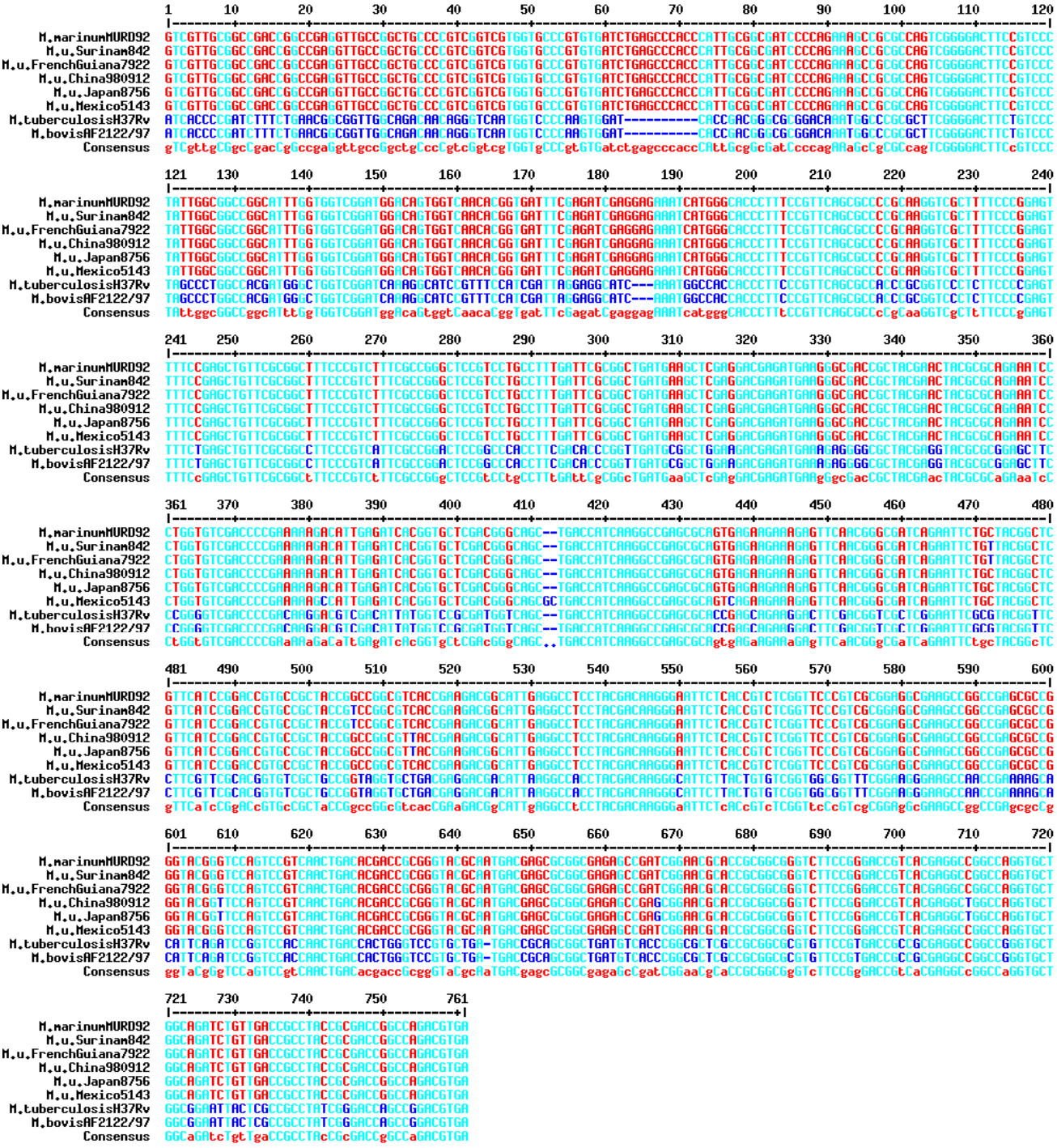



**Figure S2:** Nucleotide sequence alignment (using Multalin) of hspX in the two *M. marinum* loci and the *M. ulcerans* strains Surinam 842, French Guiana 7922, Japan 8756, China 980912 and Mexico 5143 in comparison with the respective *M. tuberculosis* and M. bovis CDSs. Nucleotides of the CDS are at position 192-623 (hspX).

## References


1. **Aagaard, C., M. Govaerts, V. Meikle, A. J. Vallecillo, J. A. Gutierrez-Pabello, F. Suarez-Guemes, J. McNair, A. Cataldi, C. Espitia, P. Andersen, and J. M. Pollock**. 2006. Optimizing antigen cocktails for detection of Mycobacterium bovis in herds with different prevalences of bovine tuberculosis: ESAT6-CFP10 mixture shows optimal sensitivity and specificity. J Clin. Microbiol. **44**:4326-4335.

2. **Ablordey, A., P. A. Fonteyne, P. Stragier, P. Vandamme, and F. Portaels**. 2007. Identification of a new variable number tandem repeat locus in Mycobacterium ulcerans for potential strain discrimination among African isolates. Clin. Microbiol. Infect.

3. **Ablordey, A., R. Kotlowski, J. Swings, and F. Portaels**. 2005. PCR amplification with primers based on IS2404 and GC-rich repeated sequence reveals polymorphism in Mycobacterium ulcerans. J. Clin. Microbiol. **43**:448-451.

4. **Adusumilli, S., A. Mve-Obiang, T. Sparer, W. Meyers, J. Hayman, and P. L. Small**. 2005. Mycobacterium ulcerans toxic macrolide, mycolactone modulates the host immune response and cellular location of M. ulcerans in vitro and in vivo. Cell Microbiol. **7**:1295-1304.

5. **Brodin, P., M. I. de Jonge, L. Majlessi, C. Leclerc, M. Nilges, S. T. Cole, and R. Brosch**. 2005. Functional analysis of early secreted antigenic target-6, the dominant T-cell antigen of Mycobacterium tuberculosis, reveals key residues involved in secretion, complex formation, virulence, and immunogenicity. J Biol Chem **280**:33953-33959.

6. **Brodin, P., K. Eiglmeier, M. Marmiesse, A. Billault, T. Garnier, S. Niemann, S. T. Cole, and R. Brosch**. 2002. Bacterial artificial chromosome-based comparative genomic analysis identifies Mycobacterium microti as a natural ESAT-6 deletion mutant. Infect. Immun. **70**:5568-5578.

7. **Brodin, P., L. Majlessi, L. Marsollier, M. I. de Jonge, D. Bottai, C. Demangel, J. Hinds, O. Neyrolles, P. D. Butcher, C. Leclerc, S. T. Cole, and R. Brosch**. 2006. Dissection of ESAT-6 system 1 of Mycobacterium tuberculosis and impact on immunogenicity and virulence. Infect. Immun. **74**:88-98.

8. **Brodin, P., I. Rosenkrands, P. Andersen, S. T. Cole, and R. Brosch**. 2004. ESAT-6 proteins: protective antigens and virulence factors? Trends Microbiol. **12**:500-508.

9. **Carver, T. J., K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell, and J. Parkhill**. 2005. ACT: the Artemis Comparison Tool. Bioinformatics. **21**:3422-3423.

10. **Champion, P. A., S. A. Stanley, M. M. Champion, E. J. Brown, and J. S. Cox**. 2006. C-terminal signal sequence promotes virulence factor secretion in Mycobacterium tuberculosis. Science **313**:1632-1636.

11. **Chemlal, K., G. Huys, P. A. Fonteyne, V. Vincent, A. G. Lopez, L. Rigouts, J. Swings, W. M. Meyers, and F. Portaels**. 2001. Evaluation of PCR-restriction profile analysis and IS2404 restriction fragment length polymorphism and amplified fragment length polymorphism fingerprinting for identification and typing of Mycobacterium ulcerans and M. marinum. J. Clin. Microbiol. **39**:3272-3278.

12. **de Jong, B. C., P. C. Hill, R. H. Brookes, S. Gagneux, D. J. Jeffries, J. K. Otu, S. A. Donkor, A. Fox, K. P. McAdam, P. M. Small, and R. A. Adegbola**. 2006. Mycobacterium africanum elicits an attenuated T cell response to early secreted antigenic target, 6 kDa, in patients with tuberculosis and their household contacts. J Infect. Dis. **193**:1279-1286.

13. **de Jonge, M. I., G. Pehau-Arnaudet, M. M. Fretz, F. Romain, D. Bottai, P. Brodin, N. Honore, G. Marchal, W. Jiskoot, P. England, S. T. Cole, and R. Brosch**. 2007. ESAT-6 from Mycobacterium tuberculosis dissociates from its putative chaperone CFP-10 under acidic conditions and exhibits membrane-lysing activity. J Bacteriol. **189**:6028-6034.

14. **Demissie, A., E. M. Leyten, M. Abebe, L. Wassie, A. Aseffa, G. Abate, H. Fletcher, P. Owiafe, P. C. Hill, R. Brookes, G. Rook, A. Zumla, S. M. Arend, M. Klein, T. H. Ottenhoff, P. Andersen, and T. M. Doherty**. 2006. Recognition of stage-specific mycobacterial antigens differentiates between acute and latent infections with Mycobacterium tuberculosis. Clin. Vaccine Immunol. **13**:179-186.

15. **DesJardin, L. E., L. G. Hayes, C. D. Sohaskey, L. G. Wayne, and K. D. Eisenach**. 2001. Microaerophilic induction of the alpha-crystallin chaperone protein homologue (hspX) mRNA of Mycobacterium tuberculosis. J Bacteriol. **183**:5311-5316.

16. **DiGiuseppe Champion, P. A. and J. S. Cox**. 2007. Protein secretion systems in Mycobacteria. Cell Microbiol. **9**:1376-1384.

17. **Gao, L. Y., S. Guo, B. McLaughlin, H. Morisaki, J. N. Engel, and E. J. Brown**. 2004. A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. Mol. Microbiol. **53**:1677-1693.

18. **Garcia-Pelayo, M. C., K. C. Caimi, J. K. Inwald, J. Hinds, F. Bigi, M. I. Romano, D. van Soolingen, R. G. Hewinson, A. Cataldi, and S. V. Gordon**. 2004. Microarray analysis of Mycobacterium microti reveals deletion of genes encoding PE-PPE proteins and ESAT-6 family antigens. Tuberculosis. (Edinb. ) **84**:159-166.

19. **Guinn, K. M., M. J. Hickey, S. K. Mathur, K. L. Zakel, J. E. Grotzke, D. M. Lewinsohn, S. Smith, and D. R. Sherman**. 2004. Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of Mycobacterium tuberculosis. Mol. Microbiol. **51**:359-370.

20. **Haile, Y., G. Bjune, and H. G. Wiker**. 2002. Expression of the mceA, esat-6 and hspX genes in Mycobacterium tuberculosis and their responses to aerobic conditions and to restricted oxygen supply. Microbiology **148**:3881-3886.

21. **Harboe, M., T. Oettinger, H. G. Wiker, I. Rosenkrands, and P. Andersen**. 1996. Evidence for occurrence of the ESAT-6 protein in Mycobacterium tuberculosis and virulent Mycobacterium bovis and for its absence in Mycobacterium bovis BCG. Infect. Immun. **64**:16-22.

22. **Hilty, M., M. Käser, J. Zinsstag, T. Stinear, and G. Pluschke**. 2007. Analysis of the Mycobacterium ulcerans genome sequence reveals new loci for variable number tandem repeats (VNTR) typing. Microbiology **in press**.

23. **Hilty, M., D. Yeboah-Manu, D. Boakye, E. Mensah-Quainoo, S. Rondini, E. Schelling, D. Ofori-Adjei, F. Portaels, J. Zinsstag, and G. Pluschke**. 2006. Genetic Diversity in Mycobacterium ulcerans Isolates from Ghana Revealed by a Newly Identified Locus Containing a Variable Number of Tandem Repeats. J Bacteriol. **188**:1462-1465.

24. **Hsu, T., S. M. Hingley-Wilson, B. Chen, M. Chen, A. Z. Dai, P. M. Morin, C. B. Marks, J. Padiyar, C. Goulding, M. Gingery, D. Eisenberg, R. G. Russell, S. C. Derrick, F. M. Collins, S. L. Morris, C. H. King, and W. R. Jacobs, Jr.** 2003. The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. Proc. Natl. Acad. Sci. U. S. A **100**:12420-12425.

25. **Hu, Y., F. Movahedzadeh, N. G. Stoker, and A. R. Coates**. 2006. Deletion of the Mycobacterium tuberculosis alpha-crystallin-like hspX gene causes increased bacterial growth in vivo. Infect. Immun. **74**:861-868.

26. **Johnson, P. D., T. Stinear, P. L. Small, G. Pluschke, R. W. Merritt, F. Portaels, K. Huygen, J. A. Hayman, and K. Asiedu**. 2005. Buruli Ulcer (M. ulcerans Infection): New Insights, New Hope for Disease Control. PLoS. Med. **2**:e108.

27. **Kaser, M., S. Rondini, M. Naegeli, T. Stinear, F. Portaels, U. Certa, and G. Pluschke**. 2007. Evolution of two distinct phylogenetic lineages of the emerging human pathogen Mycobacterium ulcerans. BMC. Evol. Biol **7**:177.

28. **Macgurn, J. A. and J. S. Cox**. 2007. A genetic screen for Mycobacterium tuberculosis mutants defective for phagosome maturation arrest identifies components of the ESX-1 secretion system. Infect. Immun. **75**:2668-2678.

29. **Mahairas, G. G., P. J. Sabo, M. J. Hickey, D. C. Singh, and C. K. Stover**. 1996. Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. J Bacteriol. **178**:1274-1282.

30. **Marmiesse, M., P. Brodin, C. Buchrieser, C. Gutierrez, N. Simoes, V. Vincent, P. Glaser, S. T. Cole, and R. Brosch**. 2004. Macro-array and bioinformatic analyses reveal mycobacterial 'core' genes, variation in the ESAT-6 gene family and new phylogenetic markers for the Mycobacterium tuberculosis complex. Microbiology **150**:483-496.

31. **McLaughlin, B., J. S. Chon, J. A. Macgurn, F. Carlsson, T. L. Cheng, J. S. Cox, and E. J. Brown**. 2007. A Mycobacterium ESX-1-Secreted Virulence Factor with Unique Requirements for Export. PLoS. Pathog. **3**:e105.

32. **Mve-Obiang, A., R. E. Lee, E. S. Umstot, K. A. Trott, T. C. Grammer, J. M. Parker, B. S. Ranger, R. Grainger, E. A. Mahrous, and P. L. Small**. 2005. A newly discovered mycobacterial pathogen isolated from laboratory colonies of Xenopus species with lethal infections produces a novel form of mycolactone, the Mycobacterium ulcerans macrolide toxin. Infect. Immun. **73**:3307-3312.

33. **Purkayastha, A., L. A. McCue, and K. A. McDonough**. 2002. Identification of a Mycobacterium tuberculosis putative classical nitroreductase gene whose expression is coregulated with that of the acr aene within macrophages, in standing versus shaking cultures, and under low oxygen conditions. Infect. Immun. **70**:1518-1529.

34. **Rondini, S., M. Käser, T. Stinear, M. Tessier, C. Mangold, G. Dernick, M. Naegeli, F. Portaels, U. Certa, and G. Pluschke**. 2007. Ongoing genome reduction in Mycobacterium ulcerans. Emerg. Infect. Dis. **13**:1008-1015.

35. **Roupie, V., M. Romano, L. Zhang, H. Korf, M. Y. Lin, K. L. Franken, T. H. Ottenhoff, M. R. Klein, and K. Huygen**. 2007. Immunogenicity of eight dormancy regulon-encoded proteins of Mycobacterium tuberculosis in DNA-vaccinated and tuberculosis-infected mice. Infect. Immun. **75**:941-949.

36. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell**. 2000. Artemis: sequence visualization and annotation. Bioinformatics. **16**:944-945.

37. **Stanley, S. A., J. E. Johndrow, P. Manzanillo, and J. S. Cox**. 2007. The Type I IFN response to infection with Mycobacterium tuberculosis requires ESX-1-mediated secretion and contributes to pathogenesis. J Immunol. **178**:3143-3152.

38. **Stinear, T., J. K. Davies, G. A. Jenkin, F. Portaels, B. C. Ross, F. Oppedisano, M. Purcell, J. A. Hayman, and P. D. Johnson**. 2000. A simple PCR method for rapid genotype analysis of Mycobacterium ulcerans. J. Clin. Microbiol. **38**:1482-1487.

39. **Stinear, T. P., H. Hong, W. Frigui, M. J. Pryor, R. Brosch, T. Garnier, P. F. Leadlay, and S. T. Cole**. 2005. Common evolutionary origin for the unstable virulence plasmid pMUM found in geographically diverse strains of Mycobacterium ulcerans. J. Bacteriol. **187**:1668-1676.

40. **Stinear, T. P., G. A. Jenkin, P. D. Johnson, and J. K. Davies**. 2000. Comparative genetic analysis of Mycobacterium ulcerans and Mycobacterium marinum reveals evidence of recent divergence. J. Bacteriol. **182**:6322-6330.

41. **Stinear, T. P., A. Mve-Obiang, P. L. Small, W. Frigui, M. J. Pryor, R. Brosch, G. A. Jenkin, P. D. Johnson, J. K. Davies, R. E. Lee, S. Adusumilli, T. Garnier, S. F. Haydock, P. F. Leadlay, and S. T. Cole**. 2004. Giant plasmid-encoded polyketide synthases produce the macrolide toxin of Mycobacterium ulcerans. Proc. Natl. Acad. Sci. U. S. A **101**:1345-1349.

42. **Stinear, T. P., M. J. Pryor, J. L. Porter, and S. T. Cole**. 2005. Functional analysis and annotation of the virulence plasmid pMUM001 from Mycobacterium ulcerans. Microbiology **151**:683-692.

43. **Stinear, T. P., T. Seemann, S. Pidot, W. Frigui, G. Reysset, T. Garnier, G. Meurice, D. Simon, C. Bouchier, L. Ma, M. Tichit, J. L. Porter, J. Ryan, P. D. Johnson, J. K. Davies, G. A. Jenkin, P. L. Small, L. M. Jones, F. Tekaia, F. Laval, M. Daffe, J. Parkhill, and S. T. Cole**. 2007. Reductive evolution and niche adaptation inferred from the genome of Mycobacterium ulcerans, the causative agent of Buruli ulcer. Genome Res.

44. **Stragier, P., A. Ablordey, L. M. Bayonne, Y. L. Lugor, I. S. Sindani, P. Suykerbuyk, H. Wabinga, W. M. Meyers, and F. Portaels**. 2006. Heterogeneity among Mycobacterium ulcerans isolates from Africa. Emerg. Infect. Dis. **12**:844-847.

45. **Tan, T., W. L. Lee, D. C. Alexander, S. Grinstein, and J. Liu**. 2006. The ESAT-6/CFP-10 secretion system of Mycobacterium marinum modulates phagosome maturation. Cell Microbiol. **8**:1417-1429.

46. **Tekaia, F., S. V. Gordon, T. Garnier, R. Brosch, B. G. Barrell, and S. T. Cole**. 1999. Analysis of the proteome of Mycobacterium tuberculosis in silico. Tuber. Lung Dis. **79**:329-342.

47. **Torrado, E., S. Adusumilli, A. G. Fraga, P. L. Small, A. G. Castro, and J. Pedrosa**. 2007. Mycolactone-mediated inhibition of tumor necrosis factor production by macrophages infected with Mycobacterium ulcerans has implications for the control of infection. Infect. Immun. **75**:3979-3988.

48. **Yip, M. J., J. L. Porter, J. A. Fyfe, C. J. Lavender, F. Portaels, M. Rhodes, H. Kator, A. Colorni, G. A. Jenkin, and T. Stinear**. 2007. Evolution of Mycobacterium ulcerans and other mycolactone-producing mycobacteria from a common Mycobacterium marinum progenitor. J Bacteriol.

49. **Yuan, Y., D. D. Crane, R. M. Simpson, Y. Q. Zhu, M. J. Hickey, D. R. Sherman, and C. E. Barry, III**. 1998. The 16-kDa alpha-crystallin (Acr) protein of Mycobacterium tuberculosis is required for growth in macrophages. Proc. Natl. Acad. Sci. U. S. A **95**:9578-9583.

# CHAPTER 3: Genetic diversity of *Neisseria meningitidis*

# Lack of antigenic diversification of outer membrane proteins during clonal waves of *Neisseria meningitidis* serogroup A colonization and disease

Charlotte Huber[1,2], Valentin Pflüger[1,2], Abdul-Wahab M.Hamid[3], Abudulai Adams Forgor[3], Abraham Hodgson[3], Ali Sie[4], Thomas Junghanss[5], Gerd Pluschke[1,2*]

[1]Swiss Tropical and Public Health Institute, Socinstr. 57, 4002 Basel, Switzerland
[2]University of Basel, Petersplatz 1, 4003 Basel, Switzerland
[3]Navrongo Health Research Centre, Ministry of Health, Navrongo, Ghana
[4]Centre de Recherche en Sante de Nouna, Nouna, Burkina Faso
[5]Section of Clinical Tropical Medicine, University Hospital Heidelberg, Heidelberg, Germany

* Corresponding author:
Prof. Gerd Pluschke, PhD
Swiss Tropical and Public Health Institute, Socinstr. 57, 4002 Basel, Switzerland
E-mail: gerd.pluschke@unibas.ch
phone: + 41 61 284 82 35
fax: + 41 61 284 81 01

Ready for submission

**Abstract**

In particular in the 'meningitis belt' of sub-Saharan Africa, epidemic meningococcal meningitis is a severe unresolved public health problem. In the past decades, serogroup A lineages have been the dominant etiologic agents, but also other serogroups, like C, W135 and X have caused outbreaks. Control of meningitis epidemics has relied so far on reactive vaccination strategies with polysaccharide vaccines. While a serogroup A polysaccharide conjugate vaccine is currently being clinically tested, a comprehensive vaccine based on sub-capsular outer membrane proteins (OMPs) is not available. Here we have investigated whether meningococcal populations overcome immune selection pressure associated with herd immunity by changing antigenic properties of their OMPs. Meningococcal isolates were collected in the context of longitudinal colonization and disease surveys in Ghana and Burkina Faso. Serogroup A strains isolated during two clonal waves of colonization and disease showed no diversification in the genes encoding their PorA, PorB, and FetA proteins. However, ST7 and ST2859 strains had different *opa*J alleles and ST2859 isolates from Ghana differed from ST2859 isolates from Burkina Faso in the *opa*A allele. This provides evidence for occasional allelic exchange of *opa* genes and selection of strains expressing Opa protein variants. No allelic difference within the epidemiologically related clonal group of isolates and no variations based on point mutations or insertional/deletional events in the OMP genes were found. However, *opa* genes showed wide variation in the number of intragenic tandem repeats. This shows that phase variation of Opa protein expression caused by slipped strand mispairing is a frequent event. Taken together our results demonstrate a remarkable antigenic stability of OMPs over years during the spread and local persistence of hyper-invasive meningococcal clones in human populations. Herd immunity thus does not seem to be a strong driving force for antigenic diversification of the major OMPs analyzed.

**Key words:** *Neisseria meningitidis*, OMP, meningitis, herd immunity

## Introduction

*Neisseria meningitidis*, a Gram-negative diplococcus, is an obligate commensal of humans, usually carried in the upper airways of healthy individuals. The bacterium is transmitted to close contacts by the spread of respiratory secretions. Occasionally, *N. meningitidis* strains might penetrate the mucosal membrane, multiply in the bloodstream and gain access to the cerebrospinal fluid. The overall incidence rate of meningococcal disease in Europe and North America is 1-3 per 100 000 person-years. The highest incidence rates are found in countries of sub-Saharan Africa, in the so-called “meningitis belt” which extends from Ethiopia to Senegal. In these countries, annual incidence rates of as high as 1000 per 100 000 person-years are recorded during the most severe epidemics (1). In the Meningitis belt, epidemics classically occur in the dry season, between December and April. They nearly always start in the early part of the dry season when it is hot, dry and dusty, build up to a peak at the end of the dry season and then stop abruptly at the onset of rains. High temperature, low absolute humidity and the harmattan (a dusty wind that blows from the Sahara) at the end of the dry season may favor the occurrence of meningococcal disease by damaging the local mucosal defenses (2). Based on different capsular polysaccharide structures, 13 serogroups of *N. meningitidis* can be distinguished. The vast majority of invasive meningococcal disease is caused by six of these serogroups (A, B, C, W-135, X and Y) (3). Most of the large epidemics in Africa in the past 100 years were associated with serogroup A (1). Clonal waves of colonization and disease are a characteristic feature of the epidemiology of meningococcal meningitis in the African meningitis belt (4). Closely related *N. meningitidis* serogroup A, genoclouds associated with the sequence types (STs) 5, 7 and 2859 have been responsible for outbreaks in the last two decades (5), (6), (7). Molecular typing approaches based on DNA sequencing allow meningococci to be distinguished and tracked (8). During the last decade, multi locus sequence typing (MLST) based on the identification of the alleles of fragments of seven meningococcal housekeeping genes has developed into the “gold standard” for typing these bacterial pathogens (9), (10).

Meningococci are naturally competent for transformation by exogenous DNA and high rates of recombination have been observed (11). However, most genetic exchange appears to takes place between very closely related meningococci and it has been suggested that recombination may be primarily a mechanism for genome repair that will only occasionally result in generation of diversity (12). Nevertheless *N. meningitidis* is genetically and antigenically highly diverse. Sequence typing has identified to date >5000 STs and hundreds of variants of vaccine candidate antigens, such as the outer membrane protein (OMP) PorA (http://neisseria.org/nm/). On the other hand, hyper-invasive lineages seem to be surprisingly stable over decades and during global spread (1). This may facilitate development of a comprehensive protein-based vaccine, effective against a broad range of hyper-virulent meningococci. Cross-reactivity could be achieved by targeting antigenically invariant sub-capsular structures or by combining a cocktail of vaccine antigens selected on the basis of molecular epidemiological studies. A number of meningococcal surface structures are thought to play a role in mucosal colonization, haematogenous spread and penetration of the blood brain barrier (13). In order to escape immune surveillance, meningococci have developed a range of mechanisms to change surface components. We assume that development of herd immunity is responsible for the complete disappearance of meningococcal clones after a few years of colonization of populations in the African meningitis belt (4). Here we have investigated whether meningococcal populations escape from immune detection by varying their OMPs PorA, PorB, FetA and Opa. The serogroup A ST7 and ST 2859 meningococci analyzed have been collected between March 2002 and April 2008 in the course of longitudinal meningococcal colonization and disease surveys in Ghana and Burkina Faso (4), (7).

## Materials and methods

### Bacterial isolates

The *N. meningitidis* isolates investigated in this study had been collected in the Kassena-Nankana District (KND) and the neighboring district of Bawku in Ghana and in the Nouna health district (NHD) in the Kossi region of Burkina Faso. Case strains were isolated from the cerebrospinal fluid of meningitis patients and carriage strains were isolated from throat swabs collected in the context of longitudinal carriage surveys. Isolation and characterization of strains has been described previously (14), (4), (7). For the analysis of genetic diversification, serogroup A ST7 and ST2859 strains isolated from cases and carriers at different times during clonal colonization and disease waves were selected from our strain collection. Included were nine ST7 carriage and seven ST7 case isolates collected between March 2002 and March 2005 in the KND of Ghana. In the case of ST2859 strains we analyzed six case and seven carriage isolates collected between March 2006 and March 2007 in the NHD of Burkina Faso, as well as three case and fifteen carriage isolates from the KND, as well as three case isolates collected between March 2007 and April 2008 in the neighboring district of Bawku.

### Genetic analysis

DNA was extracted from bacterial pellets using the Wizard® Genomic DNA Purification Kit (Promega AG, Duebendorf, Switzerland). The DNA concentration was measured using a Nano drop Spectrophotometer (Witec Ag, Littau, Switzerland). PCR was performed using 5µl of 10× BD buffer and 1µl of FirePol Taq polymerase, 1.25 mM $MgCl_2$ (Solis BioDyne, Tartu, Estonia), 100 ng of genomic DNA or the equivalent volume of nuclease-free water as a negative control, a 0.2 µM concentration of each forward and reverse primer, and a 0.2 mM concentration of each deoxynucleoside triphosphate in a total volume of 50 µl. PCRs were run in a T Professional Basic PCR machine (Biometra GmbH, Göttingen Germany). The thermal profile for PCR amplification included an initial denaturation step of 94°C for 2 min, followed by 35 cycles of 94°C for 30 s, annealing at 55°C for 30 s, and

elongation at 72°C for 1min 20 s up to 2 min. The PCRs were finalized by an extension step at 72°C for 10 min. PCR products were analyzed on 1% agarose gels by gel electrophoresis using ethidium bromide staining and the AlphaImager illuminator and AlphaImager software (Alpha Innotech, San Leandro, CA). PCR products were purified using a NucleoSpin purification kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) and subjected to direct sequencing or cloned using a TOPO TA cloning kit (Invitrogen Corp., Carlsbad, CA), transformed into Novablue competent cells (Merck, Darmstadt, Germany), and sequenced after DNA preparation (Macherey-Nagel GmbH & Co. KG, Düren, Germany). Sequencing was performed by Macrogen Inc, Seoul, Korea. For PorA, Por B and FetA typing primers were used as previously described (15) (16) (17). For the *opa* genes and their flanking regions the following primers were used: *Opa*A VNTR15af (TCATCCGCT ACATTGTGTTGA) and OpaA2r (TCGTCATTCCCACGGAAGT) for amplification and VNTR15af, OpaA4r (TTTCCTGATTTTCCGTCTTCA), OpaA5r (ATGACGGTTCGGG TATTTCC) and OpaA4f (GCGGCAGATTATG CCAGTTA) for sequencing. *Opa*B: OpaB2f (CA GGACAAGGCGACGAG) and OpaB5r (TGTCTGGACGGGGATGT) for amplification and VNTR15br (GCACACCGATATAGGGTTTGAA), OpaBf (GTGTTG AAACATCGCCACAA), OpaBr1 (GGCATTTTTCCATGCGTTT) and OpaBff (GCGAG AACTGAAGACGGAAA) for sequencing. For OpaD, OpaD5f (TCTCCGTAG AGGAAATGATGC) and OpaD3r (AAGTGGGAATCTAGGACGTAAAA) for amplification and OpDf (TCATCCGCTATATTGTGTTGA), Opa26f (TGGGTCTTGG TGTCATCG), OpA26r (GAATAATTACTTTCTTTCCATTTTCTG), OpD2f (CGCCCC AAACCTGATATAGT) and OpDr2 (GAAACGGTGGGAATTGTGTAA) for sequencing were used. For *Opa*J: Opaj5f (CGCCCCAAACCTGATATAGT) and Opaj1r (ATCT AGAACGTGGGGTTTGG) for amplification and Opaj5f, Opaj7f (TGATATAGTC CGCTCCTGCAA) Opaj8f (CGGTGCAGACAAAGACAAAA) Opaj9f (GTCGCCGG TGCTGCTA) Opaj10r (TAG CAGCACCGGCGAC) Opaj3r (TTTGGGCAACTGTT TTTATCC) for sequencing. Design of the primers used for amplification of opaA, opaB and opaD was based on the serogroup A strain Z2491genome sequence (18). and the Primer3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). The sequences obtained were analyzed making use of the *N. meningitidis* sequence

databases:(http://neisseria.org/perl/agdbnet.pl?file=poravr.xml), (http://neisseria.org/nm/typing/porb/),(http://neisseria.org/perl/agdbnet.pl?file=fetavr.xml), (http://neisseria.org/nm/typing/opa/). The multiple sequence alignment websites Multalin (http://bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html) and Clustalw2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html) were used for comparisons between *Opa* gene sequences and their flanking regions. The EMBOSS Transeq tool was used for the translation of the retrieved opa gene sequences. (http://www.ebi.ac.uki/Tools/ emboss/transeq)**.**

## Results

### PorA, PorB and FetA protein typing

*N. meningitidis* is naturally highly competent and transformation by exogenous DNA leading to allelic exchange of gene fragments and genes represents an important mechanism of genetic diversification of this species. Nevertheless all 50 serogroup A strains analyzed here shared the same *por*A, *por*B and *fet*A gene sequences (Table 1). The analyzed strains had been isolated from CSF of meningitis patients or from the pharynx of healthy carriers during two sequential colonization and disease waves in northern Ghana and during a meningococcal disease outbreak in Burkina Faso. Strains isolated from the start (2002) till the end (2005) of a ST7 colonization wave (4) were included in the analysis. In the case of ST 2859, strains isolated during outbreaks in Burkina Faso (2006-2007) (7) and in northern Ghana (2007-2008) were analyzed.

All PorA sequences had the VR1, VR2: 20,9 variable regions. The Por B sequences were invariably of class 3 and had the allele 47 and the variable region of Fet A was F3-1. In spite of microevolution of the epidemic clones, as detected by pulsed field gel electrophoresis (19), the structural composition of the analyzed OMPs thus was strikingly stable, providing no evidence for selection of OMP antigenic variants by herd immunity.

### Sequence analysis of *opa* genes and their flanking regions

The multiple copies of genes encoding the Opa proteins can be turned on and off by slipped strand mispairing of tandem CTCTT repeats present in their open reading frames. Here we have used the published genome sequence of the serogroup A ST4 strain Z2491 to design primers specific for the flanking regions of the *opa*A, *opa*B and *opa*D genes. In contrast to strain Z2491, the ST7 and ST2859 strains analyzed here all contained an *opa*J gene, as has been described for other ST5-complex/subgroup III strains (20).

All strains analyzed here (Table 1) had identical *opa*B and *opa*D alleles (253 and 296, respectively). ST7 and ST2859 isolates had different *opa*J alleles (213 and 127, respectively). While both ST7 strains from Ghana and ST2859 strains from Burkina Faso had the *opa*A allele 242, the ST2859 isolates from Ghana had *opa*A253. No variation was observed within the three individual groups of epidemiologically related isolates.

The numbers of pentamer repeats within the coding sequences of the *opa* genes were subject to much wider variation. However, strains isolated at the same location and a similar time point tended to be relatively uniform. For example, all Ghanaian ST2859 isolates from 2007, had 8 *opa*A, 8 *opa*B and 10 *opa*D repeats, while 5/6 isolates from 2008 had 10 *opa*A, 7 *opa*B and 12 *opa*D repeats. Alleles with a number of pentamer repeats that were a multiple of three were found to be functional genes (Table 1). Disease and colonization isolates did not differ significantly in the Opa protein expression patterns.

**Discussion:**

Humans are the only natural host for *N. meningitidis* and nasopharyngeal carriage rates are generally much higher than disease rates. In a longitudinal study in the northern Ghana we have observed waves of colonization and disease with hyper-virulent clones of serogroup A meningococci. About four years after their local emergence these clones disappeared completely (4). This may be attributed to the development of herd immunity against the colonizing clone in the local population. Antibodies directed against outer membrane proteins have been implicated in the development of natural immunity against meningococci. In particular antibodies directed against the PorA and PorB proteins seem to provide serosubtype-specific protection (21). This prompted us to investigate whether immune selection pressure building up in the population during colonization waves leads to an antigenic diversification of OMPs in the colonizing meningococcal population. It has been proposed that the propensity of *N. meningitidis* to accumulate mutations increases dramatically towards the end of an epidemic, presumably due to immune pressure, and that with time, variants can arise forming new genoclouds (22). However, we detected no mutational changes in the surface proteins PorA, PorB and FetA in any of the strains investigated here. The benefit of accumulating mutations in order to escape herd immunity may thus be dispossessed by the fitness cost that such mutations may confer (22). This striking lack of diversification may also be due the lack of a genetically diverse pharyngeal flora of *N. meningitidis*, in the study population (4), not allowing for horizontal genetic exchange (23).

While all PorA, PorB and FetA had the same alleles irrespective of the time point and location of isolation as well as the sequence type of the strains examined, two of the four *opa* genes analyzed were subject to some variation. The opa gene repertoire of most meningococcal isolates comprises 3-4 loci (*opaA, opaB, opaD, and opaJ*) (20).

The expression of the full length protein is controlled at the translational level by a phase-variable pentameric repeat region within the *opa* gene open reading frame (24), (25).While the OpaA genes of the Ghanaian ST7 strains as well as the ST2859 strains from Burkina Faso had the allele 242, all Ghanaian ST2859 strains had the allele 253. ST7 and ST2859 are MLST single locus variants and the ST2859 genocloud may have developed from ST7 meningococci in Africa. The change of the *opa*A allele in the ST 2859 genocloud that emerged in Ghana in 2008 (unpublished results) may have been due to homologous recombination and immune selection against the 242 allele (26), (27). In the case of the *opa*J gene all ST7 isolates had the allele 213, whereas the ST2859 strains both from Ghana and from Burkina Faso had the allele 127.

The highest variability observed in this study was due to changing numbers of pentamer repeats in the open reading frames of the *opa* genes. Whereas there was a certain conservation of numbers of pentamer repeats in strains collected at the same location and at a similar time point, the overall variability of the number of repeats was quite high. This variation of numbers of tandem repeats is due to slipped strand mispairing and results in phase variation involving on/off expression of the respective *opa* gene (28). Antigenic variation of the Opa genes has been suggested to be mediating immune evasion (29), (27). Functional open reading frames were found for all four *opa* genes, but no more than two functional *opa* genes were found in any of the strains.

Control of meningitis epidemics has relied so far on reactive vaccination strategies with polysaccharide vaccines. While a serogroup A conjugate vaccine is currently being clinically tested, a comprehensive vaccine based on sub-capsular outer membrane proteins (OMPs) is not yet available. Future analyses of the whole genomes of epidemiologically well defined collections of isolates may give insight into the driving forces behind the micro-evolution of *N. meningitidis*. Results will help identifying antigens suitable for inclusion into a multivalent subunit vaccine (30).

**Acknowledgments**

This work was supported by the Volkswagen Foundation through the Collaborative Research Grant (SFB) 544 ‘Control of Tropical Infectious Diseases’. This publication made use of the Neisseria Multi-Locus Sequence Typing website (http://pubmlst.org/neisseria/) developed by Keith Jolley and Man-Suen Chan (Jolley et al. 2004, BMC Bioinformatics, 5:86) and sited at the University of Oxford. The development of this site has been funded by the Wellcome Trust and European Union.

## Table 1

Origin and characteristics of serogroup A meningococci analyzed. The *opa* genes that were found to be in frame are framed.

| N | Source | Origin | Time | MLST | PorA | PorB | FetA | OpaA | | OpaB | | OpaD | | Opaj | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | ST | allele | allele | allele | allele | repeats | allele | repeats | allele | repeats | allele | repeats |
| 1396 | carriage | KND | Mar.2002 | 7 | 20,9 | 3-47 | F3-1 | 242 | 14 | 253 | 12 | 296 | 9 | 213 | 7 |
| 1577 | carriage | KND | Mar.2003 | 7 | 20,9 | 3-47 | F3-1 | 242 | 13 | 253 | 12 | 296 | 10 | 213 | 8 |
| 1808 | case | KND | Jan.2004 | 7 | 20,9 | 3-47 | F3-1 | 242 | 9 | 253 | 8 | 296 | 12 | 213 | 8 |
| 1813 | case | KND | Feb.2004 | 7 | 20,9 | 3-47 | F3-1 | 242 | 9 | 253 | 8 | 296 | 12 | 213 | 8 |
| 1812 | case | KND | Feb.2004 | 7 | 20,9 | 3-47 | F3-1 | 242 | 10 | 253 | 9 | 296 | 10 | 213 | 9 |
| 1822 | case | KND | Feb.2004 | 7 | 20,9 | 3-47 | F3-1 | 242 | 14 | 253 | 12 | 296 | 10 | 213 | 9 |
| 1838 | case | KND | Mar.2004 | 7 | 20,9 | 3-47 | F3-1 | 242 | 16 | 253 | 12 | 296 | 10 | 213 | 9 |
| 1902 | carriage | KND | Mar.2004 | 7 | 20,9 | 3-47 | F3-1 | 242 | 14 | 253 | 11 | 296 | 9 | 213 | 9 |
| 1990 | carriage | KND | Mar.2004 | 7 | 20,9 | 3-47 | F3-1 | 242 | 15 | 253 | 5 | 296 | 12 | 213 | 10 |
| 1991 | carriage | KND | Nov.2004 | 7 | 20,9 | 3-47 | F3-1 | 242 | 15 | 253 | 5 | 296 | 12 | 213 | 10 |
| 2008 | case | KND | Feb.2005 | 7 | 20,9 | 3-47 | F3-1 | 242 | 11 | 253 | 9 | 296 | 13 | 213 | 9 |
| 2009 | case | KND | Feb.2005 | 7 | 20,9 | 3-47 | F3-1 | 242 | 11 | 253 | 10 | 296 | 14 | 213 | 9 |
| 2018 | carriage | KND | Mar.2005 | 7 | 20,9 | 3-47 | F3-1 | 242 | 8 | 253 | 8 | 296 | 10 | 213 | 9 |
| 2019 | carriage | KND | Mar.2005 | 7 | 20,9 | 3-47 | F3-1 | 242 | 8 | 253 | 11 | 296 | 10 | 213 | 9 |
| 2020 | carriage | KND | Mar.2005 | 7 | 20,9 | 3-47 | F3-1 | 242 | 11 | 253 | 8 | 296 | 13 | 213 | 9 |
| 2021 | carriage | KND | Mar.2005 | 7 | 20,9 | 3-47 | F3-1 | 242 | 11 | 253 | 8 | 296 | 12 | 213 | 9 |
| 2173 | case | NHD | Mar.2006 | 2859 | 20,9 | 3-47 | F3-1 | 242 | 8 | 253 | 8 | 296 | 10 | 127 | 12 |
| 2171 | case | NHD | Mar.2006 | 2859 | 20,9 | 3-47 | F3-1 | 242 | 8 | 253 | 8 | 296 | 9 | 127 | 12 |
| 2172 | case | NHD | Mar.2006 | 2859 | 20,9 | 3-47 | F3-1 | 242 | 8 | 253 | 8 | 296 | 9 | 127 | 12 |
| 2174 | case | NHD | Mar.2006 | 2859 | 20,9 | 3-47 | F3-1 | 242 | 8 | 253 | 8 | 296 | 9 | 127 | 12 |
| 2175 | case | NHD | Mar.2006 | 2859 | 20,9 | 3-47 | F3-1 | 242 | 8 | 253 | 8 | 296 | 9 | 127 | 12 |
| 2176 | case | NHD | Mar.2006 | 2859 | 20,9 | 3-47 | F3-1 | 242 | 8 | 253 | 8 | 296 | 9 | 127 | 11 |
| 2202 | carriage | NHD | Mar.2006 | 2859 | 20,9 | 3-47 | F3-1 | 242 | 9 | 253 | 8 | 296 | 10 | 127 | 13 |
| 2243 | carriage | NHD | Mar.2006 | 2859 | 20,9 | 3-47 | F3-1 | 242 | 8 | 253 | 8 | 296 | 9 | 127 | 11 |
| 2365 | carriage | NHD | Nov.2006 | 2859 | 20,9 | 3-47 | F3-1 | 242 | 7 | 253 | 9 | 296 | 7 | 127 | 13 |
| 2378 | carriage | NHD | Nov.2006 | 2859 | 20,9 | 3-47 | F3-1 | 242 | 14 | 253 | 8 | 296 | 9 | 127 | 17 |
| 2554 | carriage | NHD | Mar.2007 | 2859 | 20,9 | 3-47 | F3-1 | 242 | 10 | 253 | 11 | 296 | 9 | 127 | 9 |
| 2560 | carriage | NHD | Mar.2007 | 2859 | 20,9 | 3-47 | F3-1 | 242 | 8 | 253 | 8 | 296 | 12 | 127 | 12 |
| 2587 | carriage | NHD | Mar.2007 | 2859 | 20,9 | 3-47 | F3-1 | 242 | 8 | 253 | 8 | 296 | 12 | 127 | 8 |
| 2537 | case | Bawku | Mar.2007 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 8 | 253 | 8 | 296 | 10 | 127 | 11 |
| 2539 | case | Bawku | Mar.2007 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 8 | 253 | 8 | 296 | 10 | 127 | 12 |
| 2541 | case | Bawku | Mar.2007 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 8 | 253 | 8 | 296 | 10 | 127 | 12 |
| 2545 | case | KND | Apr.2007 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 8 | 253 | 8 | 296 | 10 | 127 | 12 |
| 2502 | carriage | KND | Apr.2007 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 8 | 253 | 8 | 296 | 10 | 127 | 12 |
| 2620 | carriage | KND | Nov.2007 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 8 | 253 | 8 | 296 | 10 | 127 | 12 |
| 2622 | carriage | KND | Nov.2007 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 8 | 253 | 8 | 296 | 10 | 127 | 12 |
| 2624 | carriage | KND | Nov.2007 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 8 | 253 | 8 | 296 | 10 | 127 | 12 |
| 2626 | carriage | KND | Nov.2007 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 8 | 253 | 8 | 296 | 10 | 127 | 12 |
| 2628 | carriage | KND | Nov.2007 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 8 | 253 | 8 | 296 | 10 | 127 | 12 |
| 2630 | carriage | KND | Nov.2007 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 8 | 253 | 8 | 296 | 10 | 127 | 12 |
| 2632 | carriage | KND | Nov.2007 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 8 | 253 | 8 | 296 | 10 | 127 | 12 |
| 2635 | carriage | KND | Nov.2007 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 8 | 253 | 8 | 296 | 10 | 127 | 12 |
| 2669 | carriage | KND | Nov.2007 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 8 | 253 | 8 | 296 | 10 | 127 | 12 |
| 2671 | carriage | KND | Nov.2007 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 8 | 253 | 8 | 296 | 10 | 127 | 12 |
| 2699 | carriage | KND | Apr.2008 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 10 | 253 | 7 | 296 | 12 | 127 | 12 |
| 2700 | carriage | KND | Apr.2008 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 10 | 253 | 7 | 296 | 12 | 127 | 12 |
| 2701 | carriage | KND | Apr.2008 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 10 | 253 | 7 | 296 | 12 | 127 | 12 |
| 2703 | carriage | KND | Apr.2008 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 10 | 253 | 7 | 296 | 12 | 127 | 12 |
| 2708 | case | KND | Apr.2008 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 10 | 253 | 7 | 296 | 12 | 127 | 12 |
| 2707 | case | KND | Feb.2008 | 2859 | 20,9 | 3-47 | F3-1 | 253 | 8 | 253 | 8 | 296 | 12 | 127 | 10 |

## References

1. Caugant DA. Genetics and evolution of Neisseria meningitidis: importance for the epidemiology of meningococcal disease. Infect. Genet. Evol. 2008 Sep;8(5):558-565.

2. Greenwood B. Manson Lecture. Meningococcal meningitis in Africa. Trans. R. Soc. Trop. Med. Hyg. 1999 Aug;93(4):341-353.

3. Stephens DS, Greenwood B, Brandtzaeg P. Epidemic meningitis, meningococcaemia, and Neisseria meningitidis. Lancet. 2007 Jun 30;369(9580):2196-2210.

4. Leimkugel J, Hodgson A, Forgor AA, Pflüger V, Dangy J, Smith T, et al. Clonal waves of Neisseria colonisation and disease in the African meningitis belt: eight- year longitudinal study in northern Ghana. PLoS Med. 2007 Mar 27;4(3):e101.

5. Nicolas P, Norheim G, Garnotel E, Djibo S, Caugant DA. Molecular epidemiology of neisseria meningitidis isolated in the African Meningitis Belt between 1988 and 2003 shows dominance of sequence type 5 (ST-5) and ST-11 complexes. J. Clin. Microbiol. 2005 Oct;43(10):5129-5135.

6. Teyssou R, Muros-Le Rouzic E. Meningitis epidemics in Africa: a brief overview. Vaccine. 2007 Sep 3;25 Suppl 1:A3-7.

7. Sié A, Pflüger V, Coulibaly B, Dangy JP, Kapaun A, Junghanss T, et al. ST2859 serogroup A meningococcal meningitis outbreak in Nouna Health District, Burkina Faso: a prospective study. Trop. Med. Int. Health. 2008 Jun;13(6):861-868.

8. Jolley KA, Brehony C, Maiden MCJ. Molecular typing of meningococci: recommendations for target choice and nomenclature. FEMS Microbiol. Rev. 2007 Jan;31(1):89-96.

9. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U.S.A. 1998 Mar 17;95(6):3140-3145.

10. Brehony C, Jolley KA, Maiden MCJ. Multilocus sequence typing for global surveillance of meningococcal disease. FEMS Microbiol. Rev. 2007 Jan;31(1):15-26.

11. Jolley KA, Wilson DJ, Kriz P, McVean G, Maiden MCJ. The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in Neisseria meningitidis. Mol. Biol. Evol. 2005 Mar;22(3):562-569.

12. Caugant DA, Maiden MCJ. Meningococcal carriage and disease--population biology and evolution. Vaccine. 2009 Jun 24;27 Suppl 2:B64-70.

13. Hill DJ, Griffiths NJ, Borodina E, Virji M. Cellular and molecular biology of Neisseria meningitidis colonization and invasive disease. Clin. Sci. 2010 May;118(9):547-564.

14. Gagneux S, Hodgson A, Ehrhard I, Morelli G, Genton B, Smith T, et al. Microheterogeneity of serogroup A (subgroup III) Neisseria meningitidis during an outbreak in northern Ghana. Trop. Med. Int. Health. 2000 Apr;5(4):280-287.

15. Russell JE, Jolley KA, Feavers IM, Maiden MCJ, Suker J. PorA variable regions of Neisseria meningitidis. Emerging Infect. Dis. 2004 Apr;10(4):674-678.

16. Sacchi CT, Lemos AP, Whitney AM, Solari CA, Brandt ME, Melles CE, et al. Correlation between serological and sequencing analyses of the PorB outer membrane protein in the Neisseria meningitidis serotyping system. Clin. Diagn. Lab. Immunol. 1998 May;5(3):348-354.

17. Thompson EAL, Feavers IM, Maiden MCJ. Antigenic diversity of meningococcal enterobactin receptor FetA, a vaccine component. Microbiology (Reading, Engl.). 2003 Jul;149(Pt 7):1849-1858.

18. Parkhill J, Achtman M, James KD, Bentley SD, Churcher C, Klee SR, et al. Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491. Nature. 2000 Mar 30;404(6777):502-506.

19. Leimkugel J, Forgor AA, Dangy J, Pflüger V, Gagneux S, Hodgson A, et al. Genetic diversification of Neisseria meningitidis during waves of colonization and disease in the meningitis belt of sub-Saharan Africa. Vaccine. 2007 Sep 3;25 Suppl 1:A18-23.

20. Callaghan MJ, Jolley KA, Maiden MCJ. Opacity-associated adhesin repertoire in hyperinvasive Neisseria meningitidis. Infect. Immun. 2006 Sep;74(9):5085-5094.

21. Jordens JZ, Williams JN, Jones GR, Christodoulides M, Heckels JE. Development of immunity to serogroup B meningococci during carriage of Neisseria meningitidis in a cohort of university students. Infect. Immun. 2004 Nov;72(11):6503-6510.

22. Achtman M. Population structure of pathogenic bacteria revisited. Int. J. Med. Microbiol. 2004 Sep;294(2-3):67-73.

23. Achtman M. Epidemic spread and antigenic variability of Neisseria meningitidis. Trends Microbiol. 1995 May;3(5):186-192.

24. Kawula TH, Aho EL, Barritt DS, Klapper DG, Cannon JG. Reversible phase variation of expression of Neisseria meningitidis class 5 outer membrane proteins and their relationship to gonococcal proteins II. Infect. Immun. 1988 Feb;56(2):380-386.

25. Stern A, Meyer TF. Common mechanism controlling phase and antigenic variation in pathogenic neisseriae. Mol. Microbiol. 1987 Jul;1(1):5-12.

26. Levin BR, Cornejo OE. The population and evolutionary dynamics of homologous gene recombination in bacterial populations. PLoS Genet. 2009 Aug;5(8):e1000601.

27. Callaghan MJ, Buckee CO, Jolley KA, Kriz P, Maiden MCJ, Gupta S. The effect of immune selection on the structure of the meningococcal opa protein repertoire. PLoS Pathog. 2008 Mar;4(3):e1000020.

28. Murphy GL, Connell TD, Barritt DS, Koomey M, Cannon JG. Phase variation of gonococcal protein II: regulation of gene expression by slipped-strand mispairing of a repetitive DNA sequence. Cell. 1989 Feb 24;56(4):539-547.

29. Davidsen T, Tønjum T. Meningococcal genome dynamics. Nat. Rev. Microbiol. 2006 Jan;4(1):11-22.

30. Tan LKK, Carlone GM, Borrow R. Advances in the development of vaccines against Neisseria meningitidis. N. Engl. J. Med. 2010 Apr 22;362(16):1511-1520.

# CHAPTER 4: Genetic diversity of *Streptococcus agalactiae*

## Molecular characterisation of colonizing and invasive isolates of *Streptococcus agalactiae* from an urban hospital in Kenya

Charlotte Huber 1, Valentin Pflueger 2, Francis McOdimba 3, Claudia Daubenberger 1‡, Gunturu Revathi 3‡*

‡Both authors contributed equally

1 Swiss Tropical and Public Health Institute

Department of Medical Parasitology and Infection Biology

Socinstrasse 57

4002 Basel, Switzerland

2 Mabritec AG

Lörracherstrasse 50

P.O. Box 320, 4125 Riehen

Switzerland

3 Aga Khan University Hospital,

3rd Parklands Avenue

P.O. Box 30270-00100 Nairobi,

Kenya

Corresponding author: Prof Gunturu Revathi

To be submitted to the Journal of Clinical Microbiology

## Abstract

*Streptococcus agalactiae*, also referred to as group B streptococcus (GBS) is traditionally considered a neonatal pathogen. However, GBS is also emerging as a significant cause of morbidity in adults. In this study we examined the population structure of a clinically well documented collection of *S. agalactiae* isolates sampled between 2007 and 2010 at the Aga Khan Hospital in Nairobi, Kenya. This collection included 98 carrier and 75 clinical isolates. Using MLST and molecular serotyping, we identified 22 Sequence types, including 5 novel sequence types mainly associated with disease. A strong correlation between sequence types and distinct capsular serotypes was found with the disease isolates showing higher variability compared to the carrier isolates. The most prevalent sequence type was ST-23 in both clinical and carrier isolates, accounting for 26.6% in total. 99.4 % of the isolates were of one of the five previously described capsular serotypes Ia, Ib, II, III and V.

## Introduction

*Streptococcus agalactiae*, also known as group B streptococci (GBS) are Gram-positive, β-hemolytic, chain-forming cocci that are traditionally considered to be primarily a pathogen affecting neonates. However, *S. agalactiae* is also emerging as a pathogen responsible for significant morbidity in pregnant women and in non-pregnant adults, especially in the elderly and persons suffering from underlying conditions such as diabetes and cancer [1], [2]. In newborns, *S. agalactiae* causes sepsis, pneumonia, meningitis, and less frequently focal infections such as osteomyelitis, septic arthritis or cellulite [3]. In pregnant women, GBS cause urinary tract infection, chorio-amnionitis, endometritis, bacteremia and most likely stillbirth. Diseases in adults caused by *S. agalactiae* include mainly skin and soft-tissue infection, bacteremia, urinary tract infections, pneumonia and osteomyelitis. Meningitis, endocarditits and the streptococcal toxic shock syndrome are rare but serious clinical syndromes of *S. agalactiae* infections [1], [2], [4], [5]. *S. agalactiae* colonizes asymptomatically the lower gastrointestinal and genitourinary tracts of 30-

50 % of healthy adults [6], and an estimated 20 - 30 % of all pregnant women are carriers. *S. agalactiae* can be isolated from swabs of the vagina or the rectum and prenatal recto-vaginal screening for colonization of pregnant women is recommended [3]. Intra-partum antibiotic prophylaxis has been introduced successfully in a number of Western Countries to prevent vertical transmission to the neonate during delivery [7], [8].

The development of a multi-locus sequence based typing scheme (MLST) for *S. agalactiae* allows unambiguous comparison of the population structures of *S. agalactiae* among different geographical areas [9]. This MLST uncovers sequence variation among seven conserved housekeeping genes, classifying strains into numerous clones, or sequence types (STs). Some STs can be grouped together into clusters or clonal complexes (CCs) following phylogenetic analyses. Four STs were found to be the most prevalent in a global collection of strains, ST-1, ST-17, ST-19, and ST-23 [9]. ST-17 and strains of its clonal complex were strongly associated with neonatal disease in several populations and may have an enhanced invasiveness [9], [10]. MLST has been used to investigate the population structure of *S. agalactiae* in different regions of the United Kingdom [9], [11], the United States [12], Canada [13] Poland [14], Sweden [15], Portugal [16], France [17], Italy [17], Norway [18], and Israel [19]. Limited epidemiological studies from Kenya [20], South Africa [21], Malawi [22], Senegal and Central African Republic [23] suggest that *S. agalactiae* is emerging also in Africa as an important cause morbidity and mortality.

Imperi et al. have developed a multiplex PCR assay for the direct identification of the 10 known capsular serotypes (Ia to IX) of *S. agalactiae* that correlates with the conventional typing results [24]. Five serotypes (Ia, Ib, II, III and V) account for 96% of neonatal disease and 88% of invasive disease in adults in the US [25]. In the UK, the same five serotypes were accounting for 98% of neonatal disease and 94% of carrier isolates in pregnant women [11].

It has been shown that placental transfer of maternal antibodies against GBS capsular polysaccharides can protect infants from invasive GBS infection [1]. The introduction of a GBS vaccine for pregnant mothers has the potential to prevent GBS induced diseases in the newborn and efforts towards developing a subunit vaccine are pursued [26], [27]. Capsular polysaccharides representing each of the five major serotypes in the US have been conjugated to a protein carrier and have shown immuno-genicity in clinical vaccine trials [28]. A pentavalent conjugate vaccine might therefore be sufficient to induce antibodies against the majority of serotypes circulating in the US [28], [29], as well as England [11]. In order to develop an effective vaccine against *S. agalactiae* in different areas of the world, a precise characterization of the genetic diversity and the distribution of serotypes circulating within human populations in the respective regions is needed (http://www.who.int/vaccine_research/diseases/soa_bacterial/en/index6.html).

In this investigation we examined the population structure of a collection of 173 isolates of *S. agalactiae*. These samples have been collected and documented at the Aga Khan University Hospital of Nairobi, Kenya over the last 4 years. To our knowledge, this is the first report of GBS molecular epidemiology in urban East Africa.

## Materials and Methods

Ethical clearance for this study was obtained from the Ethikkommission beider Basel (EKBB).

The *S. agalactiae* isolates were routinely sampled and identified at the AKUH, Nairobi, Kenya from both in-patients and out patients. The species identification was performed using colony morphology, Gram staining, the CAMP test, as well as by performing the slide agglutination test of the bacterial meningitis kit (Welcogen) with specific Group B latex.

Duplicates of the isolates were transferred to the Swiss TPH. The isolates were incubated overnight at 37°C on Columbia Agar sheep blood plus plates (Oxoid). Single colonies were then selected and plated in order to grow overnight for MALDI-TOF MS confirmation of identity. A small amount of bacteria was suspended in 25 % formic acid and immobilized on the MALDI-TOF MS target by addition of a matrix (Sinapic acid or 3,5,-dimethyoxy-4-hydroxy cinnamic acid), in order to crystallizes the whole cell and brake the cellular membranes. The Samples were measured using the Axima Confidence™ (Shimadzu, Japan) with an automated measurement method. Using the SARAMIS software, automated computer-aided identification of bacteria was achieved by comparing mass spectra for individual samples against the SARAMIS SuperSpectra™ database (Anagnostec, Germany).

For MLST, strains were grown overnight at 37°C on Columbia Agar sheep blood plus plates (Oxoid), genomic DNA was extracted from the bacterial pellets using the Wizard® Genomic DNA Purification Kit (Promega AG, Duebendorf, Switzerland) and the MLST PCR amplification and sequencing was performed with the standard primers as described. [9]. Briefly, PCR reactions were done with 5µl of 10 × BD buffer and 1µl of FirePol Taq polymerase, 1.25 mM $MgCl_2$ (Solis BioDyne, Tartu, Estonia), 100 ng of genomic DNA or the equivalent volume of nuclease-free water as a negative control, a 0.2 µM concentration of each forward and reverse primer,

and a 0.2 mM concentration of each deoxynucleoside triphosphate in a total volume of 50 µl. The PCR reactions were run in a T Professional Basic PCR machine (Biometra GmbH, Goettingen Germany). PCR products were analyzed on 1% agarose gels by gel electrophoresis and then subjected to purification and direct sequencing on an ABI3730 XL automatic DNA sequencer (Macrogen, Seoul, South Korea). All alleles were sequenced on both strands and aligned with the ABI Prism Auto assembler, version 1.4.0 (Perkin-Elmer, Waltham, MA). Novel alleles and sequence types were amplified and sequenced twice independently and then submitted for allele assignment. The allele types and ST were determined by making use of the *S. agalactiae* MLST website (http://pubmlst.org/ sagalactiae/) developed by Man-Suen Chan and Keith Jolley and sited at the University of Oxford [30]. The neighbour joining tree was generated from the allelic profiles obtained using Start2 [31].

The multiplex PCR assay for the identification of the capsular types prevalent in our collection was performed as previously described. A positive control for all 10 capsular serotypes was included into each of the multiplex PCR assay [24].

## Results

A collection of 173 GBS isolates originating from Kenya were typed using MLST [9], and twenty two sequence types (STs) could be resolved (Fig. 1). Isolates were grouped into clonal complexes (CC) by use of eBURST software (by relaxing the group definition to six out of seven shared alleles), and thirteen STs were clustered in five CC, namely CC23, CC19, CC17, CC10 and CC1. The most prevalent CC was CC23 (27.7%), followed by CC17 (22.0%), CC10 (16.2%), CC19 (13.3%) and CC1 (12.1%). 91 % of the isolates could be grouped to one of these five CCs. Singletons not clustering were identified as ST-3, ST-4, ST-24, ST-26 ST-103, ST-327, ST-328, ST-485 and ST-486 (Table 1). Five novel STs could be found in this collection and were named ST-484, ST-485, ST-486, ST-492 and ST-501. All newly found STs are single or double locus variants of a known ST, namely ST-17, ST-103 or ST-23.

98 GBS isolates obtained from asymptomatic carriers and 75 clinical case isolates were MLST typed. The carrier samples could be grouped into 16 STs while the clinical isolates resolved into 19 STs (Table 1). The ST-4, ST-103, ST-327, ST-485, ST-492 and ST-501 were present exclusively in the clinical cases. ST-3, ST-24 and ST-8 could be detected only among the carriers (Table 1). ST-484, a novel single locus variant (SLV) of ST-17 was found 8 times in the case isolates and 2 times in the carrier isolates.

Preliminary work on GBS MLST typing using a global collection of isolates [9] has shown that capsular serotype does not strictly follow the ST. We used the PCR based capsular typing method [24] for assigning the capsular serotypes in our collection. In total, six different serotypes were detected namely Ia, Ib, II, III, IV and V. The most common serotype was serotype III (31.2%), followed by serotype Ia (30.0%), serotype V (17.3%), serotype Ib (13.9%), serotype II (6.9%), and 1 isolate was of the serotype IV (0.6%) (Table 1). A close correlation between capsular serotype and ST could be observed. Out of the 46 strains that were of the ST-23, 43 were serotype Ia. All 28 isolates of the ST-17 were of serotype III. 18 out of the 20

isolates that had the ST-10 were of the serotype Ib. 15 out of the 16 isolates with ST-1 were of serotype V. All ST-484 and ST-182 isolates had the serotype III. Interestingly, *S. agalactiae* obtained from clinical cases showed a higher general variability in the capsular serotypes of the same ST compared to the carrier *S. agalactiae* like ST-23, ST-12, ST-10 and ST-1 (Figure 2).

Next, we analysed the relationship between clinical disease and STs and serotypes (Table 2). Samples obtained from urine showed a clear bias towards ST-23, ST-17, ST-10 and ST-1. Interestingly, out of the thirteen isolates obtained from blood, the novel ST-484, ST-485, ST-486 and ST-492 represented 38 % (Table 2) indicating that these STs clearly have pathogenic potential.

**Discussion:**

A detailed analysis of the population structure of S. agalactiae in Africa has not been performed so far. In this work we used the well established MLST in combination with molecular serotyping to characterize strains collected from asymptomatic carriers and clinical cases at the Aga Khan University Hospital in Nairobi in the years 2007 to 2010. Our collection consisted of 98 carrier and 72 adult disease isolates as well as 3 invasive isolates from infants.

Using MLST, the 173 GBS isolates investigated resolved into 22 STs out of which 5 STs had not been previously described. 5 CCs (CC23, CC19, CC17, CC10 and CC1) were accounting for 91.3% of our isolates. These 5 CCs were also highly prevalent in other regions of the globe, such as Sweden [15], Italy [17], England [11], the US [12], the Central African Republic, Senegal [23] as well as Israel. In Israel, a high prevalence of ST-22 was also detected [19]. In a global collection of strains isolates of ST-22 were exclusively found in Israel [9] and ST-22 was absent in our collection. CC-26 which represented 15% of all identified isolates in the Central African Republic and Senegal [23] was only found three times in our collection.

Five capsular serotypes (Ia, Ib, II, III, V) were accounting for 99.4 % of all isolates in our collection. These serotypes were also commonly present in other regions of the world, including the Central African Republic and Senegal [23], England [11], Norway [18], the Unites States [25], Israel [19] and Korea [32].

Similarly to previous reports [23], [15], we observed an overall correlation between genotype defined by MLST and the capsular serotype. However, a certain diversification could be detected, and in general, the clinical isolates displayed a higher variability of both sequence types and corresponding capsular serotypes.

Also, it is worth noting that all of the 5 novel STs including ST-484, ST-485, ST-486, ST-492 and ST-501, were at least found once in disease isolates and only ST-484 and ST-486 were also found in carrier isolates. The most prevalent novel ST was ST-484, a SLV of ST-17, accounting for 10 isolates. In agreement with previous reports [33], [34], ST-17 was invariably associated with the capsular serotype III, as was ST-484. The ST-17 complex is thought to have relatively recently arisen from a bovine ancestor [34] and appears to have diverged independently with an exclusive set of virulence characteristics [35]. The enhanced invasiveness of ST-17 has been proposed to be specific to neonates and independent of capsular serotype [11]. In our study the GBS samples had mainly been collected from adults, and only 8 of the 28 isolates of the ST-17 were associated with disease. However, 8 of the 10 strains of the ST-484 were invasive isolates, including two neonatal blood isolates, three urine isolates, two pus isolates and one isolate from fat tissue. Two novel sequence types had emerged from ST-23, the most frequently detected ST in our collection. ST-492 and ST-501, both SLVs of ST-23 had the serotype Ia and were derived from adult disease isolates. ST-103, its SLV ST-486 as well as its double locus variant ST-485 were of the capsular serotype Ia, and one isolate of each ST was found in adult blood isolates of our collection. One strain of the ST-486 was also found amongst the carrier isolates. ST-103 had previously been isolated from humans as well as animals. One invasive strain of the ST-103 was found in an elderly person and carrier strains had been isolated from pregnant women [11], [35]. Animal strains

of the ST-103 had been found in a guinea pig, a cat and a cow [33]. Having its habitat in both humans and animals, *S. agalactiae* may be provided with a great available gene pool for lateral gene transfer [36],[37]. It has been demonstrated that GBS DNA segments of up to 334 kb can be transferred through conjugation, and that large DNA exchanges may contribute to the genome dynamics in the natural population challenging the GBS vaccine development [38], [27]. A pentavalent conjugate vaccine including the capsular serotypes Ia, Ib, II, III and V has been proposed for the US [29], and such a vaccine may provide protection against the majority of serotypes circulating in different areas of the world, including the Central African Republic, Senegal [23], England [11], Norway [18], Israel [19] and Korea [32], as well as our setting. Other serotypes have been reported to arise however [39], [32], and a vaccine including the capsular serotypes Ia, Ib, II, III, and V may not be suitable in regions like Japan [40]. Efforts have been undertaken to develop a protein based vaccine against GBS, using genomic and proteomic-based approaches, but no vaccine is on the market to date [41], [26], [27].

**Figure 1:**

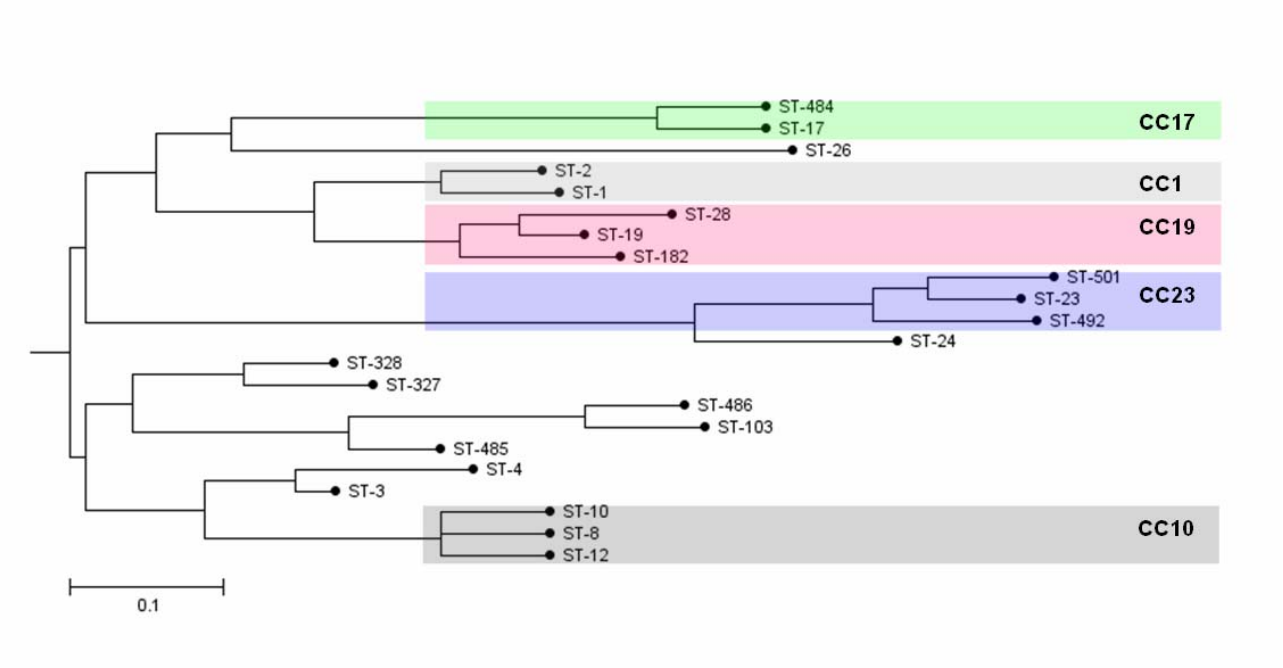



Figure 1 Neighbour-joining tree showing genetic relationships between 22 STs of group B Streptococcus isolated from patients with asymptomatic carriage, neonates and adults. The phylogeny is based on MLST data for seven housekeeping genes.

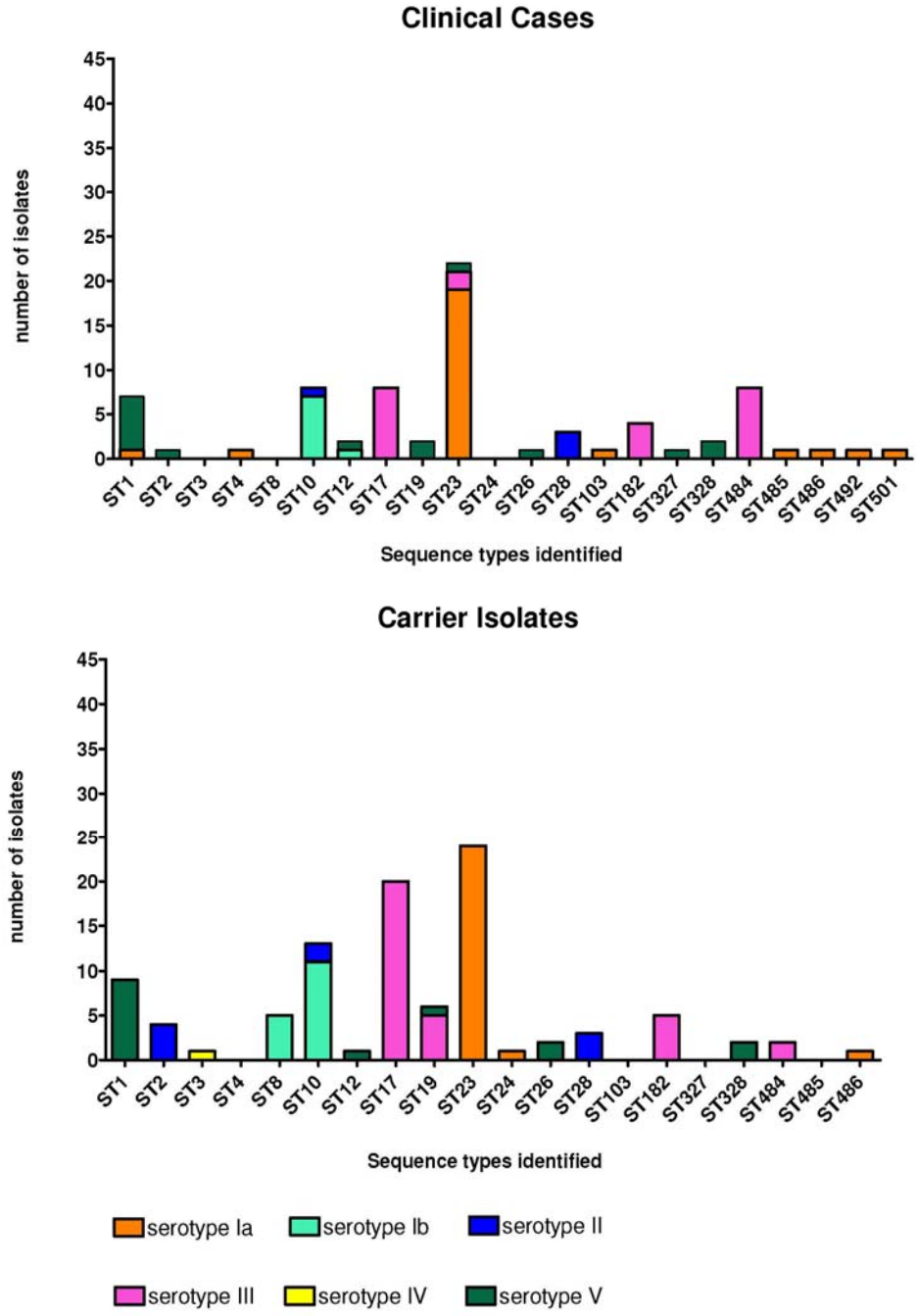



**Figure 2: Overview of distribution of STs with their capsular serotypes for carrier and clinical case isolates of S. agalactiae.**

In the 75 clinical isolates investigated, 19 different ST could be observed while in the 98 carrier isolates investigated, 16 ST were identified. For the capsular serotyptes, 6 and 5 different serotypes were found in carrier and case isolates, respectively. A higher variety of ST and corresponding capsular serotypes was detected in clinical vs carrier isolates.

| ST | No. of total isolates (%) | No. of Isolates (Carrier) | No. of Isolates (Clinical cases) | Serotypes (No. of total isolates) | Serotypes (No. of carrier isolates) | Serotypes (No. of clinical case isolates) |
|---|---|---|---|---|---|---|
| **1** | 16 (9.2) | 9 | 7 | V (15), Ia (1) | V (9) | V (6), Ia (1) |
| **2** | 5 (2.9) | 4 | 1 | II (4), V (1) | II (4) | V (1) |
| **3** | 1 (0.6) | 1 | 0 | IV (1) | IV (1) | |
| **4** | 1 (0.6) | 0 | 1 | Ia (1) | | Ia (1) |
| **8** | 5 (2.8) | 5 | 0 | Ib (5) | Ib (5) | |
| **10** | 20 (11.6) | 12 | 8 | Ib (18), II (2) | Ib (11), II (1) | Ib (7), II (1) |
| **12** | 3 (1.7) | 1 | 2 | V (2), Ib (1) | V (1) | V (1), Ib (1) |
| **17** | 28 (16.2) | 20 | 8 | III (28) | III (20) | III (8) |
| **19** | 8 (4.6) | 6 | 2 | III (5), V (3) | III (5), V (1) | V (2) |
| **23** | 46 (26.6) | 24 | 22 | Ia (43), III (2), V (1) | Ia (24) | Ia (19), III (2), V (1) |
| **24** | 1 (0.6) | 1 | 0 | Ia (1) | Ia (1) | |
| **26** | 3 (1.7) | 2 | 1 | V (3) | V (2) | V (1) |
| **28** | 6 (3.5) | 3 | 3 | II (6) | II (3) | II (3) |
| **103** | 1 (0.6) | 0 | 1 | Ia (1) | | Ia (1) |
| **182** | 9 (5.2) | 5 | 4 | III (9) | III (5) | III (4) |
| **327** | 1 (0.6) | 0 | 1 | V (1) | | V (1) |
| **328** | 4 (2.3) | 2 | 2 | V (4) | V (2) | V (2) |
| **484** | 10 (5.8) | 2 | 8 | III (10) | III (2) | III (8) |
| **485** | 1 (0.6) | 0 | 1 | Ia (1) | | Ia (1) |
| **486** | 2 (1.2) | 1 | 1 | Ia (2) | Ia (1) | Ia (1) |
| **492** | 1 (0.6) | 0 | 1 | Ia (1) | | Ia (1) |
| **501** | 1 (0.6) | 0 | 1 | Ia (1) | | Ia (1) |

TABLE 1. Overview of the ST and capsular serotypes obtained from cases and carriers in the current study.

| ST | No. of total clinical isolates | Capsular serotypes | Samples from urine with serotypes | | Samples from blood with serotypes | | Samples from Pus with serotypes | | Samples from other body locations with serotypes | |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 7 | V (6), Ia (1) | 4 | V (4) | | | 1 | Ia (1) | 2 | V (2) |
| **2** | 1 | V (1) | | | | | | | 1 | V (1) |
| **4** | 1 | Ia (1) | 1 | Ia (1) | | | | | | |
| **10** | 8 | Ib (7), II (1) | 5 | Ib (5) | | | 1 | II (1) | 2 | Ib (2) |
| **12** | 2 | V (1), Ib (1) | | | 1 | V (1) | 1 | Ib (1) | | |
| **17** | 8 | III (8) | 7 | III (7) | 1 | III (1) | | | | |
| **19** | 2 | V (2) | 2 | V (2) | | | | | | |
| **23** | 22 | Ia (19), III (2), V (1) | 10 | Ia (9), III (1) | 3 [1] | Ia (3) | 3 | Ia (2), III (1) | 6 | Ia (5), V (1) |
| **26** | 1 | V (1) | | | | | | | 1 | V (1) |
| **28** | 3 | II (3) | 3 | | | | | | | |
| **103** | 1 | Ia (1) | | | 1 | Ia (1) | | | | |
| **182** | 4 | III (4) | 2 | III (2) | 2 | III (2) | | | | |
| **327** | 1 | V (1) | | | | | | | 1 | V (1) |
| **328** | 2 | V (2) | 2 | V (2) | | | | | | |
| **484** | 8 | III (8) | 3 | III (3) | 2[2] | III (2) | 2 | III (2) | 1 | III (1) |
| **485** | 1 | Ia (1) | | | 1 | Ia (1) | | | | |
| **486** | 1 | Ia (1) | | | 1 | Ia (1) | | | | |
| **492** | 1 | Ia (1) | | | 1 | Ia (1) | | | | |
| **501** | 1 | Ia (1) | 1 | Ia (1) | | | | | | |

1; 1 infant blood sample, 2; both are infant blood samples.

Table 2: Distribution of clinical isolates of different body location origin given shown according to their STs and serotypes. Samples obtained from urine, blood and pus samples are listed separately in columns while the column with the other samples include: Amniotic fluid (1 isolate) ST-1, serotype V; Cervix swab (2 isolates, ST-23, serotype V and Ia); Endocervical swabs (2 isolates, ST-23, serotype Ia and ST-26, serotype V); Fat tissue (1 isolate, ST-484, serotype III); Genital swab, (ST-10, serotype Ib); semen (1 isolate, ST-23,serotype Ia); Urethral swabs (6 isolates,, ST-327, serotype V, ST-23, serotype Ia (2 isolates), ST-10, serotype Ib, ST-2, serotype V, ST-1, serotype V).

## References:

[1] Rajagopal L. Future microbiology 2009;4:201-221.

[2] Skoff TH, Farley MM, Petit S, Craig AS, Schaffner W, Gershman K, Harrison LH, Lynfield R, Mohle-Boetani J, Zansky S, Albanese BA, Stefonek K, Zell ER, Jackson D, Thompson T, Schrag SJ. Clinical infectious diseases 2009;49:85-92.

[3] Gibbs RS, Schrag S, Schuchat A. Obstet gynecol 2004;104:1062-1076.

[4] Farley MM. Clin. infect. dis 2001;33:556-561.

[5] Ulett KB, Benjamin WH, Zhuo F, Xiao M, Kong F, Gilbert GL, Schembri MA, Ulett GC. J clin microbiol 2009;47:2055-2060.

[6] van der Mee-Marquet N, Fourny L, Arnault L, Domelier A, Salloum M, Lartigue M, Quentin R. J clin microbiol 2008;46:2906-2911.

[7] Schrag SJ, Zywicki S, Farley MM, Reingold AL, Harrison LH, Lefkowitz LB, Hadler JL, Danila R, Cieslak PR, Schuchat A. N. engl. j. med 2000;342:15-20.

[8] Spaetgens R, DeBella K, Ma D, Robertson S, Mucenski M, Davies HD. Obstet gynecol 2002;100:525-533.

[9] Jones N, Bohnsack JF, Takahashi S, Oliver KA, Chan M, Kunst F, Glaser P, Rusniok C, Crook DWM, Harding RM, Bisharat N, Spratt BG. J clin microbiol 2003;41:2530-2536.

[10] Lin FC, Whiting A, Adderson E, Takahashi S, Dunn DM, Weiss R, Azimi PH, Philips JB, Weisman LE, Regan J, Clark P, Rhoads GG, Frasch CE, Troendle J, Moyer P, Bohnsack JF. J. clin. microbiol 2006;44:1257-1261.

[11] Jones N, Oliver KA, Barry J, Harding RM, Bisharat N, Spratt BG, Peto T, Crook DW. Clin. infect. dis 2006;42:915-924.

[12] Bohnsack JF, Whiting A, Gottschalk M, Dunn DM, Weiss R, Azimi PH, Philips JB, Weisman LE, Rhoads GG, Lin FC. J. clin. microbiol 2008;46:1285-1291.

[13] Manning SD, Springman AC, Lehotzky E, Lewis MA, Whittam TS, Davies HD. J. clin. microbiol 2009;47:1143-1148.

[14] Sadowy E, Matynia B, Hryniewicz W. J. antimicrob. chemother. 2010:dkq230.

[15] Luan S, Granlund M, Sellin M, Lagergård T, Spratt BG, Norgren M. J. clin. microbiol 2005;43:3727-3733.

[16] Martins ER, Pessanha MA, Ramirez M, Melo-Cristino J. J clin microbiol 2007;45:3224-3229.

[17] Gherardi G, Imperi M, Baldassarri L, Pataracchia M, Alfarone G, Recchia S, Orefici G, Dicuonzo G, Creti R. J. clin. microbiol 2007;45:2909-2916.

[18] Bergseng H, Afset JE, Radtke A, Loeseth K, Lyng RV, Rygg M, Bergh K. Clin. microbiol. infect 2009;15:1182-1185.

[19] Bisharat N, Jones N, Marchaim D, Block C, Harding RM, Yagupsky P, Peto T, Crook DW. Microbiology (reading, engl.) 2005;151:1875-1881.

[20] English M, Ngama M, Musumba C, Wamola B, Bwika J, Mohammed S, Ahmed M, Mwarumba S, Ouma B, McHugh K, Newton C. Arch. dis. child 2003;88:438-443.

[21] Madhi SA, Radebe K, Crewe-Brown H, Frasch CE, Arakere G, Mokhachane M, Kimura A. Ann trop paediatr 2003;23:15-23.

[22] Gray KJ, Bennett SL, French N, Phiri AJ, Graham SM. Emerg infect dis 2007;13:223-229.

[23] Brochet M, Couvé E, Bercion R, Sire J, Glaser P. J clin microbiol 2009;47:800-803.

[24] Imperi M, Pataracchia M, Alfarone G, Baldassarri L, Orefici G, Creti R. J. microbiol. methods 2010;80:212-214.

[25] Phares CR, Lynfield R, Farley MM, Mohle-Boetani J, Harrison LH, Petit S, Craig AS, Schaffner W, Zansky SM, Gershman K, Stefonek KR, Albanese BA, Zell ER, Schuchat A, Schrag SJ. Jama 2008;299:2056-2065.

[26] Maione D, Margarit I, Rinaudo CD, Masignani V, Mora M, Scarselli M, Tettelin H, Brettoni C, Iacobini ET, Rosini R, D'Agostino N, Miorin L, Buccato S, Mariani M, Galli G, Nogarotto R, Nardi Dei V, Vegni F, Fraser C, Mancuso G, Teti G, Madoff LC, Paoletti LC, Rappuoli R, Kasper DL, Telford JL, Grandi G. Science 2005;309:148-150.

[27] Johri AK, Paoletti LC, Glaser P, Dua M, Sharma PK, Grandi G, Rappuoli R. Nat. rev. microbiol 2006;4:932-942.

[28] Baker CJ, Edwards MS. Arch. dis. child 2003;88:375-378.

[29] Edwards MS. Hum vaccin 2008;4:444-448.

[30] Jolley K, Chan M, Maiden M. Bmc bioinformatics 2004;5:86.

[31] Jolley KA, Feil EJ, Chan MS, Maiden MC. Bioinformatics 2001;17:1230-1231
.

[32] Seo YS, Srinivasan U, Oh K, Shin J, Chae JD, Kim MY, Yang JH, Yoon H, Miller B,

DeBusscher J, Foxman B, Ki M. J. korean med. sci 2010;25:817-823.

[33] Brochet M, Couvé E, Zouine M, Vallaeys T, Rusniok C, Lamy M, Buchrieser C, Trieu-Cuot P, Kunst F, Poyart C, Glaser P. Microbes infect 2006;8:1227-1243.

[34] Bisharat N, Crook DW, Leigh J, Harding RM, Ward PN, Coffey TJ, Maiden MC, Peto T, Jones N. J. clin. microbiol 2004;42:2161-2167.

[35] Springman AC, Lacher DW, Wu G, Milton N, Whittam TS, Davies HD, Manning SD. J. bacteriol 2009;191:5419-5427.

[36] Lefébure T, Stanhope MJ. Genome biol 2007;8:R71.

[37] Marri PR, Hao W, Golding GB. Mol. biol. evol 2006;23:2379-2391.

[38] Brochet M, Rusniok C, Couvé E, Dramsi S, Poyart C, Trieu-Cuot P, Kunst F, Glaser P. Proc. natl. acad. sci. u.s.a 2008;105:15961-15966.

[39] Diedrick MJ, Flores AE, Hillier SL, Creti R, Ferrieri P. J. clin. microbiol 2010;48:3100-3104.

[40] Lachenauer CS, Kasper DL, Shimada J, Ichiman Y, Ohtsuka H, Kaku M, Paoletti LC, Ferrieri P, Madoff LC. J. infect. dis 1999;179:1030-1033.

[41] Doro F, Liberatori S, Rodríguez-Ortega MJ, Rinaudo CD, Rosini R, Mora M, Scarselli M, Altindis E, D'Aurizio R, Stella M, Margarit I, Maione D, Telford JL, Norais N, Grandi G. Mol cell proteomics 2009;8:1728-1737.

# CHAPTER 5: DISCUSSION

## General discussion

Humans have approximately ten times more bacteria associated with them than they have human cells in the body. The bacteria of the normal flora of the human are usually found in the exposed parts of the body such as the skin, the mucosal surfaces of the nose, the mouth and intestinal and urogenital tracts. Although bacteria of the normal flora may be harmful if they spread into previously sterile parts of the body, the normal flora prevents pathogen colonization by occupying almost all of the available ecological niches. Besides out-competing pathogens for living space, the normal flora of the skin produces fatty acids, discouraging other species from invading. Gut bacteria release a number of factors with antibacterial activity, and vaginal lactobacilli maintain an acid environment, suppressing growth of other organisms.

Commensals are usually harmless organisms residing on the body of larger species. Many bacteria, not only the normal flora but also neisserial, streptococcal as well as other bacterial species may live in coexistence with the human host without causing harm, but there are numerous ones that quite commonly cause disease. However, microorganisms are generally confronted with the antimicrobial defenses of its host and have evolved ways to overcome the defense mechanisms of both the innate and adaptive immune system.

In the work presented in this thesis, we were investigating the genetic variability of *Mycobacterium ulcerans, Neisseria meningitidis,* as well as *Streptococcus agalactiae*. The aim of this study was to contribute to the understanding of how genetic variablility allows these bacterial pathogens to evade the immune system of the human host and the challenge that genetic variability may pose for vaccine development.

## Mycobacterium ulcerans

*M. ulcerans* has arisen from *Mycobacterium marinum* by horizontal transfer of a virulence plasmid coding for the enzymes needed for mycolactone production, followed by reductive evolution. In our study, we have analyzed a worldwide collection of *M. ulcerans* strains belonging to the ancestral as well as the classical lineage. The ancestral lineage has been previously reported to be phylogenetically closer to its progenitor *M. marinum* than members of the *M. ulcerans* classical lineage [1]. While being present in the sequenced *M. marinum* strain M (http://www.sanger.ac.uk/cgi-bin/BLAST/submitblast/m_marinum), genes encoding ESAT-6, CFP-10, and HspX are absent from the genome of the sequenced Ghanaian *M. ulcerans* strain Agy99 (http://genopole.pasteur.fr/Mulc/BuruList.html) The genome of the *M. marinum* strain M harbours two *esxB-esxA* gene clusters coding for ESAT-6 and CFP-10 at distant chromosomal locations, in the so-called MURD4 (*M. marinum* M genome position 218302 to 230285) and MURD152 (*M. marinum* M genome position 6489253 to 6592034) [2]. In our study, all analyzed *M. ulcerans* strains belonging to the ancestral lineage were found to have lost only one copy of the *esxB-esxA* cassette. The strains that had been collected in Asia or South America had lost the MURD4 copy wheareas the Mexican strain had lost the MURD152 copy. Both copies were absent in the *M. ulcerans* strains that belonged to the classical lineage. The hspX gene was found to be present in the strains of the ancestral lineage and absent in the classical lineage. Additionally, pseudogenization of the hspX gene had taken place in the Mexican strain. The loss of these immunodominant proteins may have helped *M. ulcerans* to bypass the host's immunological response.

Insights into the biological functions of mycolactone, Esat-6, Cfp-10 and HspX may help explain why the acquisition of the mycolactone producing plasmid pMUM001 has rendered the virulence factors ESAT-6, CFP-10 and HspX superfluous [3].

Mycolactone is a diffusible cytotoxin with immunomodulatory properties, playing a key role in the development of the pathology of *M. ulcerans* infection [4], [5]. The cytotoxic action of mycolactone affects a broad range of cell types in different intensity, also depending on the state of maturation of the target cells. Adipose cells are extremely sensitive to mycolactone and undergo apoptosis at low toxin concentrations [6], whereas other cell types may cope better. Mycolactone has structural similarities with immunosuppressive drugs such as FK506 and rapamycin, and at nontoxic concentrations, mycolactone has been described to selectively suppress dendritc cell functions, limiting both the initiation of primary immune responses and the recruitment of inflammatory cells to the infection site [4], [7]. In human monocytes, mycolactone suppresses the expression of tumor necrosis factor alpha (TNF-α), interleukin-2, (IL-2) and interleukin-10 (IL-10) in vitro [8]. Mycolactone has also been described to impair the production of TNF-α as well as interferon γ (IFN- γ) in macrophages, contributing to immune evasion by inhibiting phagocytosis [9], [10], [11]. A recent study described a characteristic suppression of circulating chemokines in Buruli patients. Also, the capacity to produce Th1, Th2 and Th17 cytokines was largely impaired in ex vivo stimulation assays [12].

The *M. tuberculosis* proteins ESAT-6, CFP-10 and hspX are strong T- and B-cell immunogens, making them suitable for immunodiagnostic tests [13], [14], [15].
ESAT-6 and CFP-10 play a complex role in the immune response in TB. In addition to the immunodominance of their epitopes, ESAT-6 and CFP-10 may activate macrophages, dendritic cells and mast cells to release proinflammatory mediators, and induce the differentiation and maturation of dendritic cells, resulting in a specific Th1 response. However, ESAT-6 and CFP-10 may also be partly responsible for mycobacterial immune evasion by impairment of macrophage functions as well as dendritic cell capacity for optimal induction of a specific Th1 response [13], [16]. HspX is essential for the survival of *M. tuberculosis* during persistence and is targeted by the immune system during latent infection in humans, inducing very strong T- and B-cell responses [14].

A vaccine against M. ulcerans would protect persons at risk in highly endemic areas. However, whilst the immunogenicity of ESAT-6, CFP-10 and HspX makes these proteins suitable for vaccine development in *M. tuberculosis* [17], [18], they are not expressed by classical-lineage *M. ulcerans* strains that are prevalent in Africa and Australia where the vast majority of Buruli cases are found worldwide. In order to know if an effective vaccine against *M. ulcerans* should mainly stimulate a humoral or cellular immune response, the question remains to be resolved whether *M. ulcerans* is essentially an intracellular or extracellular pathogen. It is most probable that both arms of the immune response are required for optimal protection [19].

## Neisseria meningitidis

A common mechanism of masking bacterial surfaces is the expression of a carbohydrate capsule [20], and although carrier isolates of *N.meningitidis* may be encapsulated or unencapsulated, disease isolates are invariably encapsulated. Thirteen meningococcal serogroups have been described, and the majority of disease is caused by one of the five capsule types A, B, C, W-135 and Y. Interestingly, the capsular polysaccharides of the serogroups B, C, W-135 and Y contain sialic acid, commonly present on cell surfaces of the human, enabling bacteria to become less visible to the immune system. The immune response against the serogroup B capsule is particularly poor due to the structural identity of its sialic acid homopolymer with a component of the human NCAM (neural cell-adhesion molecule). In Europe and the Americas, meningococcal disease is predominantly caused by the serogroups B and C [21], whereas the large epidemics in Africa are associated with serogroup A [22].

In this current investigation we were typing the genes of the subcapsular antigens PorA, PorB and FetA of fifty different *Nm* strains of the serogroup A and ST-7 or 2859 respectively, collected at the same locations either in Ghana or Burkina Faso over multiple time points. No sequence diversification was detected for PorA, PorB and FetA throughout the waves of colonization and disease irrespective of sequence type or country of origin. Although antibodies directed against PorA and PorB had been shown to be associated with the development of an immune response following meningococcal carriage and disease [23], [24], [25] and FetA had also been demonstrated to be immunogenic [26], [27], our results do not indicate immune selection against PorA, PorB and FetA. The homogeneity of these surface proteins may partially be due to the lack of a genetically diverse pharyngeal *Neisseria* flora in the African population, not allowing for horizontal genetic exchange [28], [29].

We also investigated the sequences of the four *opa* genes OpaA, OpaB, OpaD and OpaJ together with their flanking regions in all the strains included in this study. The highest variability was found due to changing numbers of pentamer repeats within the open reading frame of the *opa* genes. The *opa* gene sequences were translated into predicted protein sequences, and all the *opa* alleles that had a number of pentamer repeats that were a multiple of three were found to be functional genes. It is worth noting that no more than two *opa* genes were found to be in frame in any of the strains investigated. Additionally, only one of the fifty strains was found to have all the four *opa* genes out of frame, indicating an essential role of the opa proteins for their survival on the host. It has been reported that certain opa proteins have the ability to suppress the activation of CD4(+) T cells when binding to the human carcinoembryonic antigen cellular adhesion molecule 1 (CEACAM1) in *Neisseria gonorrhoeae* [30].

A limited variability besides the change in the number of pentamer repeats could be detected in the open reading frame of the OpaA genes as well as the OpaJ genes. The OpaA genes of the Ghanaian strains of the ST-7 as well as the Burkina Fasian strains of the sequence type 2859 had the allele 242. The OpaA alleles of the Ghanaian strains of the ST-2859 differed from the OpaA alleles of the other strains investigated and had the allele 253, which was the same allele which was found for the OpaB gene in all of the strains studied. This change of the OpaA allele that had taken place between the Ghanaian strains of the ST-7 and the ST-2859 may have been due to homologous recombination and immune selection against the 242 allele [31], [32]. However, none of the opa 253 alleles were found to be in frame in any of the strains of the ST-2859, also indicating immune selection against the 253 allele.

The OpaJ genes of the strains of the ST-7 had the allele 213 and all the strains of the ST-2859 had the allele 127. This change of the opaJ allele may have been partially responsible for the replacement of the ST7 wave by the ST2859 wave [33], [29], and it is worth noting that 76% of all the *opa* alleles 127 were found to be in frame.

Altogether the results of this study demonstrate a striking lack of diversification of the surface proteins investigated. In order to more deeply understand the waves of colonization and disease of the meningococci, whole genome approaches may be needed, possibly in combination with the investigation of meningococcal protein expression patterns [34].

Meningitis epidemics in sub-Saharan Africa may be countered based on early detection of cases and emergency reactive vaccination of the population at risk with meningococcal polysaccharide vaccines, although the downside of most polysaccharide vaccines is their failure to induce immunological memory and their lack of immunogenicity in young children [35]. Conjugation of polysaccharides to proteins change the nature of the immune response induced from T-independent to T-dependent, and polysaccharide conjugate vaccines are effective in infants and induce immunological memory [36]. It had been decided to develop an affordable mono-valent serogroup A conjugate vaccine, based on the expectation that other serogroups, including W135 strains will not cause massive epidemics in sub-Saharan Africa, but it is becoming increasingly clear that other serogroups may have to be included [35].

Outer membrane proteins such as PorA, PorB and FetA are candidates in the search for comprehensive meningococcal vaccines. However, whilst the allele types of PorA, PorB and FetA often correlate with the MLST genotype of a meningococcal strain [37], [38], these outer membrane proteins are highly diverse between strains of different clonal complexes [38]. The technique of reverse vaccinology is used to identify genes from the *N. meningitidis* genome encoding potential surface exposed protein antigens to develop a safe and effective vaccine against serogroup B *N. meningitidis* [39], [40], with the ultimate aim to develop a universal vaccine against the pathogen [41].

## Streptococcus agalactiae

In this work we were performing a detailed analysis of the population structure of GBS strains collected from 98 asymptomatic carriers and 75 clinical cases at the Aga Khan University Hospital in Nairobi in the years 2007 to 2010, using the well established MLST in combination with molecular serotyping.

Using MLST, the 173 GBS isolates investigated in our study resolved into 22 STs out of which 5 STs had not been previously described. The high diversification of GBS may be explained by the broad habitat range of *S. agalactiae* [42], [43], and its high potential for genetic recombination [44].

The population structure of the GBS isolates in our Kenyan collection, defined by MLST and capsular serotyping, was found to be similar to the GBS population structure in many other regions of the world. Five clonal complexes (CCs) (CC23, CC19, CC17, CC10 and CC1) were accounting for 91.3% of our isolates. These 5 CCs were also highly prevalent in other regions of the globe, such as Sweden [45], Italy [46], England [47], the US [48], the Central African Republic, Senegal [49] as well as Israel [50]. Five capsular serotypes (Ia, Ib, II, III, V) were accounting for 99.4 % of all isolates in our collection. These serotypes were also commonly present in other regions of the world, including the Central African Republic and Senegal [49], England [47], Norway [51], the Unites States [52], Israel [50] and Korea [53]. Similarly to previous reports [49], [45], we observed an overall correlation between genotype defined by MLST and the capsular serotype. However, a certain diversification could be detected, and in general, the clinical isolates displayed a higher variability of both sequence types and corresponding capsular serotypes. Also, it is worth noting that the 5 novel STs were more commonly found in the case isolates than in the carrier isolates.

GBS have ten known capsular serotypes which are heavily sialylated, providing resistance to components of the host's innate immune response. The sialylation of the GBS capsule results in the impairment of the C3 deposition on the bacterial cell

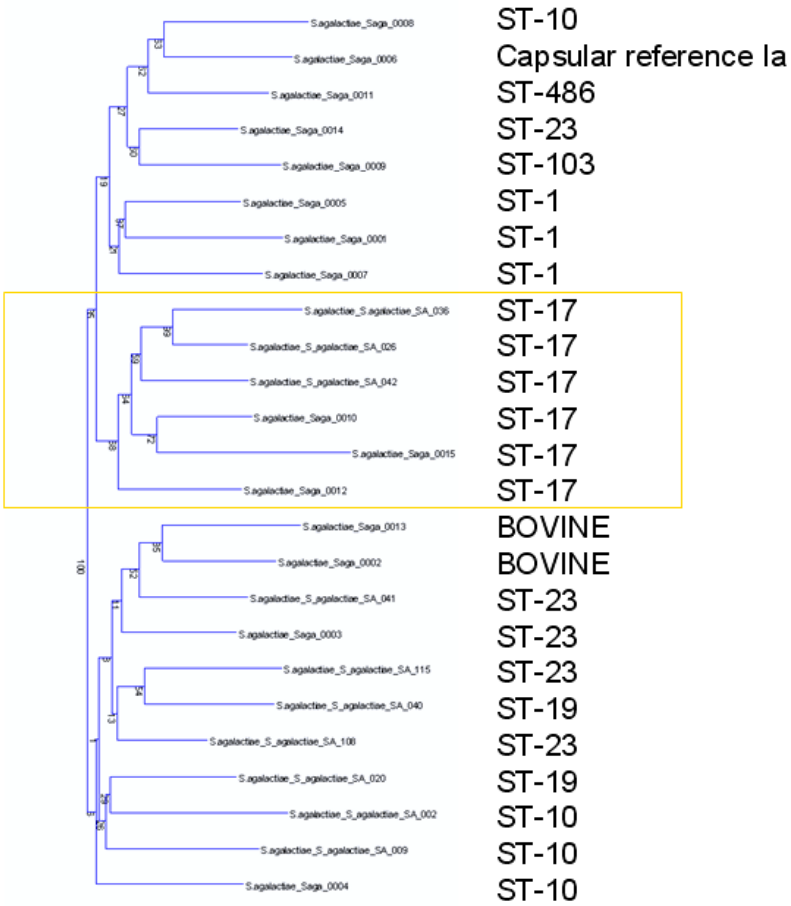
surface, thus preventing the activation of the complement alternative pathway [54], [55]. Studies have shown though that placental transfer of maternal antibodies can protect infants from invasive GBS infection [56], and conjugate vaccines of polysaccharides complexed to highly immunogenic proteins have been demonstrated to mount immune responses against GBS capsular polysaccharides [55], [57].

A pentavalent conjugate vaccine including the capsular serotypes Ia, Ib, II, III and V has been proposed for the US [58]. A vaccine including the serotypes Ia, Ib, II, III and V may also provide protection against the majority of serotypes circulating in different areas of the world, including the Central African Republic, Senegal [49], England [47], Norway [51], Israel [50] and Korea [53], as well as our setting. A vaccine including the capsular serotypes Ia, Ib, II, III, and V may not be suitable in regions like Japan though [59], and other serotypes have been reported to arise in Korea as well as the US [53], [60].

To develop a broadly protective vaccine against GBS, the genome sequences of eight *S. agalactiae* isolates were analyzed, and 312 surface proteins were subsequently tested as vaccine candidates. Four proteins elicited protection in mice and their combination proved highly protective against a high number of strains including all circulating serotypes [61]. Alternatively, a proteomics-based approach allowing the identification of surface-exposed proteins was applied to GBS. Whole viable bacterial cells of the hypervirulent serotype III strain COH1 were treated with proteases, and the released peptides were identified using mass spectroscopy. A novel highly expressed antigen, SAN_1485 was found and used for protection studies in the offspring of immunized adult female mice. SAN_1485 conferred good protection in the pups when challenged with the COH1 strain as well as the heterologous M781 strain [62].

## Outlook

MLST has a high discriminatory power, however, it is time consuming and cost intensive, and MALDI TOF MS analysis has been suggested as a more rapid and less costly alternative for bacterial typing [63], [64]. Here, we were performing the typing of 22 GBS isolates of known STs that had been sampled at the Aga Khan hospital in Nairobi, Kenya, using MALDI TOF MS. Two bovine strains and one capsular reference strain was also included in the study.




A phylogenetic tree was obtained. The 6 strains of the ST-17, as well as the three strains of the ST-1 and the two bovine stains clustered together, whereas ST-10, ST-19 and ST-23 appeared to be more heterogenic.

GBS strains of the ST-17 have previously been described to be more homogenous than strains of other STs [65], and it has been proposed that the number of loci selected for the GBS MLST scheme may be insufficient for strains of the clonal complex 23 [66].

MLST is a well-established method to study bacterial populations of sufficient genetic variability, and high quality Sanger sequencing is commonly used to sequence individual positions of a genome. However, the technology is expensive and too slow for sequencing extended genomic regions [66], [67]. To meet the increased demand for sequencing, several so-called second-generation sequencing systems such as Solexa and 454 sequencing technologies have been developed. These developments have significantly reduced the cost of sequencing whilst simultaneously increasing the DNA sequencing speed. The main downsides of these sequencing technologies are short read-lengths, higher error-rates, and the difficulty of managing massive amounts of data [68]. The informatics challenges of data obtained by second-generation sequencing technologies are mostly due to the short reads characteristic of these technologies. Novel single-molecule DNA sequencing technologies have the potential to yield reads that are in excess of 10 kilobases, easing many of the informatics challenges relating to assembly that are now experienced. However, raw read error rates may remain high and the increased information content will demand new types of mathematical models and algorithms [69].

## References

[1] Käser M, Rondini S, Naegeli M, Stinear T, Portaels F, Certa U, Pluschke G. Bmc evol. biol 2007;7:177.

[2] Tekaia F, Gordon SV, Garnier T, Brosch R, Barrell BG, Cole ST. Tuber. lung dis 1999;79:329-342.

[3] Demangel C, Stinear TP, Cole ST. Nat. rev. microbiol 2009;7:50-60.

[4] Schütte D, Pluschke G. Expert opin biol ther 2009;9:187-200.

[5] Phillips R, Sarfo FS, Guenin-Macé L, Decalf J, Wansbrough-Jones M, Albert ML, Demangel C. J. infect. dis 2009;200:1675-1684.

[6] Dobos KM, Small PL, Deslauriers M, Quinn FD, King CH. Infect. immun 2001;69:7182-7186.

[7] Coutanceau E, Decalf J, Martino A, Babon A, Winter N, Cole ST, Albert ML, Demangel C. J. exp. med 2007;204:1395-1403.

[8] Pahlevan AA, Wright DJ, Andrews C, George KM, Small PL, Foxwell BM. J. immunol 1999;163:3928-3935.

[9] Torrado E, Adusumilli S, Fraga AG, Small PLC, Castro AG, Pedrosa J. Infect. immun 2007;75:3979-3988.

[10] Coutanceau E, Marsollier L, Brosch R, Perret E, Goossens P, Tanguy M, Cole ST, Small PLC, Demangel C. Cell. microbiol 2005;7:1187-1196.

[11] Torrado E, Fraga AG, Logarinho E, Martins TG, Carmona JA, Gama JB, Carvalho MA, Proença F, Castro AG, Pedrosa J. J. immunol 2010;184:947-955.

[12] Phillips R, Sarfo FS, Guenin-Macé L, Decalf J, Wansbrough-Jones M, Albert ML, Demangel C. J. infect. dis 2009;200:1675-1684.

[13] Trajkovic V, Natarajan K, Sharma P. Microbes infect 2004;6:513-519.

[14] Roupie V, Romano M, Zhang L, Korf H, Lin MY, Franken KLMC, Ottenhoff THM, Klein MR, Huygen K. Infect. immun 2007;75:941-949.

[15] Demissie A, Leyten EMS, Abebe M, Wassie L, Aseffa A, Abate G, Fletcher H, Owiafe P, Hill PC, Brookes R, Rook G, Zumla A, Arend SM, Klein M, Ottenhoff THM, Andersen P, Doherty TM. Clin. vaccine immunol 2006;13:179-186.

[16] Ganguly N, Siddiqui I, Sharma P. Tuberculosis (edinb) 2008;88:510-517.

[17] Gao H, Yue Y, Hu L, Xu W, Xiong S. Vaccine 2009;27:5313-5319.

[18] Shi C, Chen L, Chen Z, Zhang Y, Zhou Z, Lu J, Fu R, Wang C, Fang Z, Fan X. Vaccine 2010;28:5237-5244.

[19] Huygen K, Adjei O, Affolabi D, Bretzel G, Demangel C, Fleischer B, Johnson RC, Pedrosa J, Phanzu DM, Phillips RO, Pluschke G, Siegmund V, Singh M, van der Werf TS, Wansbrough-Jones M, Portaels F. Med. microbiol. immunol 2009;198:69-77.

[20] Finlay BB, McFadden G. Cell 2006;124:767-782.

[21] Hill DJ, Griffiths NJ, Borodina E, Virji M. Clin. sci 2010;118:547-564.

[22] Caugant DA. Infect. genet. evol 2008;8:558-565.

[23] Williams JN, Skipp PJ, O'Connor CD, Christodoulides M, Heckels JE. Infect. immun 2009;77:5080-5089.

[24] Jones GR, Christodoulides M, Brooks JL, Miller AR, Cartwright KA, Heckels JE. J. infect. dis 1998;178:451-459.

[25] Williams JN, Jones GR, Christodoulides M, Heckels JE. J. infect. dis 2003;187:1433-1441.

[26] Kortekaas J, Pettersson A, van der Biezen J, Weynants VE, van der Ley P, Poolman J, Bos MP, Tommassen J. Vaccine 2007;25:72-84.

[27] Ala'Aldeen DA, Davies HA, Borriello SP. Vaccine 1994;12:535-541.

[28] Leimkugel J, Hodgson A, Forgor AA, Pflüger V, Dangy J, Smith T, Achtman M, Gagneux S, Pluschke G. Plos med 2007;4:e101.

[29] Achtman M. Trends microbiol 1995;3:186-192.

[30] Lee HSW, Ostrowski MA, Gray-Owen SD. J. immunol 2008;180:6827-6835.

[31] Levin BR, Cornejo OE. Plos genet 2009;5:e1000601.

[32] Callaghan MJ, Buckee CO, Jolley KA, Kriz P, Maiden MCJ, Gupta S. Plos pathog 2008;4:e1000020.

[33] Norheim G, Rosenqvist E, Aseffa A, Yassin MA, Mengistu G, Kassu A, Fikremariam D, Tamire W, Høiby EA, Alebel T, Berhanu D, Merid Y, Harboe M, Caugant DA. J. clin. microbiol 2006;44:861-871.

[34] Mendum TA, Newcombe J, McNeilly CL, McFadden J. Plos one n.d.;4.

[35] Okoko BJ, Idoko OT, Adegbola RA. Vaccine 2009;27:2023-2029.

[36] Girard MP, Preziosi M, Aguado M, Kieny MP. Vaccine 2006;24:4692-4700.

[37] Callaghan MJ, Jolley KA, Maiden MCJ. Infect. immun 2006;74:5085-5094.

[38] Urwin R, Russell JE, Thompson EAL, Holmes EC, Feavers IM, Maiden MCJ. Infect. immun 2004;72:5955-5962.

[39] Pizza M, Scarlato V, Masignani V, Giuliani MM, Aricò B, Comanducci M, Jennings GT, Baldi L, Bartolini E, Capecchi B, Galeotti CL, Luzzi E, Manetti R, Marchetti E, Mora M, Nuti S, Ratti G, Santini L, Savino S, Scarselli M, Storni E, Zuo P, Broeker M, Hundt E, Knapp B, Blair E, Mason T, Tettelin H, Hood DW, Jeffries AC, Saunders NJ, Granoff DM, Venter JC, Moxon ER, Grandi G, Rappuoli R. Science 2000;287:1816-1820.

[40] Bai X, Borrow R. Expert rev vaccines 2010;9:1203-1217.

[41] Serruto D, Serino L, Masignani V, Pizza M. Vaccine 2009;27:3245-3250.

[42] Lefébure T, Stanhope MJ. Genome biol 2007;8:R71.

[43] Marri PR, Hao W, Golding GB. Mol. biol. evol 2006;23:2379-2391.

[44] Brochet M, Rusniok C, Couvé E, Dramsi S, Poyart C, Trieu-Cuot P, Kunst F, Glaser P. Proc. natl. acad. sci. u.s.a 2008;105:15961-15966.

[45] Luan S, Granlund M, Sellin M, Lagergård T, Spratt BG, Norgren M. J. clin. microbiol 2005;43:3727-3733.

[46] Gherardi G, Imperi M, Baldassarri L, Pataracchia M, Alfarone G, Recchia S, Orefici G, Dicuonzo G, Creti R. J. clin. microbiol 2007;45:2909-2916.

[47] Jones N, Oliver KA, Barry J, Harding RM, Bisharat N, Spratt BG, Peto T, Crook DW. Clin. infect. dis 2006;42:915-924.

[48] Bohnsack JF, Whiting A, Gottschalk M, Dunn DM, Weiss R, Azimi PH, Philips JB, Weisman LE, Rhoads GG, Lin FC. J. clin. microbiol 2008;46:1285-1291.

[49] Brochet M, Couvé E, Bercion R, Sire J, Glaser P. J clin microbiol 2009;47:800-803.

[50] Bisharat N, Jones N, Marchaim D, Block C, Harding RM, Yagupsky P, Peto T, Crook DW. Microbiology (reading, engl.) 2005;151:1875-1881.

[51] Bergseng H, Afset JE, Radtke A, Loeseth K, Lyng RV, Rygg M, Bergh K. Clin. microbiol. infect 2009;15:1182-1185.

[52] Phares CR, Lynfield R, Farley MM, Mohle-Boetani J, Harrison LH, Petit S, Craig AS, Schaffner W, Zansky SM, Gershman K, Stefonek KR, Albanese BA, Zell ER, Schuchat A, Schrag SJ. Jama 2008;299:2056-2065.

[53] Seo YS, Srinivasan U, Oh K, Shin J, Chae JD, Kim MY, Yang JH, Yoon H, Miller B, DeBusscher J, Foxman B, Ki M. J. korean med. sci 2010;25:817-823.

[54] Severi E, Hood DW, Thomas GH. Microbiology (reading, engl.) 2007;153:2817-2822.

[55] Rajagopal L. Future microbiology 2009;4:201-221.

[56] Baker CJ, Kasper DL. N. engl. j. med 1976;294:753-756.

[57] Baker CJ, Edwards MS. Arch. dis. child 2003;88:375-378.

[58] Edwards MS. Hum vaccin 2008;4:444-448.

[59] Lachenauer CS, Kasper DL, Shimada J, Ichiman Y, Ohtsuka H, Kaku M, Paoletti LC, Ferrieri P, Madoff LC. J. infect. dis 1999;179:1030-1033.

[60] Diedrick MJ, Flores AE, Hillier SL, Creti R, Ferrieri P. J. clin. microbiol 2010;48:3100-3104.

[61] Maione D, Margarit I, Rinaudo CD, Masignani V, Mora M, Scarselli M, Tettelin H, Brettoni C, Iacobini ET, Rosini R, D'Agostino N, Miorin L, Buccato S, Mariani M, Galli G, Nogarotto R, Nardi Dei V, Vegni F, Fraser C, Mancuso G, Teti G, Madoff LC, Paoletti LC, Rappuoli R, Kasper DL, Telford JL, Grandi G. Science 2005;309:148-150.

[62] Doro F, Liberatori S, Rodríguez-Ortega MJ, Rinaudo CD, Rosini R, Mora M, Scarselli M, Altindis E, D'Aurizio R, Stella M, Margarit I, Maione D, Telford JL, Norais N, Grandi G. Mol cell proteomics 2009;8:1728-1737.

[63] Murray PR. Clin. microbiol. infect 2010;16:1626-1630.

[64] Wolters M, Rohde H, Maier T, Belmar-Campos C, Franke G, Scherpe S, Aepfelbacher M, Christner M. Int j med microbiol 2010.

[65] Bisharat N, Crook DW, Leigh J, Harding RM, Ward PN, Coffey TJ, Maiden MC, Peto T, Jones N. J. clin. microbiol 2004;42:2161-2167.

[66] Sørensen UBS, Poulsen K, Ghezzo C, Margarit I, Kilian M. Mbio 2010;1.

[67] Baker S, Hanage WP, Holt KE. Curr. opin. microbiol 2010;13:640-645.

[68] Gupta PK. Trends biotechnol 2008;26:602-611.

[69] Schadt EE, Turner S, Kasarskis A. Hum. mol. genet 2010;19:R227-240.

# CURRICULUM VITAE

| | | |
|---|---|---|
| Name and Address | | Huber, Charlotte<br>Rütistrasse 40<br>8032 Zürich<br>Switzerland |
| Date of birth | | July 18, 1973 |
| Place of birth | | Winterthur |
| Nationality | | Swiss |


| | | |
|---|---|---|
| Education | 1994-2001 | ETH Zurich, study of pharmacy |
| | Nov 2001 | Swiss Federal pharmacist diploma |
| Languages | | German (mother tongue)<br>English (“Certificate of Proficiency in English”)<br>French (fluent)<br>Italian (basic knowledge) |
| Work Experience | 2002-2003 | Ship Pharmacist on the Mercy Ship Anastasis in Europe (shipyard, procurement and publicity), as well as in West Africa (Sierra Leone and Togo) where free surgical procedures to indigenous people were provided |
| | Oct 2003 | Election to the board of Mercy Ships Switzerland |
| | 2004-2005 | Coordination of drug procurement for Mercy Ships International, including purchase of drugs as well as coordination of shipments of donated medicines in Switzerland |
| | March 2005 until March 2006 | Pharmacist with Sun Store MED AG (In the German -as well as the French-speaking part of Switzerland) |
| | Mai 2006 – Dec 2010 | Doctoral Thesis at the Swiss TPH, Basel, Switzerland<br>Publication: Huber CA, Ruf MT, Pluschke G, Käser M „Multiple Loss of Immunogenic Proteins in Mycobacterium ulcerans Suggests Immune Evasion“ Clin Vaccine Immunol. 2008 Apr;15(4):598-606.<br><br>Thesis committee: Prof. G. Pluschke, Swiss TPH<br>PD Dr. C. Daubenberger, Swiss TPH<br>Prof. U. Certa, Roche<br>Prof. R. Brun, Swiss TPH |