Innovative DNA microarray design for bacterial flora composition evaluation

Huyghe, Antoine

UNIVERSITÉ DE GENÈVE

FACULTÉ DE MÉDECINE
Service des Maladies Infectieuses — Professeur Jacques Schrenzel

FACULTÉ DES SCIENCES
Département d'informatique — Professeur Bastien Chopard

Innovative DNA microarray design for bacterial flora composition evaluation

THÈSE
présentée à la Faculté des sciences de l'Université de Genève
pour obtenir le grade de Docteur ès sciences, mention bioinformatique

par

Antoine HUYGHE

de

Lille (France)

Thèse N° 4159

GENÈVE
2009

UNIVERSITÉ
DE GENÈVE

FACULTÉ DES SCIENCES

Doctorat ès sciences
mention bioinformatique

Thèse de *Monsieur Antoine HUYGHE*

intitulée :

" Innovative DNA Microarray Design for Bacterial Flora Composition Evaluation "

La Faculté des sciences, sur le préavis de Messieurs J. SCHRENZEL, professeur et codirecteur de thèse (Faculté de médecine, Service des maladies infectieuses), B. CHOPARD, professeur ordinaire et codirecteur de thèse (Département d'informatique), P. LINDER, professeur ordinaire (Faculté de médecine, Département de microbiologie et de médecine moléculaire) et J. FREY, professeur (Institut de bactériologie vétérinaire, VetSuisse, Université de Berne, Suisse), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 4 décembre 2009

Thèse - 4159 -

Le Doyen, Jean-Marc TRISCONE

N.B.- La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "Informations relatives aux thèses de doctorat à l'Université de Genève".

Nombre d'exemplaires à livrer par colis séparé à la Faculté : - 4 -

Remerciements - Acknowledgements

Je tiens chaleureusement à remercier le Professeur Daniel Lew pour m'avoir accueilli au sein du service des maladies infectieuses de l'hôpital cantonal de Genève durant ma thèse.

Chaleureux remerciement au Professeur Jacques Schrenzel pour m'avoir encadré durant toutes ces années et pour sa disponibilité. Merci pour ton enthousiasme, tes nombreux conseils et tes idées motivantes.

Je souhaite remercier le Professeur Bastien Chopard pour son encadrement et ses conseils.

Je tiens également à remercier les membres du jury pour la relecture et le temps consacré.

Merci au Docteur Patrice François pour m'avoir soutenu et encadré durant ces années. Je te remercie pour tes nombreuses idées et le temps que tu as passé avec moi à discuter de science et de plein d'autres choses, disons moins scientifiques. Merci également pour ton optimisme très communicatif, et pour toujours avoir trouvé des solutions aux problèmes qui se sont posés.

Aussi, je tiens à remercier tous les membres du Genomic Research Laboratory pour leur bonne humeur et disponibilité: Manuela Tangomo, Myriam Girard, Eve-Julie Bonetti, Dr Katerine Whiteson, Marie Beaume, Adrien Fischer, Ludwig Stenz et Dr Vladimir Lazarevic.

Je tiens particulièrement à remercier le Docteur David Hernandez pour son soutien et ses nombreux conseils durant ma thèse. Merci aussi pour m'avoir supporté (au sens français du terme) dans la promiscuité du labo durant ces derniers mois.

Je souhaite remercier ma famille, et plus particulièrement mes parents et grands-parents, pour leur soutien infaillible et pour m'avoir motivé les moments opportuns.

Merci à Charles et Christian pour leur amitié.

Akhirnya, saya ingin berterima kasih kepada Tina atas pertolongannya yang penuh dengan kasih sayang dan keceriaan. Terima kasih kepadamu kerana memberi motivasi diwaktu saya memerlukannya. Kamu memberikan saya kekuatan untuk menyelesaikan kerjaku

Foreword

This thesis starts with a general introduction (Chapter 1) on microarrays. The first part of the introduction introduces various topics such as the nature of the probes, available platforms, probe design and possible applications of microarrays. It is followed by a review on the characterization of microbial pathogens using microarrays. This review was published in Infection, Genetics and Evolution in 2008 and presents further applications of microarrays.

The second part of the introduction details the development of the two microarrays that were used to study a disease of unknown etiology (noma).

The results of this thesis have been already published or will be submitted soon (Chapter 2). These articles are first summarized and put in the perspective of this thesis, and then directly included.

The results are followed by a conclusion (Chapter 3) completed with perspectives.

The contribution section (Chapter 4) contains various articles related to the use of microarrays for which I contributed during my thesis but not directly in the scope of the main topic. The articles are first summarized (abstract) and then directly included.

Note: Author's contributions are underlined in the cited references included in the text of this manuscript.

Résumé

Conception de puces à ADN novatrices, dédiées à l'évaluation de la composition de la flore bactérienne

Au cours de la précédente décennie, l'avènement de nouvelles techniques moléculaires a permis d'énormes progrès en biologie, notamment avec le développement des puces à ADN. Cette technologie permet de mesurer simultanément l'expression de plusieurs milliers de gènes dans un organisme donné.

Dans cette thèse, deux approches de puces à ADN dédiées à la caractérisation de la flore bactérienne sont décrites. La première approche, dénommée puce phylogénétique, a permis de développer un ensemble de sondes reconnaissant des séquences signatures spécifiques de chacun des nœuds de l'arbre phylogénétique bactérien. Cette stratégie permet d'étendre le champ d'action des puces à ADN à l'ensemble du règne bactérien, tout en permettant de détecter des microorganismes connus, mais aussi inconnus, dans des échantillons biologiques. De plus, comparées au clonage-séquençage, les puces à ADN phylogénétiques permettent d'observer une diversité bactérienne plus importante dans les échantillons analysés. Une seconde stratégie a permis de développer une puce à ADN de basse-densité ciblant le gène de l'ARNr 16S, développée spécifiquement pour étudier les espèces bactériennes retrouvées dans la flore gingivale d'enfants atteint de noma, une maladie dévastatrice dont l'étiologie est inconnue.

Dans à le but d'identifier le ou les agent(s) causatif(s) du noma, ces deux méthodologies ont été utilisées pour caractériser la flore gingivale des enfants souffrant de cette maladie. Les résultats obtenus permettent d'écarter *Fusobacterium necrophorum*, l'organisme longtemps suspecté d'être l'agent causatif du noma. De plus, dans les lésions gingivales des patients noma nombre de pathogènes oraux connus, apparaissent plus abondants, notamment *Atopobium* spp., *Peptostreptococcus* spp., *Prevotella intermedia*, *Streptococcus pyogenes* et *S. anginosus*. En outre, la diversité bactérienne dans les lésions noma est moindre que celle rencontrée dans la gingivite nécrosante aigüe (GNA), soutenant l'hypothèse déjà formulée que la GNA est précurseur du noma.

Tout en donnant un meilleur aperçu de la bactériologie du noma, le travail accompli durant cette thèse démontre que les deux stratégies de puce à ADN développées permettent d'étudier et de caractériser méthodologiquement des communautés microbiennes de composition complexe.

Summary

During the past decade, the advent of new molecular techniques has led to enormous progress in biology, notably with the development of DNA microarray technology. This technology allows monitoring simultaneously the expression of thousands of genes from a given organism. DNA microarrays have been used in a variety of applications, including the characterization of bacteria in biological samples.

In this thesis, two distinct DNA microarray approaches for the characterization of bacterial flora are introduced. The first approach, termed phylogenetic microarrays, consists in a probe set recognizing specific sequence signatures for each node of the bacterial phylogenetic tree. This strategy, based on sequence information, allows extending the scope of microarrays to the whole bacterial kingdom and detecting both known and unknown microorganisms in biological samples. Moreover, phylogenetic microarrays permit detection of a broader bacterial diversity compared to the classical cloning-sequencing approach. The second strategy consists in a low-density 16S rRNA gene microarray specifically designed to monitor bacterial species found in the gingival flora of African children suffering from noma, a devastating disease of unknown etiology.

In an attempt to identify the causative agent(s) of noma, these two methodologies were applied for the characterization of the gingival flora of children suffering from the disease. Observations made during this study allowed exonerating *Fusobacterium necrophorum*, considered by some experts as the causative agent of noma. Moreover, various oral pathogens were recovered in higher abundance in noma lesions, notably *Atopobium* spp., *Peptostreptococcus* spp., *Prevotella intermedia*, S*treptococcus pyogenes* and *Streptococcus anginosus*. In addition, noma lesions exhibited a lower bacterial diversity compared to acute necrotizing gingivitis (ANG), thus supporting a previous hypothesis that ANG might precede acute noma.

The accomplished work, while giving better insights on the bacteriology of noma, demonstrates the power of the two developed approaches to explore and systematically characterize complex microbial communities.

Table of Content

Abbreviations

ANG: Acute necrotizing gingivitis

cDNA: Complementary deoxyribonucleic acid

CGH: Comparative genomic hybridization

DNA: Deoxyribonucleic acid

FISH: Fluorescence *in situ* hybridization

IVT: *In vitro* transcription

MLST: Multi-locus sequence typing

MM: Mismatch

mRNA: Messenger ribonucleic acid

ORF: Open Reading Frame

PCR: Polymerase chain reaction

PM: Perfect match

RDP: Ribosomal Database Project (http://rdp.cme.msu.edu/)

RNA: Ribonucleic acid

rRNA: Ribosomal ribonucleic acid

SNP: Single nucleotide polymorphism

VNTR: Variable number of tandem repeat

Tm: Temperature of melting

Figure Table

Chapter 1

Introduction

1.1 Introduction to microarray technology

Modern diagnostics of bacterial etiologies (etiology: the cause of a disease) involved in human diseases rely on conventional gene identification or expression assays which, despite accurate, appear limited by the number of target genes that can be treated simultaneously. DNA microarrays, or 'DNA chips', overcome this limitation by providing the possibility to identify or quantify the expression of thousands of genes in a single experiment. Microarrays take advantage of the hybridization phenomenon occurring between two complementary single nucleotide strands: each sample sequence (target) hybridizes with its complementary strand on the array (probe) which reveals the presence of the gene in the studied sample.

DNA microarrays consist of multiple nucleic acid probes immobilized on a surface (generally glass slides) allowing specific interaction with their cognate target sequence. To assess the presence of a gene, sample sequences are generally labeled with fluorescent dyes, either chemically or during enzymatic reactions. After hybridization (i.e. contact of the labeled mixture with the array surface under defined duration with proper chemical and temperature conditions), the mixture is washed off to leave only strong (i.e. theoretically specific) hybridizations on the surface of the array. The resulting fluorescence is measured by a confocal microarray scanner which produces high definition images. The fluorescence content of each feature (i.e. location of a specific probe set) is then extracted and quantified using dedicated image analysis software.

Microarray technology relies heavily on bio-informatics. There is a great need for specialized software methods for i) the optimal design of microarrays, ii) the normalization of array datasets, iii) the analysis of array data, iv) the data mining or interpretation, and v) the handling of the huge databases resulting from analysis.

Historically, the first published article (Schena et al., 1995) describing the use of microarrays was published in 1995. In this article, authors spotted 48 *Arabidopsis thaliana* cDNAs (complementary DNA) on a microscope glass slide and performed the first differential expression measurements using microarrays. Since then, the technology rapidly evolved on several levels (number of features per array, scanner resolution, surface chemistry, samples preparation, etc.) and has been used in a large variety of applications such as comparative genome hybridization (CGH) (Koessler et al., 2006)[1], transcriptomics (Garzoni et al., 2007), single nucleotide polymorphism (SNP) analysis (Fan et al., 2000), resequencing (Leski et al., 2009) or detection of microbial pathogens (Palacios et al., 2007). The field of application depends of the underlying strategy and marker genes used during the probe design.

1.2 Technical aspects of microarray technology

1.2.1 Nature of the probes

Basically, the nature of the probes can be distinguished in 3 groups: PCR products, cDNA probes and oligonucleotides, described below.

1.2.1.1 PCR products

This method was initially developed before the availability of complete genome sequences. Only partial knowledge of sequences is therefore needed, as only two short primers are mandatory to amplify PCR products. Following purification and quantification, the PCR product will be deposited onto the glass surface. This method provides sensitive microarray detection, as the products are generally 0.5-2 kb long but shows poor capacity to find punctual differences due to the large size of hybridization capture elements. This strategy performed on

[1] Author's contributions are underlined in the cited references included in the text of this manuscript

sequenced organisms allows the production of high density arrays whereas non-sequenced organisms generally yield to (low) densities limited to known genes only.

1.2.1.2 cDNA probes

A cDNA library is a collection of cloned cDNA fragments. To produce such libraries bacterial mRNAs are isolated from the bacterial cell of interest; then, double-stranded cDNAs are created from mRNAs using the reverse transcriptase enzyme; the cDNA fragments are then inserted into bacterial plasmids. Consequently, the cloned cDNA fragments can be conveniently selected, commonly through the use of antibiotic, and sequenced.

cDNA clones from cDNA libraries were used to design the first expression arrays (Schena et al., 1995). After extraction of the cDNA inserts from the vectors, single-stranded cDNAs can be used for spotting. Amplification of single-stranded cDNAs is generally performed using 5'-modified primers allowing covalent coupling of single-stranded fragments onto the surface. After purification and quality control, products are immobilized on a surface (nylon membrane, glass slide, etc.) using a spotting-robot. However, this method produces very long probes whose physico-chemical properties are poorly controlled, yielding to probes with low specificity (see chapter 1.2.2). Moreover, several investigators (Halgren et al., 2001; Kothapalli et al., 2002; Taylor et al., 2001) observed that a large number of sequences from various cDNA libraries are incorrect, making sequence verification a prerequisite before spotting and thus greatly complicating microarray manufacturing.

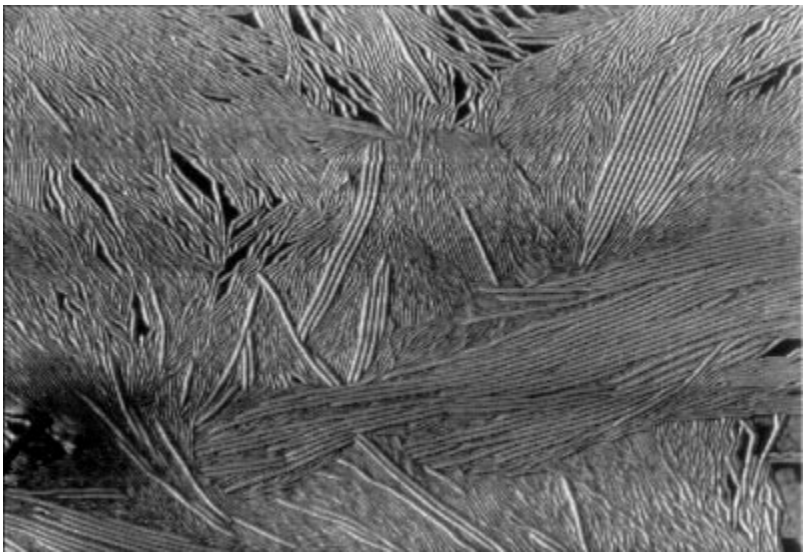


Fig. 1: Atomic force microscopy of a cDNA microarray. This is a micrograph of a portion of a hybridization probe from a yeast microarray, taken after the array was subjected to hybridization. The width of the picture represents a scanned distance of 2µm. Figure and legend adapted from (Duggan et al., 1999).

1.2.1.3 Oligonucleotide probes

Compared to the two aforementioned cDNA probes and PCR products, oligonucleotide probes are chemical synthesized, meaning that the physico-chemical properties are supervised (length, *Tm*, secondary structures, specificity; see chapter 1.2.2). Thus, although more expensive, the main advantage of this technology is the easiest handling of specificity (limited cross-hybridization with non-targets) and sensitivity (detection of low-abundant targets). Moreover, oligonucleotide probes give the opportunity to target a specific region of the target.

Oligo-probes can be either synthesized and then spotted on a surface, or directly synthesized *in situ* on the surface. Although the spotting approach is cost-appealing, *in situ* synthesis allows better chip-to-chip reproducibility and efficiency. For illustrative purposes, three commercial microarray platforms are described below.

A – Clondiag (http://www.clondiag.com/)

Clondiag provides a low-density microarray platform (335 probes per array) that uses oligonucleotide probes deposited on a glass surface using spotting robot. The latest product of this company is the ArrayStrip, consisting of a standard 8-well strip integrating a small microarray in the bottom of each well (see figure 2). The format is identical to a 96-well plates providing ease of multiple pipetting steps mandatory during hybridization, washing and staining of the microarray.

Fig. 2. **Clondiag ArrayStrip.** (Picture from www.clondiag.com)

Labeling of targets consists of the following steps: biotinylation of the target, hybridization of the labeled targets to the array; conjugation of horseradish peroxydase (HRP) – streptavidin to the bound target molecules; and finally, addition of the HRP-substrate tetramethylbenzidine (TMB). HRP initiates the conversion of TMB into a dark precipitate which can be detected by a dedicated optical scanner (see figure 3 below).

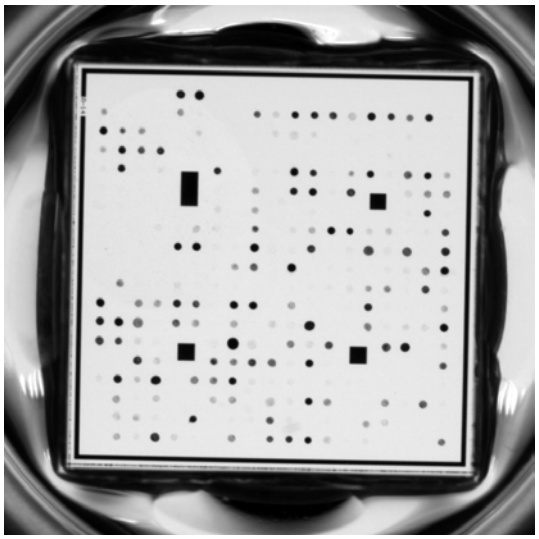


Fig. 3: Scanned Clondiag microarray. Each dot represents multiple copies of a determined probe whose feature intensity is relative to the target concentration in the hybridized sample. The geometrical features (squares and rectangle) on the array are used as positional marks to facilitate data acquisition.

Due to its very affordable cost per array (approx. 20 €) and its high-throughput capability, this platform is ideal for studies including a large number of samples, yet with relatively low number of features.

B – Affymetrix (http://www.affymetrix.com/index.affx)

With this technology, oligonucleotides are synthesized *in situ* on a silicon surface using photolithography, a technique used to built microprocessors. Basically, a mask is used to hide or unprotect feature positions from UV-light. The chip is flooded with a given nucleotide that couples to the unprotected positions. This step is repeated several times with different masks and with the 4 nucleotides in order to obtain short probes (25-mer typically) distributed all over the array (see figure 4 below).

Another unique feature is the use of paired probes; each perfect-match (PM) probe is paired to a mismatch (MM) probe differing by a complementary base located in the 13th position to quantify and deflate effects of cross-hybridization.

The advantage of this approach is the very high probe density, which can reach over 5 million probes on a surface of 1.6 cm^2. The main concerns are the high price and low flexibility of the platform, mainly explained by the use of several dedicated masks per array (~70).

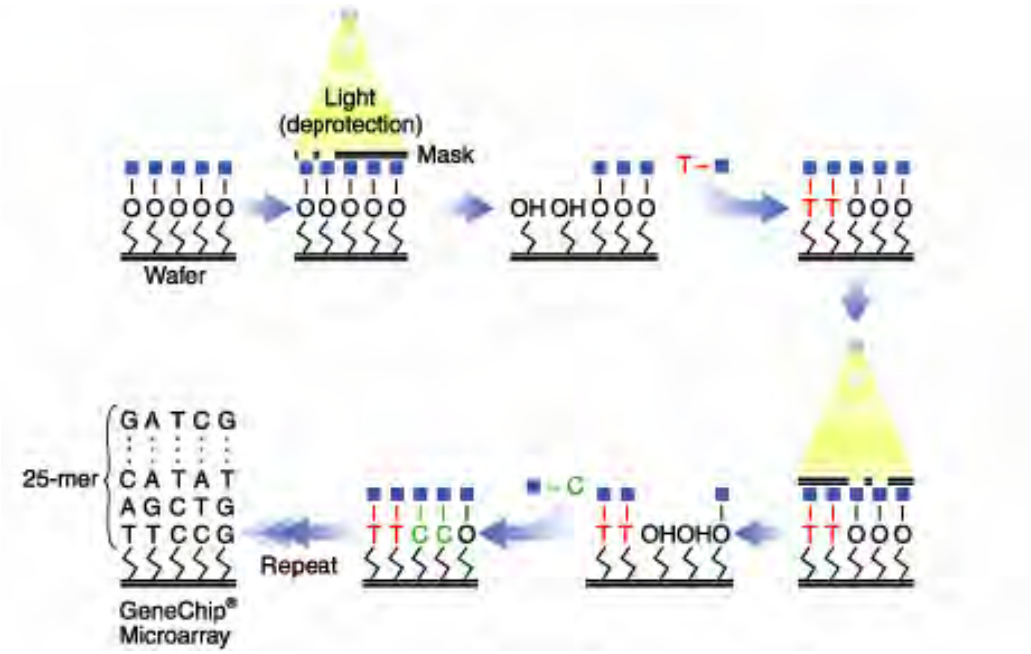


Fig. 4: Schematic mechanism of the photolithographic synthesis of oligonucleotide probes used by Affymetrix. Light is shine onto a chip through the holes of a mask, removing a protective group at specific positions. Subsequently, the chip is flooded with a given nucleotide that couples to the unprotected positions. These steps are repeated until 25-mer nucleotides are synthesized. Illustration from the German cancer research institute website (www.dkfz.de).

C – Agilent (http://www.home.agilent.com/agilent/home.jspx)

Agilent uses the inkjet printing technology (SurePrint) developed by Hewlett Packard to synthesize probes *in situ*. Without going too much into details, the glass slide is silane-coated with the hydroxyl groups available to initiate the oligo construction at a precise position (Coleman and Tsongalis, 2005). The printer head scans the whole surface while delivering a small volume of single nucleotide species; specific chemistry (phosphoramidite chemistry) is

used to elongate the oligonucleotide thus allowing obtaining probes with the desired sequence (see figures 5 and 6). On this platform, probes are longer (60-mer) than on the Affymetrix platform (25-mer), and no probe set methodology (paired probes) is used. As for the Affymetrix platform, fluorescent dyes are used to label the targets prior to hybridization. Labeling of target molecules is made with two fluorescent cyanine dyes: Cy-3 and Cy-5. Typically, when one wants to compare the expression levels between a wild-type and a mutant, the targets of the two cell lines will be labeled with either one of the dyes. During hybridization of the two target species on the microarray, a competition to form a stable probe-target duplex between the two species will occur. If only one cell line expresses a given gene, the probe will emit the corresponding fluorescence (red for Cy-5; green for Cy-3); however, if both cell lines express the same gene the resulting fluorescence will be a combination of the two wavelengths (yellow).

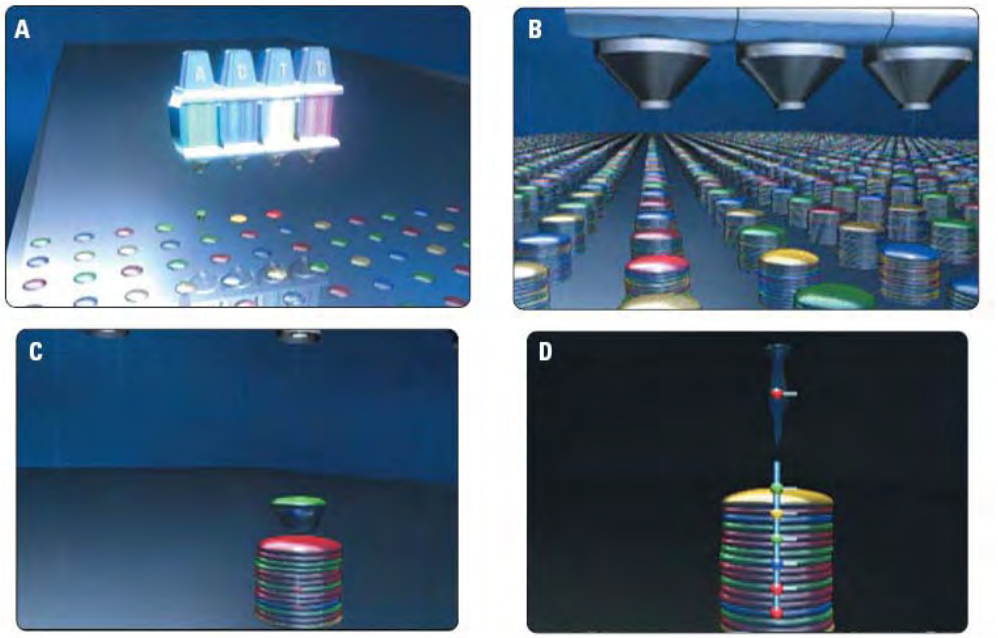


Fig. 5: General mechanism for oligo synthesis via inkjet printing. A: the first layer of nucleotides is deposited on the activated microarray surface. B: growth of the oligos is shown after multiple layers of nucleotides have been precisely printed. C: close-up of one oligo as a new base is being added to the chain, which is shown in figure D. Illustration and legend from Agilent website.

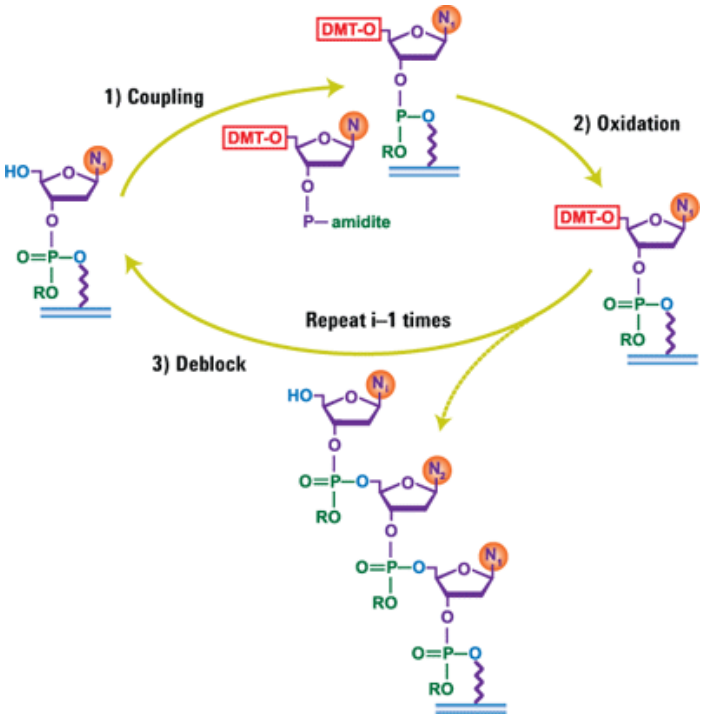


Fig.6: The general cycle of oligo synthesis via phosphoramidite chemistry. The process is repeated 60 times. Illustration and legend from Agilent website.

Compared to Affymetrix chips, where design of new chips requires manufacturing a new series of expensive masks, this approach makes design of custom chips more flexible and more affordable. Moreover, the Agilent platform allows co-hybridization of two samples (labeled in Cy-3 and Cy-5), while the Affymetrix only allows hybridization of one sample per array.

However, the probe density achieved with the inkjet technology is rather limited compared to Affymetrix. On a single microscope slide surface, up to 1 million probes can be synthesized for the Agilent platform, compared to more than 5 million (5’362’207) probes for the Affymetrix platform.

1.2.2 Oligonucleotide probe design

The selection of adequate probes is the cornerstone of any microarray design. Basically, two points must be considered:

- *Specificity*: Probes must permit hybridization of the target while avoiding cross-hybridization with non-target sequences.
- *Sensitivity*: Probes must be able to detect low-abundant target sequences in the sample, including in complex nucleic acids mixtures.

To address these expectations, probes must exhibit peculiar physico-chemical properties and display limited similarity to all non-target sequences.

1.2.2.1 Probe length

Oligonucleotide length plays a critical role in the specificity/sensitivity of the probe. Short probes (< 35 nucleotides) tend to be more specific but tend also to be less sensitive than longer nucleotide probes (60-80 nucleotides) (Chou et al., 2004; Hughes et al., 2001; Relogio et al., 2002; Suzuki et al., 2007). In a study addressing the optimization of oligonucleotide probes for DNA microarray, Relogio *et al.* (Relogio et al., 2002) reported that 60-nt probes are on average 7-fold more sensitive than 25-mer, but 4-fold less specific. However, spacers on short oligonucleotide probes can be used to obtain a good balance between specificity and sensitivity. Spacers, such as thymidine tails, increase the length of short probes thus reducing steric hindrance with the microarray surface (Peplies et al., 2003) and increasing probe sensitivity. The choice of the appropriate probe length depends of the application of the microarray. For example, short probes are more adequate for SNP analysis because it enables detection of a single nucleotide variation (Bischoff et al., 2008).

1.2.2.2 Probe thermodynamics

From a structural point of view, probe-target hybridization is influenced by the formation of inter- and/or intra-molecular secondary structures (stem loops, helices, dimerization, etc.) and by the affinity of the probe to bind to a target; which follows thermodynamic predictions. During hybridization, the binding affinity of the target to the probe, covalently fixed on a surface, is characterized by the difference in the binding free energy (ΔG) and by the melting temperatures (*Tm*). To allow probe-target duplex formation and stability, probes exhibiting low values in ΔG (Fish et al., 2007) and a homogenous range of *Tm* will be favored.

A study by He *et al.* (He et al., 2005), showed that a binding free energy threshold < -30 Kcal/mol could be set in order to select efficient 50-mer probes (< -40 Kcal/mol for 70-mer). Moreover, probes that can form intra-molecular secondary structures that could therefore affect the formation of probe-target duplexes must be excluded.

The most accurate model to predict thermodynamic properties is certainly the nearest-neighbor model (SantaLucia, 1998). In this model the ΔG is calculated as a sum of 'stacking' parameters associated to di-nucleotides. However, this model was described with experimental observations performed in solution and may not reflect the actual properties of probes attached on a solid surface (Li and Stormo, 2001). This was confirmed by Pozhitkov and colleagues (Pozhitkov et al., 2006) who observed a poor statistical correlation between predicted free energies and signal intensities of oligoarray probes.

1.2.2.3 Probe specificity

Probe specificity plays a crucial role in microarray experiments: an ideal probe set must maximize identity between probes and targets, while limiting identity of probes with non-targets (cross-hybridization) in another part of the considered organism genome. Series of parameters must be taken into account in order to dispose of a successful microarray design.

Using synthetic transcripts hybridized on an array composed of 60-mer oligonucleotide probes with mismatches and deletions, Hughes *et al.* (Hughes et al., 2001) characterized probes hybridization efficiency. This study suggested that oligonucleotides with more than 75% of identity with non-targets may produce cross-hybridization.

In addition, Kane et al. (Kane et al., 2000) reported that for a 50-mer probe with a 15-base, 20-base, 30-base or 35-base stretch identical with non-targets, non-targets contribution to the signal corresponds to approximately 1%, 15%, 30% and 50%, respectively. Consequently, for a 50-mer probe a maximum of 14 continuous perfectly complementary base pairs is recommended to avoid cross-hybridization with non-specific target. Of course, the stretch length should be extended for longer probes (up to 20 bases for 70-mer (He et al., 2005)).

However, considering the overall identity of probes with non-targets to prevent cross-hybridizations is not sufficient. Indeed, mismatch position (i.e. the position of a base in a strand that is not complementary to the base of the other strand) plays also a crucial role in probe specificity and probe-target duplex stability. In a simplified approach, the importance of a mismatch is proportional to its distance from the microarray surface. On classical oligoarray platforms, such as Agilent, probes are synthesized directly on the glass support with the 3' end anchored to the array surface. Hughes *et al.* (Hughes et al., 2001), showed that 5 or more mismatches that are randomly distributed along a 60-mer probe reduced the signal to <50%, and that the same effect was measured with a single mismatch positioned at the 5' end of the probe (solution end). On figure 4 (below), we clearly see that a mismatch near the 5' end will have more effect on probe signal (i.e. probe specificity) than a mismatch on the surface end of the probe.

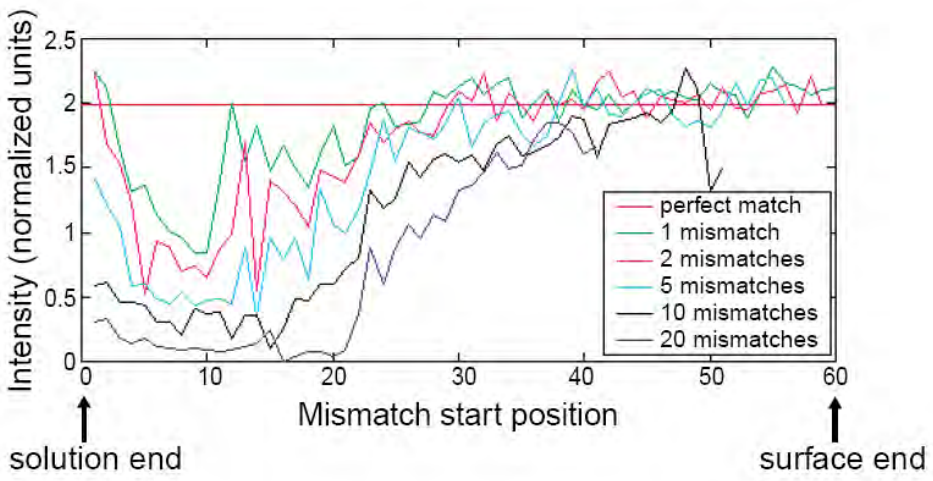


Fig. 7: Dependence of hybridization intensity on location of mismatches. Mean hybridization intensities as a function of the starting position for the indicated number of contiguous nucleotide mismatches in the 60-mer sequence. Figure and legend from Hughes *et al.* (Hughes et al., 2001).

1.2.3 Design strategies and markers used in microbiology

The field of application of a microarray depends primarily on the strategy and the marker genes used to design the probes. The chosen strategy is thus tributary of the context of the study. Below is a non-exhaustive review of the possible strategies.

1.2.3.1 Whole-genome ORF arrays

Whole-genome ORF arrays contain probes that will ideally match each ORF of one or several genomes. I have been personally involved in the conception of various whole-genome ORF arrays designed to study various strains of *Staphylococcus aureus* (Garzoni et al., 2007), *Rickettsia conorii* (Renesto et al., 2008), *E. coli* (Grasselli et al., 2008), Staphylococcus *epidermidis* (unpublished) or *Neisseria meningitidis* (unpublished). Typically, with this kind of approach, we could globally cover around 92% to 96% of the ORF's in each genome. Full coverage of the genome can be hard to reach because:

- The candidate probes sequence can be repeated in the genome (e.g. in case of paralogous genes, i.e. homologous gene sequences from the same genome that diverged by gene duplication),
- Or because, no probe sequence can be found in a (short) sequence due to the specified physico-chemical properties.

This approach is suitable to monitor gene expression (mRNA, transcriptomics studies) of one cell line under several growth conditions or during time-series experiments (Charbonnier et al., 2005; Garzoni et al., 2007; Renesto et al., 2008). Alternatively, the same arrays can be exploited to compare the genomic content (DNA) of two related strains/species in a comparative genomic hybridization (CGH) study. In this case, the DNA of the two studied strains/species is labeled with two different fluorescent dyes and co-hybridized on the same array (Koessler et al., 2006). Comparison of the log-ratio values for all the fluorescent signal intensities makes possible the direct comparison of the genomic content of the two studied strains/species. The CGH approach is valuable to identify core genes shared by several strains

and strain-specific accessory genes which result from microevolution events, such as lateral transfer gene acquisitions, deletions and rearrangements (Grasselli et al., 2008; Leonard et al., 2003; Lindsay et al., 2006). In 2002, a study (Tettelin et al., 2002) used CGH to measure the relatedness of *Streptococcus agalactiae with the related human pathogens Streptococcus pneumonia and Streptococcus pyogenes. In this study, the authors showed that 50% of the genes had homolog(s) in the three genomes*, and that gene synteny (i.e. both sequence and gene order are conserved between closely related species) was more pronounced between *S. agalactiae* and *S. pyogenes* than between *S. agalactiae* and *S. pneumonia. This study also identified* genes unique to *S. agalactiae* that are expected to play a role in colonization or disease; many of these being associated with mobile genetic elements (bacteriophages, transposons and mobile elements). Similarly, a CGH study focusing on the genomic characterization of community- and hospital-acquired methicillin-resistant *S. aureus* (Koessler et al., 2006) showed that this approach is well suited for discovering novel marker genes and chromosomal regions in a collection of clinical isolates. CGH is an ideal genomotyping tool (genomotyping: characterization of the genomic DNA of unsequenced bacterial strains by comparison to a sequenced reference genome) with a better resolution than more typical molecular techniques such as characterizing variable numbers of tandem repeats (VNTR, this technique measures in a genome the number of repetitions in a tandem repeat. This approach can be conveniently used for determining the clonality of bacterial strains, for example in epidemiologic studies (Francois et al., 2005; Francois et al., 2008; Harbarth et al., 2005; Mégevand et al., 2009)) or multilocus sequence typing (MLST, this technique consists in sequencing multiple housekeeping genes – usually seven – of a bacterial species. Each sequence is compared to a database in order to determine the allelic profile, allowing assessment of the evolutionary relatedness between strains (Islam et al., 2009; Köck et al., 2009; Neumann and Rehberger, 2009; Pitondo-Silva et al., 2009; Varshney et al., 2009)). Note that these two methods rely on the characterization of a limited number of targets compared to the CGH.

1.2.3.2 Resequencing microarrays

This approach consists in sequencing a DNA region of an organism that shares partial identity with a related sequenced genome. Practically, a series of overlapping (tiled) probes

are designed all along a reference sequence. For each probe, four variant probes are included that vary in the middle probe position by a different nucleotide. Comparison of the probe signals allows determining which probe anneals the best and reveals the sequence of the studied organism.

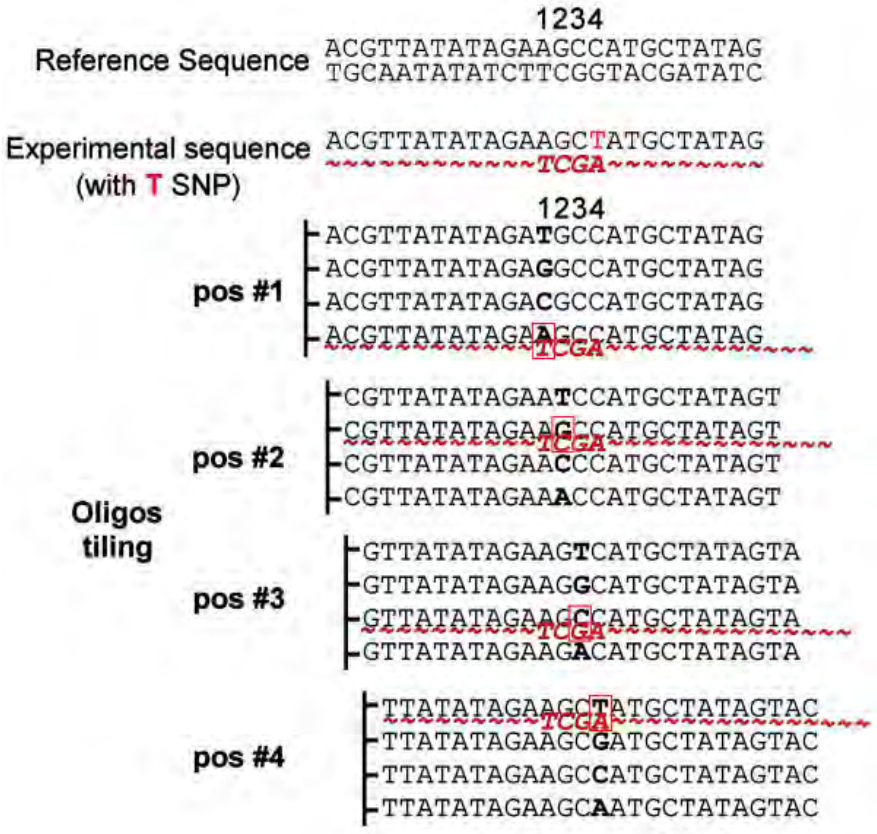


Fig. 8: Schematic representation of the design of a resequencing microarray

(Adapted from http://www.mun.ca/biology/scarr/DNA_Re-sequencing_with_a_Microarray.html)

In a study by Zwick *et al.* (Zwick et al., 2004), a panel of 56 *Bacillus anthracis* strains was resequenced using this approach. This study highlighted 37 SNPs demonstrating a rather low level of DNA sequence variation between the studied strains. In another study, Wang *et al.* (Wang et al., 2006) used this approach to identify influenza viruses involved in respiratory infections. This team developed a respiratory pathogen microarray including probes targeting partial sequences from haemagglutinin (HA), neuraminidase (NA) and matrix (M) genes of influenza A and B viruses. Using a random amplification protocol, this study demonstrated that the sequences obtained with the array were identical to conventional sequencing except for ambiguous base calls, and enabled comparison of the sequences of the isolated virus to

previously sequenced reference viruses. Resequencing microarrays not only permit the characterization of respiratory viruses at the species level (using pattern recognition algorithms for instance), but also at the strain level, which makes this a powerful tool for epidemiologic surveillance (Wang et al., 2006).

Resequencing microarrays provide a fast and affordable alternative to shotgun sequencing approaches (shotgun sequencing consists in shearing the genomic DNA in random fragments before sequencing. Obtaining short genomic fragments allows bypassing methodological constraints inherent to sequencing. After sequencing, assembly software are used to reconstruct the original sequence from the resulting reads). However, implementation of resequencing microarrays can be complicated. Firstly, because selection of a series of overlapping probes respecting specificity criteria can be difficult to obtain. Secondly, early publications demonstrated a high rate of false positive calls: studies on the human genome reported that between 12% and 45% of the detected SNP variants were false (Cargill et al., 1999; Halushka et al., 1999; Wang et al., 1998). This last point emphasizes the need to implement dedicated analysis software, such as ABACUS (Cutler et al., 2001), to reduce the SNP false-positive rates.

1.2.3.3 Functional gene arrays

Functional gene arrays (FGAs) are DNA microarrays containing probes that target a conserved region within a family of genes of interest, displaying specific functions. FGAs can be used to ensure a higher phylogenetic granularity (phylogeny: evolutionary relatedness among different species) when the degree of similarity between species/strains is too narrow to perform reliable identification using only ribosomal regions (see section 1.2.3.4, below). Secondly, this approach appears particularly useful for assessing the presence and activity of key enzymes, antibiotics resistance genes or virulence factors from microbial samples, which may be more important, in some instances, than identifying species using other phylogenetic criteria.

Primarily used to study the metabolic potential of environmental bacterial samples (e.g. biodegradation microbial communities (Rhee et al., 2004), biogeochemical processes (He et al., 2007) or nitrogen cycling (Tiquia et al., 2004)), this approach successfully detected genes involved in virulence and antibiotic resistance mechanisms (Jaing et al., 2008; Sergeev et al.,

2006; Volokhov et al., 2002). For instance, Sergeev *et al.* (Sergeev et al., 2006) implemented a FGA testing for the presence of 18 virulence factor genes in the subtype *Bacillus* group and they showed that this strategy could indeed provide strain typing information in addition to the inventory of virulence factors. A similar strategy was also described for different bacteria of clinical importance (Perreten et al., 2005; Saunders et al., 2004). The description of a high density arrays targeting a comprehensive collection of genes involved in virulence and antibiotics resistance mechanisms was recently reported (Jaing et al., 2008). For the first generation array, this team identified 1'245 target sequences representing 160 virulence and antibiotics resistance gene families derived from two *Escherichia coli* strains (CFT073 and K12), *E. faecalis* and *S. aureus*. To validate the approach and determine the detection limit of the array, authors spiked known amounts of bacterial DNA in aerosols samples. Two important observations emerged from this study. Firstly, when using proper whole-genome amplification, this array was able to detect low quantities (3.1 fg) of purified DNA spiked in aerosol samples. Secondly, this FGA proved successful in detecting orthologs (orthologs: gene sequences sharing homology between two or more organisms) of different species in the same genus. Thus, while providing strain-level identification, this array could also be used as a tool to describe virulence factors in environmental samples containing yet un-sequenced organisms.

1.2.3.4 rRNA gene arrays

Evolution can be considered as a branching process, where populations are altered over time and may speciate into separate branches. Phylogenetics is the study of the evolutionary relatedness between different organisms, bacteria in our case, based on protein or nucleic acid sequences. The underlying assumption is that lateral gene transfer is the dominant force that shapes prokaryotic genomes. The ribosomal rRNA gene is the marker the most commonly used for inferring the bacterial phylogeny (see figure 9). Modern molecular diagnosis tests established the ribosomal RNA, and more precisely the 16S rRNA and 23S rRNA genes as a gold-standard for bacterial characterization. Several characteristics make this structural RNA suitable for bacterial identification. Firstly, its structural role as a central component of the ribosome is essential to all bacteria and makes this gene ubiquitous. Secondly, the combination of conserved and divergent sequence regions and extremely rare lateral transfer

(Claridge et al., 1997) makes this gene usable as a “molecular clock” suitable for phylogenetic identification (Claridge et al., 1997; Woese, 1987). Finally, this gene is highly abundant in bacterial cells: it is found in multi-copy and is highly transcribed. A multitude of techniques take advantage of these properties such as denaturing gradient gel electrophoresis (DGGE), fluorescence *in situ* hybridization (FISH), single strand conformation polymorphism (SSCP), terminal restriction fragment length polymorphism (T-RLFP) and, of course, sequencing.

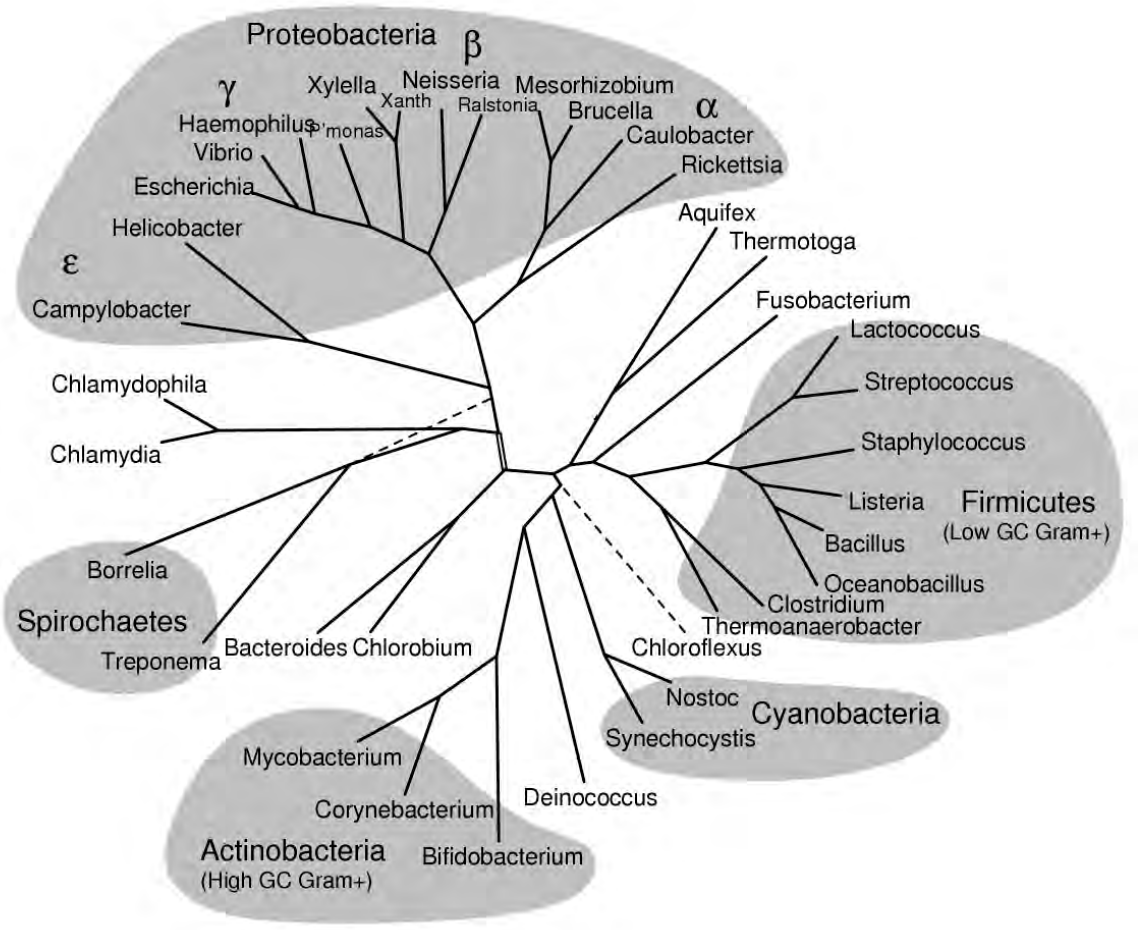


Fig. 9: Simplified phylogenetic dendrogram based on the 16S rRNA gene.

For microbial identification, the 16S rRNA gene is undoubtedly the most popular rRNA as it benefits from a large number of dedicated sequence databases such as Greengenes (DeSantis et al., 2006), RDP (Maidak et al., 2001) or RIDOM (Harmsen et al., 2002); and

mainly explains the popularity of 16S rRNA gene microarrays. This kind of approach is used for single bacteria identification in a clinical sample, and can also be conveniently used in metagenomic studies (metagenomic: the comprehensive range of all DNA sequences of a bacterial consortium in a particular ecological niche). Note that 16S rRNA microarrays are always oriented towards specific bacteria or microbiota (microbiota: community of microorganisms inhabiting a particular ecological niche).

A 16S rRNA microarray, consisting of long and short oligonucleotide probes, was developed to monitor human gastrointestinal tract microbiota. This microarray not only proved to be able to rapidly detect uncultivable bacteria in the fecal flora, but also displayed an excellent analytical sensitivity determined to be $8.8 \cdot 10^4$ bacterial cells/g of fecal sample. This claimed sensitivity is extremely low compared to the total gut bacterial population of 10^{12} to 10^{13} cells/g and exceeds the detection limit of the fluorescence *in situ* hybridization (FISH) technique (10^6 cells (Zoetendal et al., 2004)).

The Human Oral Microbe Identification Microarray (Colombo et al., 2009), developed at the Forsyth institute, is another good example of microbiota-specific 16S rRNA-based microarray. This microarray consists of 400 oligonucleotide probes targeting more than about 300 bacterial species encountered in the human oral flora. Using this platform, Colombo *et al.* (Colombo et al., 2009) analyzed subgingival plaque samples taken from 67 subjects with periodontitis and 20 individuals with a healthy periodontal status. In this study, this group concluded that patients with periodontitis exhibited a broader bacterial diversity compared to individuals with a healthy periodontal status; moreover, this platform was able to characterize subtypes of gingivitis on the basis of their respective bacterial profiles.

Another study by François *et al.* (Francois et al., 2006) used classification algorithms on a Ta_2O_5-coated microarray coated with 19-mer signature oligonucleotides targeting the 5'-end of 16S rRNA genes of human pathogenic bacteria. In an approach unique to this study, the authors hybridized RNA directly, chemically coupled to fluorescent dyes in order to avoid biases related to enzymatic labeling methods (reverse transcription or *in vitro* transcription) enzymatic reverse-transcription. In this study, classification of hybridization patterns was performed using decision trees and artificial neural networks yielding to >96% confidence in bacterial identification. Interestingly, the authors of both studies did not record any cross-hybridization due to the presence of human nucleic acids but highlighted the need for sophisticated analytical methods instead of simply monitoring signal levels of cognate probes.

The main limitation of the 16S rRNA gene relies in its high degree of conservation in several regions, even between distant organisms. In some cases, it can be impossible to design

species-specific probes. A relevant example exists in the *Enterobacteriaceae* family. To overcome this problem, alternative regions within the ribosomal operon have been proposed, such as the 5S and 23S (Hong et al., 2004) as well as the rDNA internal transcribed spacer (ITS) region (Leaw et al., 2007; Nübel et al., 2004). Various properties make the ITS region especially suitable for the identification of fungi: (i) this region is ubiquitous to all fungi, (ii) it is present in multiple copies, and (iii) the ITS region offers a higher level of variability as compared to rRNA genes, thus making this region more suitable for identification at the species level. In order to characterize fungal pathogens, a large variety of microarrays were developed on the basis of the ITS region with good reproducibility and specificity (Hsiao et al., 2005; Huang et al., 2006; Leinberger et al., 2005).

1.2.3.5 Phylogenetic arrays

Characterizing the content of complex microbial samples is usually achieved by enumerating small rRNA subunit (16S rRNA) molecules after PCR amplification using "universal" primers followed by cloning and sequencing (Kassinen et al., 2007; Kuhbacher et al., 2006; Manichanh et al., 2005; Zhu et al., 2002). Still considered as the "gold standard method", this approach remains laborious, costly and the number of clones required to obtain a comprehensive inventory of the bacterial content of a sample can be tremendously large. Note also that the first amplification step introduces – by definition – a bias between the original sample composition and the final amplification products (Baker et al., 2003; Blackwood et al., 2005; Isenbarger et al., 2008), due to the utilization of "universal" primers.

Taking advantage of the high parallelism property of microarrays, phylogenetic microarrays have been proposed as a convenient alternative to study complex bacterial flora. The underlying concept is to design probes recognizing sequence signatures specific for each node of the phylogenetic tree, extending the scope of oligoarrays to the whole bacterial kingdom. Many phylogenetic markers can be used for the design, such as the following genes *groEL* (Wertz et al., 2003; Wong and Chow, 2002), *gyrA/gyrB* (Antwerpen et al., 2007), *rpoB* (Drancourt et al., 2004; Troesch et al., 1999) or *rpsA* (Martens et al., 2007), but once again the 16S rRNA gene remains the preferred candidate because of the number and prosperity of available databases dedicated to this marker (see preceding chapter). Several teams developed and validated this approach (Brodie et al., 2007; Huyghe et al., 2008; Huyghe et al., 2009;

Palmer et al., 2006; Palmer et al., 2007), proving its efficiency in monitoring the contents of complex bacterial mixtures in a quantitative (fractional abundance <0.1% (Palmer et al., 2006)) and qualitative fashion. Moreover, compared to the gold-standard cloning-sequencing, several contributors (DeSantis et al., 2007; Huyghe et al., 2008; Huyghe et al., 2009) observed a broader bacterial diversity in microbial samples using phylogenetic microarrays.

Applying a rRNA phylogenetic microarray to monitor the temporal and geographical dynamics of bacterial populations in urban aerosols, Brodie (Brodie et al., 2007) and colleagues detected a surprising variety of human and animal pathogenic bacteria in the air of two American cities (Austin and San Antonio). The causative agents of glanders and melioidosis, *Burkholderia mallei* and *B. pseudomallei* respectively, were regularly detected during the study; similarly, different pathogens such as *Campylobacteraceae*, *Clostridium botulinum*, *Helicobacter* and tick-borne *Rickettsia* were also detected in aerosol samples. This study showed that phylogenetic oligoarrays can be easily used to characterize bacterial pathogens from a dense background of non-pathogenic bacteria and could, for example, be exploited as part of a bio-surveillance initiative.

The same strategy was applied, along with cloning sequencing, to trace the development of the human infant gastrointestinal microbiota (Palmer et al., 2007). In this interesting study, fecal samples from 14 healthy full-term babies were profiled at defined intervals. In addition, parental samples (breast milk, stool and vaginal swabs) were also characterized in order to investigate possible origins of the infant gastrointestinal microbiota. The authors demonstrated that a small subset of phylotypes colonize the gastrointestinal tract: Actinobacteria, *Bacteroides*, Firmicutes, Proteobacteria and Verrucomicrobia. Also, they showed that in the early months of life the gastrointestinal microbiota differ greatly between individuals and are mainly composed of aerobes (e.g. *Staphylococcus*, *Streptococcus* and Enterobacteria). Progressively, these idiosyncratic microbial profiles become more similar to one another and converge toward an adult-like profile essentially composed of anaerobes (Eubacteria and Clostridia). Moreover, the authors of this study suggest that environmental and genetic factors play crucial role in the establishment of the gastrointestinal microbiota in human infants. From a technical point of view, this study also demonstrated an excellent concordance between the observations made with the phylogenetic array and the cloning-sequencing method, excepted for some cross-hybridization events observed with highly abundant taxa.

Infection, Genetics and Evolution 9 (2009) 987–995

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Discussion

Characterization of microbial pathogens by DNA microarrays

Antoine Huyghe [a,b,*], Patrice Francois [a], Jacques Schrenzel [a]

[a] Genomic Research Laboratory, Infectious Diseases Service, University of Geneva Hospitals, Micheli-du-Crest 24, 1211 Geneva 14, Geneva, Switzerland
[b] Swiss Institute of Bioinformatics, Michel-Servet 1, 1211 Geneva 4, Geneva, Switzerland

ARTICLE INFO

Article history:
Received 4 July 2008
Received in revised form 21 October 2008
Accepted 26 October 2008
Available online 17 November 2008

Keywords:
Microarray
Pathogen
Detection
CGH
Phylogeny
Bacteria

1. Introduction

Searching for the keyword 'microarray' in the NCBI database (June 2008 release) retrieved more than 25,000 hits, illustrating the popularity of this relatively recent technology that was first described in 1995 (Schena et al., 1995). Microarrays consist of multiple probes (nucleic acids, proteins, carbohydrates, antibodies) deposited or directly synthesized on a surface in an ordered fashion. This approach has several advantages over classical probe-based methods: high-throughput, parallelism, miniaturization and automation.

Conventional DNA microarrays consist of nucleic acid probes deposited on a planar glass surface. The surface is usually coated with chemically reactive groups (epoxy, poly-L-Lysine or aldehyde) to ensure efficient binding of nucleotidic probes on the surface. To assess the presence of target genes, nucleic acid samples are labeled, either chemically or by an enzymatic reaction. Labeled samples are then hybridized onto the array, and washed using different stringency buffers. The remaining signal resulting from specific interactions between probes and target nucleic acids is measured using a confocal microarray scanner. Only probes hybridized to a labeled target will yield signal, thus revealing the presence of the cognate nucleic acid motif in the sample.

High-density DNA microarrays have been used in a broad variety of applications such as transcriptomics, comparative genome hybridization (CGH), resequencing, drug discovery, microbial community characterization or single nucleotide polymorphism (SNP) analysis. The field of application depends primarily on the strategy and the marker genes used to design the probes. A variety of genes (virulence factors, phylogenetic markers, antibiotic resistance genes, etc.) have been employed on microbial diagnostic microarrays, depending on the question raised by the researcher.

Microarrays have thus been instrumental in the detection of known pathogens as well as in the discovery of novel infectious agents, such as SARS (Wang et al., 2003). Their performance has also been described in the detection of co-infecting agents during mixed infections (Lin et al., 2006, 2007). To the best of our knowledge, the utilization of microarrays is currently limited to research laboratories and no such technology is yet employed in routine clinical practice. In this review, we will inventory the proposed strategies for the detection of microbial pathogens.

2. Strategies and markers

2.1. Genome-oriented strategies

2.1.1. Whole-genome ORF arrays

A whole-genome ORF array contains probes recognizing the majority of the ORF content of one or several genomes. This approach is applicable to many different types of studies,

* Corresponding author.
E-mail address: antoine.huyghe@genomic.ch (A. Huyghe).

1567-1348/$ – see front matter
doi:10.1016/j.meegid.2008.10.016

explaining its popularity. Typically, this strategy is implemented for gene expression profiling studies (i.e. transcriptomics) to monitor the expression of mRNAs in the organism of interest under different growth conditions or during time-series analyses. Alternatively, whole-genome ORF arrays can be exploited for comparative studies, now relying on the hybridization of genomic DNA. Such comparative genome hybridization (CGH) studies can permit assessment of the genetic relatedness between a collection of organisms from the same genus and/or species, by evaluating at the genome scale the presence of various microevolution events (acquisition of genes by lateral transfer as well as deletions and rearrangements). Typically, the DNA of the reference genome is co-hybridized with the strain being characterized (Fig. 1A), in order to provide a positive control on each probe (Koessler et al., 2006). Comparison of log ratio values for all fluorescent signal intensities makes it possible to distinguish the presence/absence of genes in the studied strain (Kim et al., 2002). Whole-genome comparisons typically identify sets of core genes that are shared by a large majority of strains in the same species (Grasselli et al., 2008; Leonard et al., 2003; Lindsay et al., 2006), as well as accessory genes that are present in a subset of strains from the species that typically result from lateral gene acquisition. These differences can often be used to identify genes and/or genetic islands related to 'gain-of-function traits' in pathogenic strains.

A CGH study (Fitzgerald et al., 2001) that was carried out to investigate molecular population genetics on a panel of *Staphylococcus aureus* clinical isolates showed that about 22% of the genes are accessory. Most of the species-specific genes are located in chromosomal regions carrying genes involved in virulence or antibiotic resistance mechanisms. This study also underlined the importance of horizontal gene transfers in the evolution of *S. aureus* genome, notably in the acquisition and evolution of the

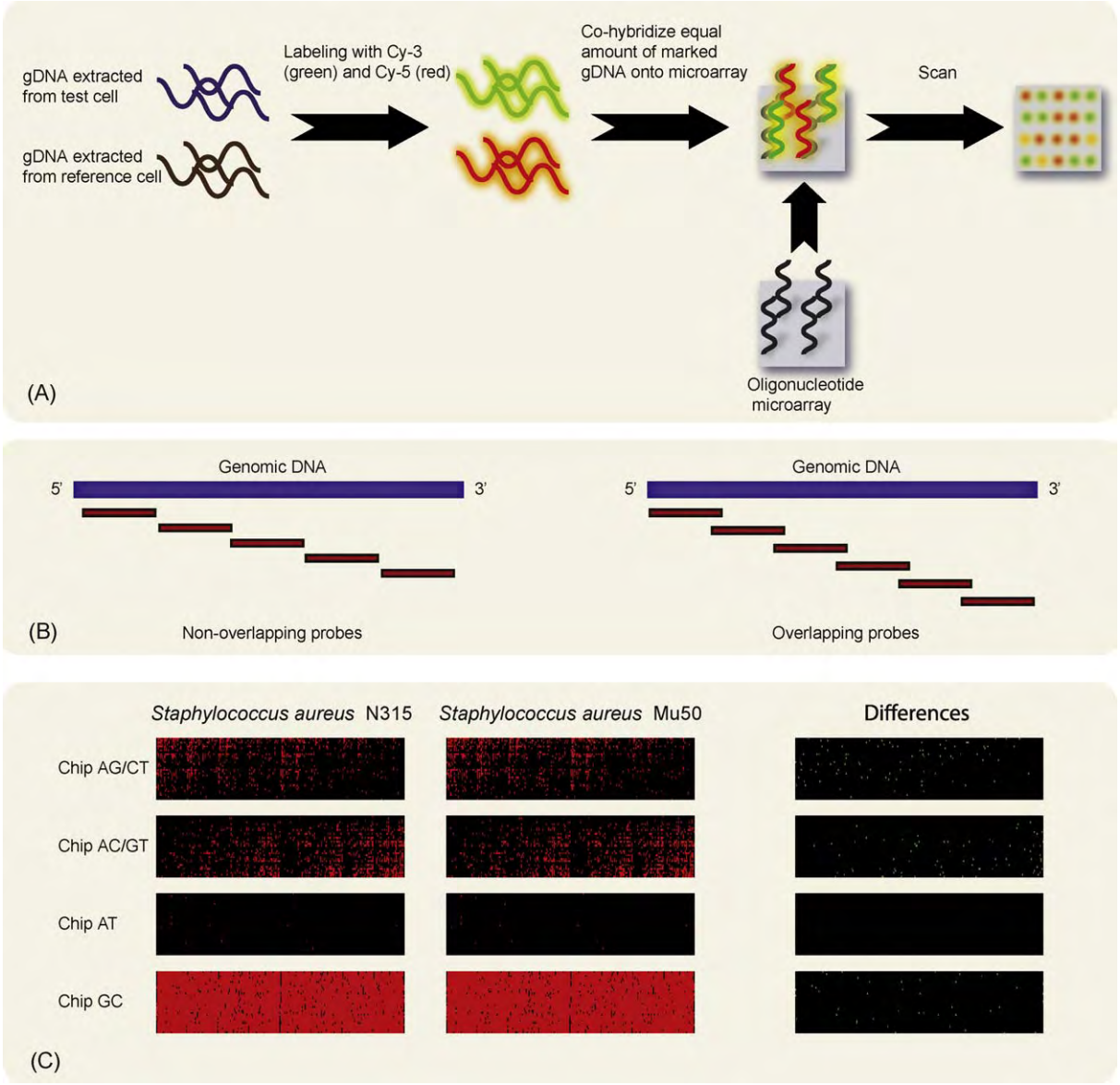


Fig. 1. (A) Principle of the comparative genomic hybridization (CGH). Equal amounts of genomic material appear in orange on the scanned array. (B) Probes disposition in a tiling array design. (C) Virtual hybridization of two methicillin-resistant *S. aureus* strains on NCHS chips and their differential hybridization patterns. This figure shows *in silico* hybridization (red) of the related *S. aureus* strains N315 and Mu50 (displaying decreased susceptibility to glycopeptides) on the four different types of NCHS chips (13-mers di-nucleotides). The differential pattern (green) represents probes that hybridize exclusively to either one of the two strains, yielding to a total of 216 informative probes for the four NCHS chips.

mecA gene responsible for methicillin resistance. Similarly, a CGH study focusing on the genomic characterization of community- and hospital-acquired methicillin-resistant *S. aureus* (Koessler et al., 2006) showed that this approach is well suited for discovering novel marker genes and chromosomal regions in a collection of clinical isolates. CGH is an ideal genomotyping tool with a better resolution than more typical molecular techniques such as characterizing variable numbers of tandem repeats (VNTR) or multilocus sequence typing (MLST).

2.1.2. Tiling arrays

With this strategy, probes are tiled along the entire genome length in an overlapping fashion or in close proximity (Fig. 1B). Although the number of probes required to cover a whole genome can rapidly become unwieldy, tiling arrays are conveniently used in resequencing and for single nucleotide polymorphisms (SNP) detection studies (Herring and Palsson, 2007; Kozal et al., 1996; Lin et al., 2006; Wong et al., 2004).

Use of resequencing microarrays allows generating sequences and determining SNPs to provide a way to track genetic variants, making this approach ideally suited for outbreaks surveillance. Typically, a resequencing microarray is composed of probes tiled on a sequence of interest with the central nucleotide permutated with all four possible combinations. The introduction of this central polymorphic site makes this approach ideal to define SNPs in a gene (Hu et al., 2008).

Wang et al. (2006) used this approach to identify influenza viruses involved in respiratory infections. This team developed a respiratory pathogen microarray including probes targeting partial sequences from hemagglutinin (HA), neuraminidase (NA) and matrix (M) genes of influenza A and B viruses. Using a random amplification protocol, this study demonstrated that the sequences obtained with the array were identical to conventional sequencing except for ambiguous base calls, and enabled comparison of the sequence of the isolated virus to previously sequenced reference viruses. Resequencing microarrays not only permit characterization of respiratory viruses at the species level (pattern recognition), but also at the strain level, which makes this a powerful tool for epidemiologic surveillance.

2.2. Metagenomic strategies

2.2.1. rRNA gene arrays

Conventional molecular diagnosis tests designed to identify bacterial strains usually rely on the amplification of the universal rRNA genes. These genes are particularly relevant targets for identification purposes for several reasons: (i) universality among bacteria, (ii) a combination of both conserved and divergent sequence regions, (iii) abundance in cells (frequently present as multi-copy genes and very highly transcribed, leading to high rRNA concentration), (iv) and the fact that lateral transfer of rRNA genes is extremely rare, explaining why they can be used as phylogenetic markers (Claridge et al., 1997). The resulting amplicons can be analyzed by various techniques such as denaturing gradient gel electrophoresis (DGGE), single strand conformation polymorphism (SSCP), terminal restriction fragment length polymorphism (T-RFLP) or sequencing. Unsurprisingly, a plethora of microarray designs take advantage of the intrinsic potential of this gene as a phylogenetic marker to characterize microbial pathogens.

Among rRNA molecules, the bacterial 16S rRNA is particularly popular and benefits from a large array of public sequence databases (ARB (Ludwig et al., 2004), Greengenes (DeSantis et al., 2006) or RDP-II (Maidak et al., 2001) among others), making this gene particularly appealing for routine bacterial identification. A variety of microarrays using probes targeting the 16S rRNA gene have been successfully validated for the identification of bacteria from clinical samples (Francois et al., 2006; Liu et al., 2005; Zhu et al., 2007) or environmental niches (Neufeld et al., 2006; Brodie et al., 2006; Loy et al., 2005). A diagnostic array based on 16S rRNA was developed to detect 23 bacterial bloodstream pathogens (Wiesinger-Mayr et al., 2007) using 74 short (20–30 bp) oligonucleotidic probes. This team showed that the approach was able to detect and identify closely related species in a range of 10^1–10^3 bacteria from pure cultures, and 10^1–10^5 bacteria per ml of spiked blood. The observed performance was markedly dependant on the bacterial species. Although all probes did not show species-specific signals, a machine learning algorithm (*k*-nearest neighbor) was used to classify hybridization patterns by similarity to hybridizations of known organisms. Similarly, Francois et al. (2006) used classification algorithms on a Ta_2O_5-coated microarray coated with 19-mer signature oligonucleotides targeting the 5′-end of 16S rRNA genes of human pathogenic bacteria. In an approach unique to this study, the authors hybridized labeled RNA directly coupled to fluorescent dye to avoid biases related to the amplification method. In this study, classification of hybridization patterns was performed using decision tree and artificial neural networks, most recently achieving a confidence >96% in bacterial identification. Interestingly, the authors of both studies did not record any cross-hybridization due to the presence of human nucleic acid but highlighted the need for sophisticated analytical methods instead of simply monitoring signal levels of cognate probes.

However, the main limitation of the 16S rRNA gene relies in its high degree of conservation. In some cases, it can be impossible to design species-specific probes. A notable example exists in the *Enterobacteriaceae* family. To overcome this problem, alternative regions within the ribosomal operon have been proposed, such as the 5S and 23S (Hong et al., 2004) as well as the rDNA internal transcribed spacer (ITS) region (Leaw et al., 2007; Nübel et al., 2004). Various properties make the ITS region especially suitable for the identification of fungi: (i) this region is ubiquitous to all fungi, (ii) it is present in multiple copies, and (iii) ITS regions offer a higher level of variability compared to rRNA genes, thus making this region suitable for identification at the species level. In order to characterize fungal pathogens, a large variety of microarrays were developed on the basis of the ITS region with good reproducibility and specificity (Hsiao et al., 2005; Leinberger et al., 2005; Huang et al., 2006).

2.2.2. Functional gene arrays

Functional gene arrays (FGAs) are DNA microarrays containing probes that target a conserved region within a family of genes of interest, displaying specific functions. FGAs can be used to ensure a higher phylogenetic granularity when the degree of similarity between species/strains is too narrow to perform reliable identification using only ribosomal regions. Secondly, this approach appears particularly useful for assessing the presence and activity of key enzymes, antibiotics resistance genes or virulence factors from microbial samples, which may be more important, in some instances, than identifying species with other phylogenetic criteria. Primarily used to study the metabolic potential of environmental bacterial samples (e.g. biodegradation microbial communities (Rhee et al., 2004), biogeochemical processes (He et al., 2007) or nitrogen cycling (Tiquia et al., 2004)), this approach successfully detected genes involved in virulence and antibiotic resistance mechanisms (Jaing et al., 2008; Sergeev et al., 2006; Volokhov et al., 2002). For instance, Sergeev et al. (2006) implemented a FGA testing for the presence of 18 virulence factor genes in the subtype *Bacillus* group and they showed that this strategy can indeed provide strain typing information in addition to the inventory of virulence factors. A similar strategy was also described for different bacteria of clinical

importance (Saunders et al., 2004; Perreten et al., 2005). The description of a high density arrays targeting a comprehensive collection of genes involved in virulence and antibiotics resistance mechanisms was recently reported (Jaing et al., 2008). For the first generation array, this team identified 1245 target sequences representing 160 virulence and antibiotics resistance gene families derived from 2 *Escherichia coli* strains (CFT073 and K12), *E. faecalis* and *S. aureus*. Two important observations emerge from this study. Firstly, when using proper whole-genome amplification, this array was able to detect low quantities (3.1 fg) of purified DNA spiked in aerosol samples. Secondly, this FGA proved successful in detecting orthologs of different species in the same genus. Thus, while providing strain-level identification, this array could also be used as a tool to describe virulence factors in environmental samples containing yet un-sequenced organisms.

2.2.3. Phylogenetic arrays

Characterizing the content of complex microbial samples is usually achieved by sequencing cloned PCR products, after amplification of the small rRNA subunit (16S rRNA) using "universal" primers. Still considered the 'gold standard method', this approach remains laborious, costly and the number of clones required to obtain a comprehensive inventory of the bacterial content of a sample can be tremendously large. Note also that the first amplification step introduces – by definition – a bias between the original sample composition and the final amplification products. Taking advantage of the high parallelism property of microarrays, phylogenetic microarrays have been proposed as a convenient alternative to study complex bacterial flora. The underlying concept is to design probes recognizing sequence signatures specific for each node of the phylogenetic tree, extending the scope of oligoarrays to the whole bacterial kingdom. Many phylogenetic markers can be used for the design, such as groEL (Wertz et al., 2003; Wong and Chow, 2002), gyrA/gyrB (Antwerpen et al., 2007), rpoB (Drancourt et al., 2004; Troesch et al., 1999) or rpsA (Martens et al., 2007), but once again the rRNA gene remains the preferred candidate because of the number of available databases dedicated to this marker (see Section 2.2.1). Several teams developed and validated this approach (Palmer et al., 2006; Huyghe et al., 2008; Brodie et al., 2007), proving its efficiency in monitoring the contents of complex bacterial mixtures in a quantitative (fractional abundance $<0.1\%$) and qualitative fashion. Our group recently published a validation of a bacterial phylogenetic microarray (Huyghe et al., 2008) composed of oligonucleotidic probes targeting the 16S rRNA gene. Microarray design was performed by selecting about 9500 short oligonucleotides (25-mers, poly(T)-tailed to reach an overall length of 60 nucleotides) specific to nodes matching the different levels of the whole bacterial phylogenetic tree (Fig. 2). As part of the validation process, this array was used to characterize the bacterial gingival flora of two healthy subjects and results showed good agreement with previous cloning sequencing studies. Compared to cloning libraries, we also noted that phylogenetic microarrays revealed a broader bacterial diversity, a finding already reported by similar studies (Brodie et al., 2007; DeSantis et al., 2007), which could be explained by cloning bias, or by the paucity of clones sequenced. We developed this strategy in order to elucidate the etiological agent(s) of noma (Baratti-Mayer et al., 2003), a gangrenous disease affecting children and leading to the destruction of hard and soft tissues of the face. Our preliminary results (manuscript in preparation) suggest an imbalance in bacterial composition when comparing healthy children and noma patients, with a broader bacterial diversity in unaffected children. In contrast, *Peptostreptococcus* spp., *Prevotella* spp. and *Spirochaetaceae* are dominant phylotypes in noma lesions. Briefly, this study illustrates the usefulness of phylogenetic strategies in the understanding of unknown etiologies when a microbial origin is suspected.

Applying an rRNA phylogenetic microarray to monitor the temporal and geographical dynamics of bacterial populations in urban aerosols, Brodie et al. (2007) detected a surprising variety of human and animal pathogenic bacteria in the air of two American cities (Austin and San Antonio). The causative agents of glanders and melioidosis, *Burkholderia mallei* and *B. pseudomallei*, respectively, were regularly detected during the study; similarly, different pathogens such as *Campylobacteraceae*, *Clostridium botulinum*, *Helicobacter* and tick-borne *Rickettsia* were also detected in aerosol samples. This study showed that phylogenetic oligoarrays can be easily used to characterize bacterial pathogens from a dense background of non-pathogenic bacteria and could, for example, be exploited as part of a bio-surveillance initiative.

2.3. Universal pathogen diagnostic microarrays

In comparison to bacteria, the absence of ubiquitous gene targets makes the design of viral microarrays much more complex. An interesting approach was proposed by Wang et al. (2002) using a pan-viral microarray composed of approximately 1600 probes representing the most highly conserved 70-mers from 140 animal and human viruses. Probes were selected from each virus in order to maximize potential cross-hybridization to conserved regions and allowing detection of non-sequenced or unknown viruses. This strategy was combined with random PCR amplification, i.e. without use of sequence-specific primers. It proved very useful in identifying a previously uncharacterized coronavirus as the causative agent of an outbreak of severe acute respiratory syndrome (SARS) in 2003 (Wang et al., 2003). An unknown virus recovered from a SARS patient was cultured and hybridized on this pan-viral array. Probes with the highest hybridization intensities suggested that the virus belonged to – but was likely a new member of – the coronavirus family. Another original feature of this study was that the hybridized material was again retrieved from the array, cloned and sequenced for further characterization of this novel virus, highlighting the potential of microarrays as preparative molecular biology tools. In a previous study (Wang et al., 2002), the same team demonstrated that this approach, besides being able to distinguish between related viral serotypes, seems well adapted for detecting previously uncharacterized viruses. Indeed, maximizing potential cross-hybridization in conserved regions within the same viral genus allows identification of unsequenced viruses.

To date, the most comprehensive multi-microbial detection microarray was proposed by Palacios et al. (2007). This team developed a panmicrobial microarray comprising oligonucleotide probes targeting 1710 virus species and 135 bacterial, 73 fungal and 63 parasite genera. In addition, a few hundred probes recognizing host immune response genes were added to the design to potentially assess the host response. Unsurprisingly, the ribosomal RNA region was selected as a target marker for the detection of bacterial (16S), fungal (18S) and parasitic (18S) pathogens, while viral probes were selected from highly conserved regions. Using viral RNA extracted from cultured cells, detection limits of this microarray ranged from 1900 RNA copies for RNA viruses to 10,000 RNA copies for DNA viruses. The performance of the array was tested on clinical samples from patients suffering from various bacterial and viral diseases, and results correlated well with determinations performed by using classical methods. This panmicrobial approach has many interesting aspects, especially the possibility to identify unexpected pathogens. For example, the authors identified *Plasmodium falciparum* (the causative agent of malaria) in a patient formerly suspected of viral hemorrhagic-like disease.

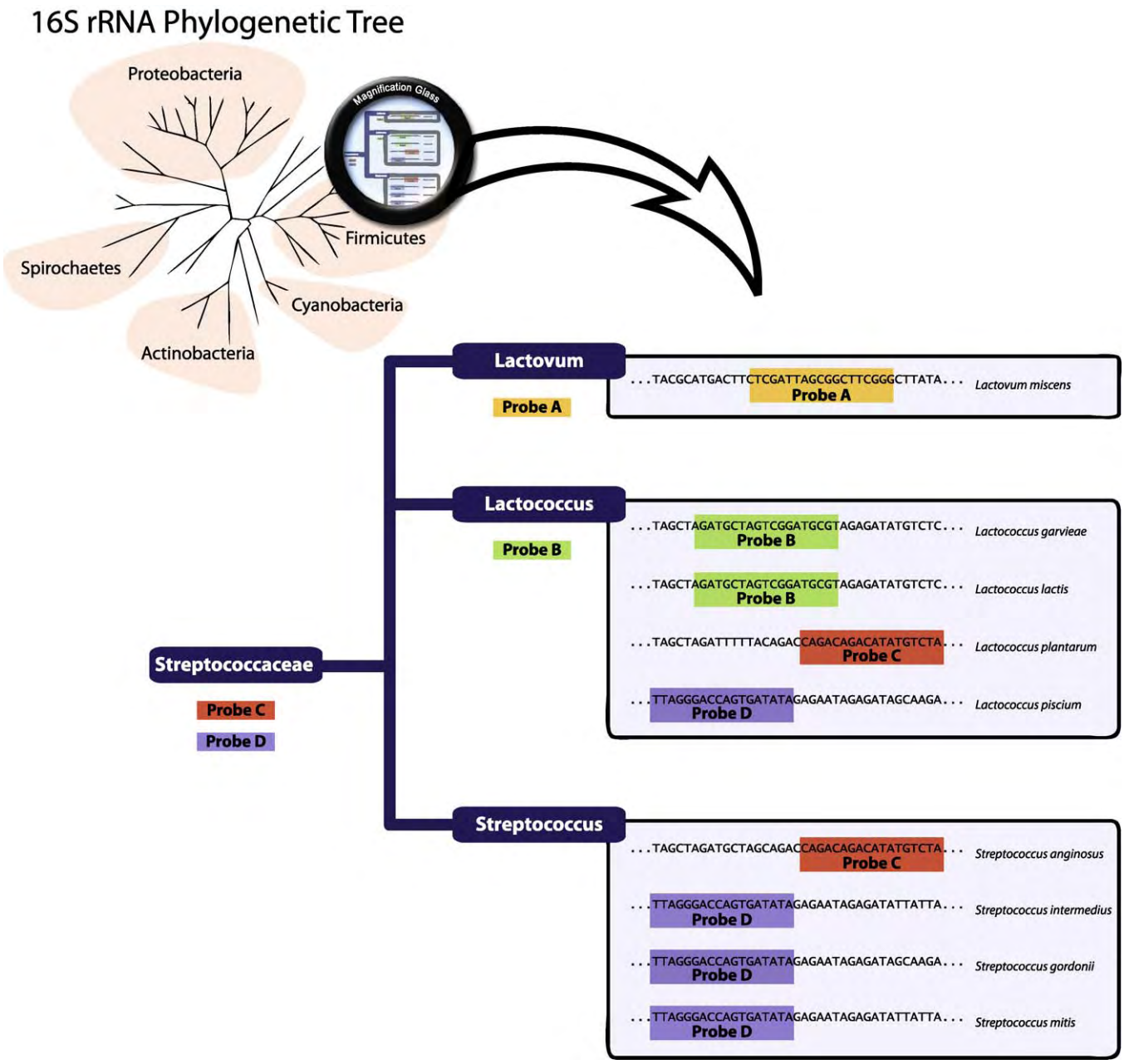


Fig. 2. Schematic representation of the probe selection process characterizing the 16S rRNA phylogenetic microarray as proposed by Huyghe et al. (2008). Candidate probes are compared to a 16S rRNA sequences database and assigned to the most distal common node. For example, probe D, which is common to the *Streptococcus* and *Lactococcus* genera, is assigned to the *Streptococcaceae* family level. In contrast, probe B is specific to some *Lactococcus* species, and it is therefore assigned to the genus level.

The design of the broadest panmicrobial diagnosis microarray involves the extensive knowledge of pathogen sequences, meaning that classical probe designs relying on prior knowledge of nucleotide sequences are limited by the number of known sequenced species and the quality of sequencing. To overcome this limitation we proposed a non-cognate approach relying on a probe design scheme that enables targeting multiple organisms without having prior knowledge of their genomic content. This experimental strategy, Non-cognate hybridization system (NCHS; patent WO/2000/075377; manuscript in preparation), is based on the utilization of two microarrays containing short probes (Fig. 1C) providing all combinations of only two nucleotides (i.e. A and C for the first array; A and G for the second array). Hybridization of genomic DNA from bacterial strains using this strategy results in a strain-specific probe pattern allowing direct identification of bacterial contents in samples. Though pattern identification relies on sequence databases, pattern recognition algorithms (such as self-organizing maps) can be used to identify unknown pathogens by homology to closely related and known organisms. However, due to the high variation in the GC content of capture elements, experimental conditions have to be tailored to limit potential bias during hybridization. The utilization of chemical compounds such as formamide (Sarkar et al., 1990), betaine (Henke et al., 1997) or tetramethylammonium chloride (Sorg et al., 1991) during hybridization or washing steps generally limit effect related to heterogeneity in melting temperatures of probes. Although this approach needs extensive enhancement and validation, it should prove efficient in detecting a broad variety of microbial pathogens.

3. General considerations for probe selection

The key point for setting up oligonucleotide microarray studies is the selection of an adequate set of probes. Ideally, three points must be considered during the probe selection process. First, all probes must be highly specific for the target gene while avoiding cross-hybridization with non-target sequences (specificity). Secondly, probes must display a high sensitivity in order to detect low abundant target sequences from a sample. Finally, the set of probes must exhibit homogenous thermodynamic properties ensuring similar hybridization behavior.

Oligonucleotide length plays a crucial role in the probe tolerance to mismatches towards its target. Short probes (25–35 nt) tend to be more adapted to identify mismatched position**s**, and are thus probably more suitable for phylogenetic microarrays. However, longer probes display better sensitivity but weaker specificity (Relogio et al., 2002). A good strategy to increase sensitivity relies on the use of spacers, such as thymidine tails, on short oligonucleotides. Increasing the length of short probes reduces steric hindrance by pushing the probe away from the array surface (Peplies et al., 2003).

Presence of mismatches at the central position of short oligonucleotides tends to affect probe-target duplex stability. For longer oligonucleotides, mismatches in the distal region (solution end) have less impact on the probe-target duplex stability than mismatches on the proximal region (surface end) of the probe (Hughes et al., 2001).

Both PCR products (cDNA) and oligonucleotides can be used as probes for microarray fabrication. Although PCR is relatively simple to implement in an average laboratory, oligonucleotide probes are more reliable in terms of sensitivity and specificity (Li et al., 2002; Kothapalli et al., 2002; Barczak et al., 2003), and provide a flexible design.

From a structural point of view, probe-target hybridization is influenced by the formation of inter- and/or intra-molecular secondary structures (stem loops, helices, dimerization) and by the affinity of the probe to bind to a target; which follow thermodynamic predictions. Various free tools, such as Hyther (Bommarito et al., 2000), Mfold (Zuker, 2003), or RNAfold (Hofacker and Stadler, 2006), are available for the estimation of thermodynamic properties of probes and probe-target duplexes, including prediction of Gibbs free energies (ΔG) and melting temperatures (T_m). However these tools rely on nearest-neighbor models and have been described with experimental observations performed in solution and may not reflect the actual properties of probes attached on a solid surface. This difference was recently illustrated by Pozhitkov et al. (2006) who observed a poor statistical correlation between prediction of Gibbs free energies and signal intensities of oligoarray probes targeting eukaryotic rRNA genes.

In summary, thermodynamic evaluation of oligonucleotide properties helps for optimizing hybridization sensitivity and specificity; nonetheless, careful wet-lab testing is required to evaluate hybridization properties of oligonucleotide probes before considering their use in other studies.

4. Sample preparation

Several caveats should be considered when discussing potential application of diagnostic arrays for routine bacterial identification. The risk of false-negative results is real and should be minimized. Most published microarray methods use amplification strategies prior to hybridization that are based on the polymerase chain reaction (Saiki et al., 1988; Mullis et al., 1986), a proven and sensitive technique. However, PCR is susceptible to numerous inhibitory compounds commonly present in clinical materials (Monteiro et al., 1997; Fredricks and Relman, 1998; Longo et al., 1990); fastidious sample preparation is thus required to obtain robust detection (Klein et al., 1997). In addition, some microorganisms potentially responsible for infections harbor specific envelopes or cell-walls which are either difficult to disrupt or further release inhibitors of enzymatic reactions. All these parameters should be considered carefully depending on the intended application and the types of samples to be analyzed, which both affect the various steps needed for detection or identification purposes.

Labeled nucleic acids consist of either DNA or RNA. Most of the time, nucleic acids require enzymatic reactions to be performed prior to hybridization, to either amplify and/or label the products. Characterization or genotyping microarrays typically rely on randomly amplified DNA fragments, often generated by the Klenow fragment (Koessler et al., 2006; Charbonnier et al., 2005) or a phage polymerase. The latter allows converting minute amounts of DNA to μg of hybridizable products (Pinard et al., 2006). Expression microarrays use either labeled RNA (Francois et al., 2006) or products of these RNAs obtained by reverse-transcription (Scherl et al., 2006). Recently, alternative RNA amplification strategies have been used, allowing for a reduction of the amount of starting cell numbers (Francois et al., 2007; La et al., 2007), an important issue in many biological applications.

At least two different strategies are currently available to fluorescently label nucleic acids, either directly or indirectly. The direct incorporation uses nucleotide derivatives that are directly coupled to fluorescent molecules (Scherl et al., 2006). An enzymatic amplification enables the incorporation of such residues and yields to fluorescently-labeled nucleic acids that can be resolved by hybridization on the array. Indirect labeling is a two-step reaction (Schroeder et al., 2002): the first step consists of incorporating nucleotides containing small chemically reactive groups, generally aminoallyl dUTP. Subsequently, activated fluorescent dyes are coupled (generally *N*-hydroxysuccinimidyl ester cyanine, such as Cy-3 and Cy-5) yielding labeled nucleic acids. This two-step strategy reduces the risk of biases introduced during the enzymatic incorporation of modified nucleotides, mostly due to the steric effects related to the presence of large aromatic fluorescent molecules.

While the use of fluorescent dyes is well established, alternative non-fluorescent labeling methods are available. An affordable alternative is to couple biotin-derivatives to target molecules, and then capture them with a streptavidin-conjugate enzyme (e.g. horseradish peroxidase HRP – streptavidin). Afterwards, numerous HRP chromogenic (e.g. tetramethylbenzidine) or chemiluminescent substrates can be used to detect hybridized targets. Resonance light scattering (RLS) can also be used as a more sensitive alternative to fluorescent labeling methods. When a gold (or silver) nanoparticle is illuminated with a white light beam, scattering produces an emitted light characterized by a defined wavelength. RLS can be exploited as a labeling method after hybridization of biotinylated targets and incubation with nanoparticles coupled with streptavidin. The procedure was tested against conventional fluorescence methods (Francois et al., 2003) to directly label and detect bacterial RNA without prior enzymatic amplification, thus limiting potential amplification bias and processing time. Compared to fluorescent labeling, RLS produces a consistently higher signal-to-noise ratio while providing a gain of about 50-fold in detection sensitivity.

5. Conclusion

DNA microarray technology, consisting of the hybridization of nucleic acid fragments (either from chemical or enzymatic

Table 1
Comparison of microarrays used for microbial pathogens characterization.

Array type	Target	Prior knowledge of sequence	Applicable to uncharacterized organisms	Highest resolution	Sensitivity	Limits
Whole-genome arrays	ORF (and inter-ORF)	Yes	Yes	Strains	–	Organism specific
Functional gene arrays	Family of functional genes	Yes	Yes	Species	3.1 fg (Jaing et al., 2008) • 10^1 bacteria (Wiesinger-Mayr et al., 2007)	Limited to specific functional families
Conventional rRNA arrays	rRNA gene	Yes	Yes	Species	• 10^1 bacteria (Wiesinger-Mayr et al., 2007) • 2.5–5.0 fmole of *S. aureus* 16S (Francois et al., 2006)	Design of species-discriminating probes can be tricky
Phylogenetic microarrays	rRNA gene or housekeeping genes	Yes	Yes	Species	0.1% of total community (Palmer et al., 2006)	Limited to phylogenetic studies
Tiling arrays	Whole genome	Yes	Yes	Strains-SNP	–	Organism specific
Universal microarrays						
Palacios et al.	rRNA gene and functionally related sequences	Yes	Yes	Species	• RNA viruses: 1900 RNA copies • DNA viruses: 10,000 RNA copies	Design of species-discriminating rRNA probes can be tricky
NCHS	Whole genome	No	Yes	Strains	To be determined	Inhomogeneous GC content requires use of additives

synthesis) immobilized onto a solid surface with the nucleic acid of a test sample, was first reported at the end of the 1990s (Schena et al., 1995). Constant improvements have been made regarding on properties of chip surfaces, nucleic acid density, the design of capture element, the dyes used to label the samples as well as in the sensitivity of detection devices. The format of DNA microarrays appears particularly adapted to the identification, characterization or genotyping of isolated microbial pathogens as well as to more clinical questions such as the assessment of virulence or antibiotic resistance genes. However, direct analysis of nucleic acid is generally not possible and the lack of sensitivity of the technology requires the use of amplification methods, introducing some potential biases in the content of the sample. To date, no commercial microbial diagnostic microarrays, approved by the food and drug administrations, are available on the clinical diagnostics market.

One of the major challenges in that field derives from the nature of the sample. Biological fluids are complex medium potentially containing pathogenic organisms (often rare) in a community of commensal organisms and in the presence of variable amounts of host material. In such cases, amplification of the pathogenic material is essential while the detection strategy must avoid amplified host nucleic acids. In addition, characterization and identification of co-infective agents represents another major challenge.

The high parallelism of microarray platforms makes it possible to address thousands of questions during a single experiment; they appear therefore particularly suited for bacterial detection or identification. Furthermore, isolated microorganisms can be further characterized using high resolution genotyping (Quiñones et al., 2007; Zhang et al., 2006), by providing an extensive documentation of virulence factors (Lindsay et al., 2006; Saunders et al., 2004) or for specific epidemiological survey purposes (Garaizar et al., 2006). Note that the depth or precision of data collected is generally dependent on the number of capture elements present on the surface of the microarray. This number varies from hundreds (Saunders et al., 2004) for virulence factors to hundreds of thousands for a resolution at the nucleotide level (Zhang et al., 2006). For specific applications, a resolution at the nucleotide level is also achievable using more elaborated strategies such as padlock probes (Szemes et al., 2005).

Generally, the question to be answered will guide the choice of both the approach and the markers (see Table 1) to assess identification. For example, rRNA genes are undoubtedly the most common gene marker used in microbial detection microarrays. However, this phylogenetic marker is not always adapted for species determination, due to: (i) its high conservation between certain closely related species, and (ii) lack of accurate thermodynamic properties to assess quality of probes targeting rRNA genes (Pozhitkov et al., 2006). Alternative phylogenetic markers may be preferable in that specific situation. In addition, the design of microarrays based on cognate approaches is entirely dependent on the quality of sequence databases available to researchers. To date, a small fraction of the microbial world has been sequenced and characterization of unknown pathogens using microarray involves making assumptions from sequenced genomes. These different challenges appear surmountable using adapted design strategies, technical procedures and specific analytical solutions.

Over the past year, high throughput DNA sequencing technology has become more accessible to the scientific community and could, in the very near future, become an alternative to microbial diagnostics microarrays. However, microarrays will not necessarily be put aside as they could, for example, be exploited as a preparative platform to capture nucleic acids (Porreca et al., 2007; Wang et al., 2003). Clinical microbiology is rapidly evolving and molecular methods are now extensively challenging conventional microbiology techniques. The diagnostic field will benefit from this evolution as microarray technology appears to be the ideal tool to assess the diversity of the microbial world as well as to improve our knowledge of host–pathogen interactions.

Acknowledgements

The authors' efforts in preparing this review were supported by the COST C05.0103 (JS) (Action B28: Array Technologies for BSL3 and BSL4 Pathogens–Workgroup 1) and by grants from the Swiss

National Science Foundation 3100A0-116075/1 (PF) and 3100A0-112370/1 (JS). AH is supported by the Geneva Study Group on Noma (GESNOMA), a research group funded by the Hirzel Foundation. Many thanks to Dr Katrine Whiteson for careful reading of the manuscript.

References

Antwerpen, M.H., Schellhase, M., Ehrentreich-Förster, E., Bier, F., Witte, W., Nübel, U., 2007. DNA microarray for detection of antibiotic resistance determinants in Bacillus anthracis and closely related Bacillus cereus. Mol. Cell. Probes 21, 152–160.

Baratti-Mayer, D., Pittet, B., Montandon, D., Bolivar, I., Bornand, J.E., Hugonnet, S., Jaquinet, A., Schrenzel, J., Pittet, D., 2003. Noma: an "infectious" disease of unknown aetiology. Lancet Infect. Dis. 3, 419–431.

Barczak, A., Rodriguez, M.W., Hanspers, K., Koth, L.L., Tai, Y.C., Bolstad, B.M., Speed, T.P., Erle, D.J., 2003. Spotted long oligonucleotide arrays for human gene expression analysis. Genome Res. 13, 1775–1785.

Bommarito, S., Peyret, N., SantaLucia, J., 2000. Thermodynamic parameters for DNA sequences with dangling ends. Nucleic Acids Res. 28, 1929–1934.

Brodie, E.L., DeSantis, T.Z., Joyner, D.C., Baek, S.M., Larsen, J.T., Andersen, G.L., Hazen, T.C., Richardson, P.M., Herman, D.J., Tokunaga, T.K., Wan, J.M., Firestone, M.K., 2006. Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. Appl. Environ. Microbiol. 72, 6288–6298.

Brodie, E.L., DeSantis, T.Z., Parker, J.P.M., Zubietta, I.X., Piceno, Y.M., Andersen, G.L., 2007. Urban aerosols harbor diverse and dynamic bacterial populations. Proc. Natl. Acad. Sci. 104, 299–304.

Charbonnier, Y., Gettler, B., François, P., Bento, M., Renzoni, A., Vaudaux, P., Schlegel, W., Schrenzel, J., 2005. A generic approach for the design of whole-genome oligoarrays, validated for genomotyping, deletion mapping and gene expression analysis on *Staphylococcus aureus*. BMC Genomics 6, 95.

Claridge, M.F., Dawah, H.A., Wilson, M.R., Association, S., 1997. Species: The Units of Biodiversity. Springer.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72, 5069–5072.

DeSantis, T.Z., Brodie, E.L., Moberg, J.P., Zubieta, I.X., Piceno, Y.M., Andersen, G.L., 2007. High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. Microb. Ecol. 53, 371–383.

Drancourt, M., Roux, V., Fournier, P.E., Raoult, D., 2004. rpoB gene sequence-based identification of aerobic Gram-positive cocci of the genera Streptococcus, Enterococcus, Gemella, Abiotrophia, and Granulicatella. J. Clin. Microbiol. 42, 497–504.

Fitzgerald, J.R., Sturdevant, D.E., Mackie, S.M., Gill, S.R., Musser, J.M., 2001. Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. Proc. Natl. Acad. Sci. 98, 8821–8826.

Francois, P., Bento, M., Vaudaux, P., Schrenzel, J., 2003. Comparison of fluorescence and resonance light scattering for highly sensitive microarray detection of bacterial pathogens. J. Microbiol. Methods 55, 755–762.

Francois, P., Charbonnier, Y., Jacquet, J., Utinger, D., Bento, M., Lew, D., Kresbach, G.M., Ehrat, M., Schlegel, W., Schrenzel, J., 2006. Rapid bacterial identification using evanescent-waveguide oligonucleotide microarray classification. J. Microbiol. Methods 65, 390–403.

Francois, P., Garzoni, C., Bento, M., Schrenzel, J., 2007. Comparison of amplification methods for transcriptomic analyses of low abundance prokaryotic RNA sources. J. Microbiol. Methods 68, 385–391.

Fredricks, D.N., Relman, D.A., 1998. Improved amplification of microbial DNA from blood cultures by removal of the PCR inhibitor sodium polyanetholesulfonate. J. Clin. Microbiol. 36, 2810–2816.

Garaizar, J., Rementeria, A., Porwollik, S., 2006. DNA microarray technology: a new tool for the epidemiological typing of bacterial pathogens? FEMS Immunol. Med. Microbiol. 47, 178–189.

Grasselli, E., François, P., Gutacker, M., Gettler, B., Benagli, C., Convert, M., Boerlin, P., Schrenzel, J., Piffaretti, J.C., 2008. Evidence of horizontal gene transfer between human and animal commensal *Escherichia coli* strains identified by microarray. FEMS Immunol. Med. Microbiol 53 (3), 351–358.

He, Z., Gentry, T.J., Schadt, C.W., Wu, L., Liebich, J., Chong, S.C., Huang, Z., Wu, W., Gu, B., Jardine, P., Criddle, C., Zhou, J., 2007. GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. ISME J. 1, 67–77.

Henke, W., Herdel, K., Jung, K., Schnorr, D., Loening, S.A., 1997. Betaine improves the PCR amplification of GC-rich DNA sequences. Nucleic Acids Res. 25, 3957–3958.

Herring, C.D., Palsson, B., 2007. An evaluation of comparative genome sequencing (CGS) by comparing two previously-sequenced bacterial genomes. BMC Genomics 8, 274.

Hofacker, I.L., Stadler, P.F., 2006. Memory efficient folding algorithms for circular RNA secondary structures. Bioinformatics (Oxford, England) 22, 1172–1176.

Hong, B.X., Jiang, L.F., Hu, Y.S., Fang, D.Y., Guo, H.Y., 2004. Application of oligonucleotide array technology for the rapid detection of pathogenic bacteria of foodborne infections. J. Microbiol. Methods 58, 403–411.

Hsiao, C.R., Huang, L., Bouchara, J.P., Barton, R., Li, H.C., Chang, T.C., 2005. Identification of medically important molds by an oligonucleotide array. J. Clin. Microbiol. 43, 3760–3768.

Hu, G., Liu, I., Sham, A., Stajich, J.E., Dietrich, F.S., Kronstad, J.W., 2008. Comparative hybridization reveals extensive genome variation in the AIDS-associated pathogen Cryptococcus neoformans. Genome Biol. 9, R41.

Huang, A., Li, J.W., Shen, Z.Q., Wang, X.W., Jin, M., 2006. High-throughput identification of clinical pathogenic fungi by hybridization to an oligonucleotide microarray. J. Clin. Microbiol. 44, 3299–3305.

Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., Kobayashi, S., Davis, C., Dai, H., He, Y.D., Stephaniants, S.B., Cavet, G., Walker, W.L., West, A., Coffey, E., Shoemaker, D.D., Stoughton, R., Blanchard, A.P., Friend, S.H., Linsley, P.S., 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat. Biotechnol. 19, 342–347.

Huyghe, A., Francois, P., Charbonnier, Y., Tangomo-Bento, M., Bonetti, E.J., Paster, B.J., Bolivar, I., Baratti-Mayer, D., Pittet, D., Schrenzel, J., 2008. Novel microarray design strategy to study complex bacterial communities. Appl. Environ. Microbiol. 74, 1876–1885.

Jaing, C., Gardner, S., McLoughlin, K., Mulakken, N., Alegria-Hartman, M., Banda, P., Williams, P., Gu, P., Wagner, M., Manohar, C., Slezak, T., 2008. A functional gene array for detection of bacterial virulence elements. PLoS ONE 3, e2163.

Kim, C.C., Joyce, E.A., Chan, K., Falkow, S., 2002. Improved analytical methods for microarray-based genome-composition analysis. Genome Biol. 3 (11) RESEARCH0065.

Klein, A., Barsuk, R., Dagan, S., Nusbaum, O., Shouval, D., Galun, E., 1997. Comparison of methods for extraction of nucleic acid from hemolytic serum for PCR amplification of hepatitis B virus DNA sequences. J. Clin. Microbiol. 35, 1897–1899.

Koessler, T., Francois, P., Charbonnier, Y., Huyghe, A., Bento, M., Dharan, S., Renzi, G., Lew, D., Harbarth, S., Pittet, D., Schrenzel, J., 2006. Use of oligoarrays for characterization of community-onset methicillin-resistant *Staphylococcus aureus*. J. Clin. Microbiol. 44, 1040–1048.

Kothapalli, R., Yoder, S.J., Mane, S., Loughran, T.P., 2002. Microarray results: how accurate are they? BMC Bioinform. 3, 22.

Kozal, M.J., Shah, N., Shen, N., Yang, R., Fucini, R., Merigan, T.C., Richman, D.D., Morris, D., Hubbell, E., Chee, M., Gingeras, T.R., 1996. Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. Nat. Med. 2, 753–759.

La, M.V., François, P., Rovery, C., Robineau, S., Barbry, P., Schrenzel, J., Raoult, D., Renesto, P., 2007. Development of a method for recovering rickettsial RNA from infected cells to analyze gene expression profiling of obligate intracellular bacteria. J. Microbiol. Methods 71, 292–297.

Leaw, S.N., Chang, H.C., Barton, R., Bouchara, J.P., Chang, T.C., 2007. Identification of medically important candida and non-Candida yeast species by an oligonucleotide array. J. Clin. Microbiol. 45, 2220–2229.

Leinberger, D.M., Schumacher, U., Autenrieth, I.B., Bachmann, T.T., 2005. Development of a DNA microarray for detection and identification of fungal pathogens involved in invasive mycoses. J. Clin. Microbiol. 43, 4943–4953.

Leonard, E.E., Takata, T., Blaser, M.J., Falkow, S., Tompkins, L.S., Gaynor, E.C., 2003. Use of an open-reading frame-specific Campylobacter jejuni DNA microarray as a new genotyping tool for studying epidemiologically related isolates. J. Infect. Dis. 187, 691–694.

Li, J., Pankratz, M., Johnson, J.A., 2002. Differential gene expression patterns revealed by oligonucleotide versus long cDNA arrays. Toxicol. Sci. 69, 383–390.

Lin, B., Malanoski, A.P., Wang, Z., Blaney, K.M., Ligler, A.G., Rowley, R.K., Hanson, E.H., von Rosenvinge, E., Ligler, F.S., Kusterbeck, A.W., Metzgar, D., Barrozo, C.P., Russell, K.L., Tibbetts, C., Schnur, J.M., Stenger, D.A., 2007. Application of broad-spectrum, sequence-based pathogen identification in an urban population. PLoS ONE 2, e419.

Lin, B., Wang, Z., Vora, G.J., Thornton, J.A., Schnur, J.M., Thach, D.C., Blaney, K.M., Ligler, A.G., Malanoski, A.P., Santiago, J., Walter, E.A., Agan, B.K., Metzgar, D., Seto, D., Daum, L.T., Kruzelock, R., Rowley, R.K., Hanson, E.H., Tibbetts, C., Stenger, D.A., 2006. Broad-spectrum respiratory tract pathogen identification using resequencing DNA microarrays. Genome Res. 16, 527–535.

Lindsay, J.A., Moore, C.E., Day, N.P., Peacock, S.J., Witney, A.A., Stabler, R.A., Husain, S.E., Butcher, P.D., Hinds, J., 2006. Microarrays reveal that each of the ten dominant lineages of *Staphylococcus aureus* has a unique combination of surface-associated and regulatory genes. J. Bacteriol. 188, 669–676.

Liu, Y., Han, J.X., Huang, H.Y., Zhu, B., 2005. Development and evaluation of 16S rDNA microarray for detecting bacterial pathogens in cerebrospinal fluid. Exp. Biol. Med. 230, 587–591.

Longo, M.C., Berninger, M.S., Hartley, J.L., 1990. Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. Gene 93, 125–128.

Loy, A., Schulz, C., Lücker, S., Schöpfer-Wendels, A., Stoecker, K., Baranyi, C., Lehner, A., Wagner, M., 2005. 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "Rhodocyclales". Appl. Environ. Microbiol. 71, 1373–1386.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., Konig, A., Liss, T., Lussmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A.,

Lenke, M., Ludwig, T., Bode, A., Schleifer, K.H., 2004. ARB: a software environment for sequence data. Nucleic Acids Res. 32, 1363–1371.
Maidak, B.L., Cole, J.R., Lilburn, T.G., Parker, C.T., Saxman, P.R., Farris, R.J., Garrity, G.M., Olsen, G.J., Schmidt, T.M., Tiedje, J.M., 2001. The RDP-II (Ribosomal Database Project). Nucleic Acids Res. 29, 173–174.
Martens, M., Weidner, S., Linke, B., de Vos, P., Gillis, M., Willems, A., 2007. A prototype taxonomic microarray targeting the rpsA housekeeping gene permits species identification within the rhizobial genus Ensifer. Syst. Appl. Microbiol. 30, 390–400.
Monteiro, L., Bonnemaison, D., Vekris, A., Petry, K.G., Bonnet, J., Vidal, R., Cabrita, J., Mégraud, F., 1997. Complex polysaccharides as PCR inhibitors in feces: *Helicobacter pylori* model. J. Clin. Microbiol. 35, 995–998.
Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H., 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb. Symp. Quant. Biol. 51 (Pt 1), 263–273.
Neufeld, J.D., Mohn, W.W., de Lorenzo, V., 2006. Composition of microbial communities in hexachlorocyclohexane (HCH) contaminated soils from Spain revealed with a habitat-specific microarray. Environ. Microbiol. 8, 126–140.
Nübel, U., Schmidt, P.M., Reiss, E., Bier, F., Beyer, W., Naumann, D., 2004. Oligonucleotide microarray for identification of Bacillus anthracis based on intergenic transcribed spacers in ribosomal DNA. FEMS Microbiol. Lett. 240, 215–223.
Palacios, G., Quan, P.l., Jabado, O.J., Conlan, S., Hirschberg, D.L., Liu, Y., Zhai, J., Renwick, N., Hui, J., Hegyi, H., Grolla, A., Strong, J.E., Towner, J.S., Geisbert, T.W., Jahrling, P.B., Büchen-Osmond, C., Ellerbrok, H., Sanchez-Seco, M.P., Lussier, Y., Formenty, P., Nichol, M.S., Feldmann, H., Briese, T., Lipkin, W.I., 2007. Panmicrobial oligonucleotide array for diagnosis of infectious diseases. Emerg. Infect. Dis. 13, 73–81.
Palmer, C., Bik, E.M., Eisen, M.B., Eckburg, P.B., Sana, T.R., Wolber, P.K., Relman, D.A., Brown, P.O., 2006. Rapid quantitative profiling of complex microbial populations. Nucleic Acids Res. 34, e5.
Peplies, J., Glöckner, F.O., Amann, R., 2003. Optimization strategies for DNA microarray-based detection of bacteria with 16S rRNA-targeting oligonucleotide probes. Appl. Environ. Microbiol. 69, 1397–1407.
Perreten, V., Vorlet-Fawer, L., Slickers, P., Ehricht, R., Kuhnert, P., Frey, J., 2005. Microarray-based detection of 90 antibiotic resistance genes of gram-positive bacteria. J. Clin. Microbiol. 43, 2291–2302.
Pinard, R., de Winter, A., Sarkis, G.J., Gerstein, M.B., Tartaro, K.R., Plant, R.N., Egholm, M., Rothberg, J.M., Leamon, J.H., 2006. Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. BMC Genomics 7, 216.
Porreca, G.J., Zhang, K., Li, J.B., Xie, B., Austin, D., Vassallo, S.L., LeProust, E.M., Peck, B.J., Emig, C.J., Dahl, F., Gao, Y., Church, G.M., Shendure, J., 2007. Multiplex amplification of large sets of human exons. Nat. Methods 4, 931–936.
Pozhitkov, A., Noble, P.A., Domazet-Loso, T., Nolte, A.W., Sonnenberg, R., Staehler, P., Beier, M., Tautz, D., 2006. Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. Nucleic Acids Res. 34, e66.
Quiñones, B., Parker, C.T., Janda, J.M., Miller, W.G., Mandrell, R.E., 2007. Detection and genotyping of Arcobacter and Campylobacter isolates from retail chicken samples by use of DNA oligonucleotide arrays. Appl. Environ. Microbiol. 73, 3645–3655.
Relogio, A., Schwager, C., Richter, A., Ansorge, W., Valcarcel, J., 2002. Optimization of oligonucleotide-based DNA microarrays. Nucleic Acids Res. 30, e51.
Rhee, S.K., Liu, X., Wu, L., Chong, S.C., Wan, X., Zhou, J., 2004. Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. Appl. Environ. Microbiol. 70, 4303–4317.
Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A., 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487–491.
Sarkar, G., Kapelner, S., Sommer, S.S., 1990. Formamide can dramatically improve the specificity of PCR. Nucleic Acids Res. 18, 7465.
Saunders, N.A., Underwood, A., Kearns, A.M., Hallas, G., 2004. A virulence-associated gene microarray: a tool for investigation of the evolution and pathogenic potential of *Staphylococcus aureus*. Microbiology 150, 3763–3771.
Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270, 467–470.
Scherl, A., François, P., Charbonnier, Y., Deshusses, J.M., Koessler, T., Huyghe, A., Bento, M., Stahl-Zeng, J., Fischer, A., Masselot, A., Vaezzadeh, A., Gallé, F., Renzoni, A., Vaudaux, P., Lew, D., Zimmermann-Ivol, C.G., Binz, P.A., Sanchez, J.C., Hochstrasser, D.F., Schrenzel, J., 2006. Exploring glycopeptide-resistance in *Staphylococcus aureus*: a combined proteomics and transcriptomics approach for the identification of resistance-related markers. BMC Genomics 7, 296.
Schroeder, B.G., Peterson, L.M., Fleischmann, R.D., 2002. Improved quantitation and reproducibility in Mycobacterium tuberculosis DNA microarrays. J. Mol. Microbiol. Biotechnol. 4, 123–126.
Sergeev, N., Distler, M., Vargas, M., Chizhikov, V., Herold, K.E., Rasooly, A., 2006. Microarray analysis of Bacillus cereus group virulence factors. J. Microbiol. Methods 65, 488–502.
Sorg, U., Enczmann, J., Sorg, R., Wernet, P., 1991. Rapid non-radioactive TMACl hybridization protocol employing enzymatically labeled oligonucleotides. Nucleic Acids Res. 19, 4782.
Szemes, M., Bonants, P., de Weerdt, M., Baner, J., Landegren, U., Schoen, C.D., 2005. Diagnostic application of padlock probes—multiplex detection of plant pathogens using universal microarrays. Nucleic Acids Res. 33, e70.
Tiquia, S.M., Wu, L., Chong, S.C., Passovets, S., Xu, D., Xu, Y., Zhou, J., 2004. Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. BioTechniques 36, 664–670 672, 674.
Troesch, A., Nguyen, H., Miyada, C.G., Desvarenne, S., Gingeras, T.R., Kaplan, P.M., Cros, P., Mabilat, C., 1999. Mycobacterium species identification and rifampin resistance testing with high-density DNA probe arrays. J. Clin. Microbiol. 37, 49–55.
Volokhov, D., Rasooly, A., Chumakov, K., Chizhikov, V., 2002. Identification of listeria species by microarray-based assay. J. Clin. Microbiol. 40, 4720–4728.
Wang, D., Coscoy, L., Zylberberg, M., Avila, P.C., Boushey, H.A., Ganem, D., DeRisi, J.L., 2002. Microarray-based detection and genotyping of viral pathogens. Proc. Natl. Acad. Sci. 99, 15687–15692.
Wang, D., Urisman, A., Liu, Y.T., Springer, M., Ksiazek, T.G., Erdman, D.D., Mardis, E.R., Hickenbotham, M., Magrini, V., Eldred, J., Latreille, J.P., Wilson, R.K., Ganem, D., DeRisi, J.L., 2003. Viral discovery and sequence recovery using DNA microarrays. PLoS Biol. 1, e2.
Wang, Z., Daum, L.T., Vora, G.J., Metzgar, D., Walter, E.A., Canas, L.C., Malanoski, A.P., Lin, B., Stenger, D.A., 2006. Identifying influenza viruses with resequencing microarrays. Emerg. Infect. Dis. 12, 638–646.
Wertz, J.E., Goldstone, C., Gordon, D.M., Riley, M.A., 2003. A molecular phylogeny of enteric bacteria and implications for a bacterial species concept. J. Evol. Biol. 16, 1236–1248.
Wiesinger-Mayr, H., Vierlinger, K., Pichler, R., Kriegner, A., Hirschl, A., Presterl, E., Bodrossy, L., Noehammer, C., 2007. Identification of human pathogens isolated from blood using microarray hybridisation and signal pattern recognition. BMC Microbiol. 7, 78.
Wong, C.W., Albert, T.J., Vega, V.B., Norton, J.E., Cutler, D.J., Richmond, T.A., Stanton, L.W., Liu, E.T., Miller, L.D., 2004. Tracking the evolution of the SARS coronavirus using high-throughput, high-density resequencing arrays. Genome Res. 14, 398–405.
Wong, R.S.Y., Chow, A.W., 2002. Identification of enteric pathogens by heat shock protein 60 kDa (HSP60) gene sequences. FEMS Microbiol. Lett. 206, 107–113.
Zhang, W., Qi, W., Albert, T.J., Motiwala, A.S., Alland, D., Hyytia-Trees, E.K., Ribot, E.M., Fields, P.I., Whittam, T.S., Swaminathan, B., 2006. Probing genomic diversity and evolution of *Escherichia coli* O157 by single nucleotide polymorphisms. Genome Res. 16, 757.
Zhu, L.X., Wang, D., Zhang, G.B., Jiang, D., Zhang, Z.W., Zhang, Q., Mitchelson, K., Cheng, J., 2007. Development of a base stacking hybridization-based microarray method for rapid identification of clinical isolates. Diagn. Microbiol. Infect. Dis. 59, 149–156.
Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.

1.3 Studying bacteriology of Noma using microarrays

1.3.1 Noma: a disease of unknown etiology

Noma (from the Greek *nomein* - to devour) is a gangrenous disease of unknown etiology that affects the hard and soft tissues of the mouth and face. This disease principally affects children in the less developed countries suffering from bad nutrition and poor oral hygiene, and is associated with high mortality rates; around 100% before the antibiotic era to 25-30% nowadays (Baratti-Mayer et al., 2003).

The disease begins with an ulceration and necrosis of the gingival papilla (the gum filling space between adjacent teeth), spreading rapidly to the adjacent tissues, causing bone and cheek mucosa destruction (see figure 10). The resulting mutilations prevent children from drinking and eating, and many die because of starvation or secondary infections.

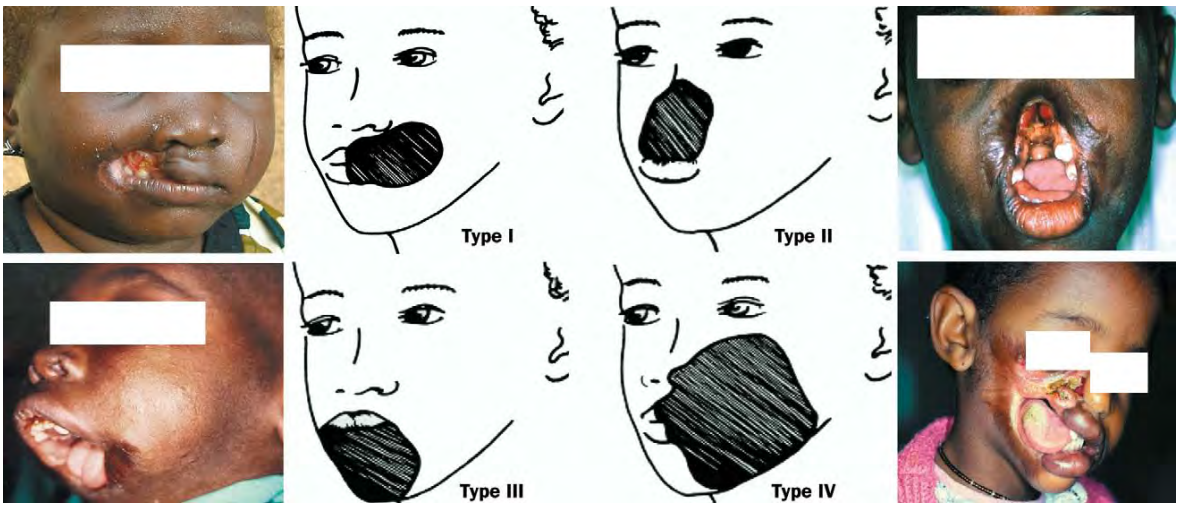


Fig. 10: Classification of noma sequelae. Adapted from (Baratti-Mayer et al., 2003).

1.3.1.1 Epidemiology

Noma was common in Europe until the 19th century, and cases were reported in the 20th century in Nazi concentration camps (Enwonwu et al., 2006). Nowadays, noma mainly affects African children inhabiting the sub-Saharian belt, but cases have been reported in Asia (Bourgeois and Leclercq, 1999; Srour et al., 2008) and in Latin America (Barmes et al., 1997). Acute noma is seen in young children (1-4 years), but late stages can be seen in teenagers and adults (Enwonwu et al., 2006). The exact incidence of children affected by noma is unknown, but in 1998 the World Health Organization estimated the worldwide incidence to 140'000 cases per year (Bourgeois and Leclercq, 1999). In the same year, the WHO estimated a mortality rate of 90% for untreated children.

1.3.1.2 Etiopathogenesis

Several risk factors play a role in the pathogenesis of noma such as malnutrition, poor oral hygiene, weakened immune functions and previous infections such as malaria, measles and tuberculosis (Enwonwu et al., 1999).

Chronic malnutrition is unfortunately frequent in less developed countries and begins during infancy. Maternal under-nutrition results in intra-uterine growth retardation (IUGR) and low birth weight. IUGR limits development of the thymus gland and the production of interleukin 7 and its receptor. Knowing that interleukin 7 promotes T-cell development, this results in children with a weakened immune system (Enwonwu, 2007). Moreover, malnourished mothers cannot properly breast-feed their children, promoting infantile malnutrition and susceptibility to infections (breast feeding in the first months of life strengthens thymic function).

Several investigators (Enwonwu et al., 1999; Falkler et al., 1999b; Horning, 1996) suggested that acute necrotizing gingivitis (ANG) is a precursor of noma in less developed countries. ANG is a gingivitis that begins with a gingival edema which rapidly evolves into necrosis of the gingival papilla. It is believed that, under certain unknown stress conditions and predispositions, ANG could evolve into noma if left untreated (Enwonwu et al., 1999).

Although the exact etiology is unknown, the proposed theory (Falkler et al., 1999a) suggests that multiple factors such as malnutrition, weakened immune functions and prior viral infections (measles, Herpesviridae), all worsened by poor oral hygiene, could play in unison to lower the host resistance and promote the development of oral lesions (e.g. ANG). The lesions would then serve as sites of entrance for a trigger microorganism which would initiate infection and favor conditions for polymicrobial growth and tissues destruction (see figure 11).

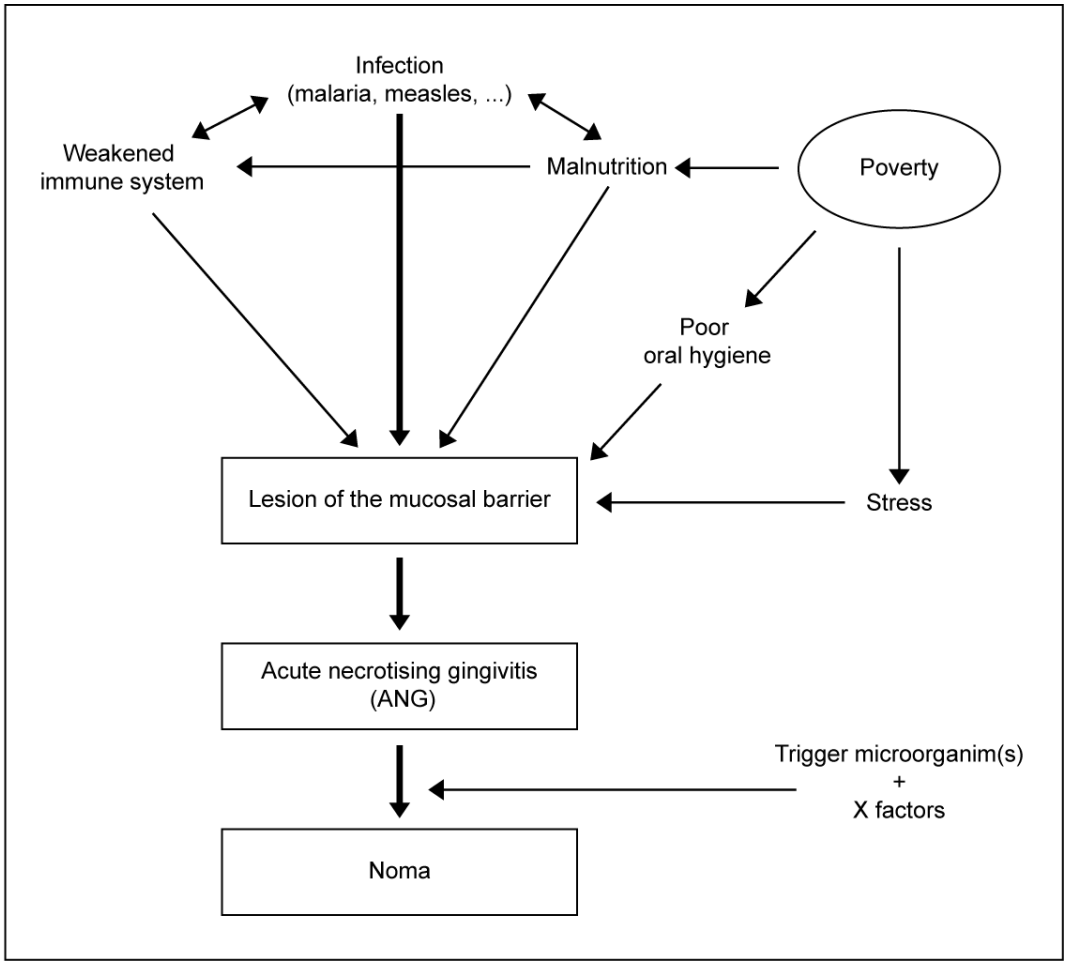


Fig. 11: Risks factors and proposed theory. Adapted from (Baratti-Mayer et al., 2003)

Classical microbiological identification methods allowed characterization of probable bacteria that could play a part in the development of noma such as: *Peptostreptococcus spp.* (Bolivar et al., 2009), *Prevotella intermedia* (Bolivar et al., 2009; Falkler et al., 1999a) and *Fusobacterium necrophorum* (Falkler et al., 1999a). Compared to healthy children, anaerobes – particularly *P. intermedia* – are more present in the oral cavity of malnourished children (Falkler et al., 1999b). Moreover, since advanced noma lesions are open to the environment, a

variety of phylotypes not commonly associated to the oral flora can be detected; for example: *Flavobacterium*, *Microbacterium*, *Sphingomonas*, *Bacillus*, *Paenibacillus*, and *Rhizobium* (Paster et al., 2002).

The necrotizing bacteria *Fusobacterium necrophorum* is involved in necrobacillosis in wallabies and the similarity of this disease with noma in humans is striking, resulting in the proposal that this species might be involved in the etiology of noma (Falkler et al., 1999a). *F. necrophorum* produces a variety of toxins, proteolytic enzymes and a growth-stimulating factor for *P. intermedia* (Price and McCallum, 1986). *P. intermedia* could add to tissue destruction by the production of lipid-degrading molecules and proteolytic enzymes (Slots et al., 1986).

1.3.2 Presentation of the GESNOMA initiative

The GESNOMA (Geneva study group of noma) group, supported by the Hirzel Foundation, conducts research on noma. Its goal is to understand the causes and mechanisms of this disease and to define appropriate treatments. This pluridisciplinary group initiated a large case-control study, including a sampling campaign in Niger followed by a complete analysis of the oral samples (serology and microbiology).

Bacteriological analysis of collected samples has been accomplished using three different culture-independent molecular techniques:

- *Cloning-sequencing*: 78 oral samples from Nigerien children were pooled in seven libraries according to child status (noma, ANG or healthy), child gender and site status (lesion or non-lesion site). Each library was then PCR-amplified using “universal” primers targeting the 16S rRNA gene. Amplified fragments were cloned and sequenced in an attempt to characterize the etiologic agent(s) of noma. This part was accomplished by Bolivar and colleagues from the Institut für Angewandte Immunologie (IAI) in Geneva.

- *Phylogenetic microarray*: Cloning-sequencing is considered as the gold-standard for bacterial identification but it is poorly adapted to the study of large sample collections and it also suffers from PCR primers bias. Indeed, the specificity of the so-called "universal" primers can be limited for certain taxa (Baker et al., 2003; Blackwood et al., 2005; Isenbarger et al., 2008) and 16S rRNA libraries are not necessarily representative of true prokaryotic diversity. The goal of the phylogenetic microarray (Huyghe et al., 2008) developed in our laboratory was to get rid of this primer bias by using *in vitro* transcription (IVT; enables amplification of total RNA using poly-A tailing and the T7 RNA-polymerase), and to observe a broader bacterial diversity (Brodie et al., 2007; DeSantis et al., 2007; Huyghe et al., 2008) in the collected oral samples. Using the Agilent platform, this unique approach allowed us to analyze the bacterial content of 84 samples.

- *Low-density 16S rRNA microarray*: In addition to the phylogenetic microarray, we designed a 16S rRNA array based on the most relevant sequences inventoried by the previously described cloning sequencing study (see above) of Bolivar *et al.* (Bolivar et al., 2009). The purpose of this array was to extrapolate the results obtained by this previous study on a larger scale. Low-density arrays appeared particularly appealing for this study as they enable monitoring of a small number of targets (335 probes for the Clondiag platform, up to 44'000 probes for the Agilent platform) at a limited cost (approximately 20 € / array for the Clondiag platform, and approx. 350 € / array for the Agilent platform, including reagents). The affordability of this platform allowed us to analyze 673 oral samples.

The study conducted by Bolivar *et al.* is achieved and the corresponding research article is currently under revision (Bolivar et al., 2009). A total of 757 samples were analyzed using these two microarray approaches and the research article depicting our results (in preparation (Huyghe et al., 2009)) is included in this manuscript (see Chapter 2).

1.3.3 Microarray designs

1.3.3.1 Low-density 16S rRNA microarray

Based on the most relevant sequences inventoried by the previously described cloning sequencing study by Bolivar *et al.* (Bolivar et al., 2009), as well as on the results of high-density microarrays, we designed a low-density 16S rRNA gene array (Huyghe et al., 2009) in order to extrapolate the observations of this study on a larger scale (in our study we analyzed 673 samples, split in 5 distinct groups: healthy controls (n = 443), noma lesion site (n = 83), noma non-lesion site (n = 82), gingivitis lesion site (n = 31), and gingivitis non-lesion site (n = 34)). In the study of Bolivar *et al.*, the dataset contains 1'237 partial 16S rRNA gene sequences representing 339 different phylotypes. On this dataset, we used an arbitrary cutoff of 1% to select the most abundant sequences detected in the libraries of Bolivar *et al.* in order to design one or more probes for each sequence (Clondiag platform, 335 probes available). Using this cutoff, the 132 most abundant 16S rRNA gene sequences were scanned for probes respecting peculiar physico-chemical properties (Tm = 65±5°C; probe length = 23–50 nt; < -5.0 kcal/mol for hairpins; < -8.0 kcal/mol for self-dimers; and dinucleotide repeats shorter than 5 bp) using a commercial software (Array Designer™ 2.0 by Premier Biosoft). Consequently, the resulting candidate probes were tested for specificity using an improved version of Olicheck (Charbonnier et al., 2005) with default parameters, leading to a final set of 271 probes. Additionally, probes specific to various Archaea (downloaded from probeBase (Loy et al., 2007)) and negative control probes were added to the probe set.

1.3.3.2 Phylogenetic microarray

The design of a "classical" identification microarray requires having at least one probe targeting a given gene, such as the 16S rRNA gene, of a particular organism while avoiding cross-hybridizations with other microorganisms (see chapter 1.2.2). With phylogenetic arrays,

the probe design process is less trivial. Indeed, probes must be sensitive enough to detect all bacteria present in a microbiota, even in low abundance, and specific to only one phylotype. The originality of phylogenetic arrays is that probes are specific to virtually all phylotypes; each probe targets a specific level of the phylogenetic tree (domain, phylum, class, order, family, genus, and species). Thus, this approach not only allows detection of known bacterial species, but also unsequenced species since the concept of 16S rRNA phylogenetic tree involves that bacterial species share evolutionary relationships and a common ancestor. Thus, an unknown bacterium would be detected and mapped to its corresponding node(s) in the phylogeny.

In-house software (written in Delphi) was developed (Huyghe et al., 2008) to produce the best probe set maximizing node coverage while respecting the number of available array features on the Agilent platform.

The program proceeds in four steps:

(1) *Parsing of 16S sequences.* The input of the program is a list of the 194,696 aligned bacterial small-subunit (SSU) rRNA gene sequences obtained from the Ribosomal Database Project (RDP; release 9.34) (Maidak et al., 2001). Additionally, a XML file from the same database that describes the 32 different taxonomic groups is parsed and loaded by the program.

(2) *Search of candidate probes*. The program scans each SSU rRNA gene sequence from its 5' to 3' end in order to extract all possible subsequences of length = 25. This length was selected as it provides an excellent specificity for single nucleotide mismatch (data not shown).

(3) Melting temperatures (*Tm*) were determined for each candidate probe by using thermodynamic parameters based on a nearest-neighbor model (SantaLucia, 1998). Candidate probes with a predicted *Tm* of 60 ± 5°C were stored in a hash table structure to eliminate duplicate sequences and to permit rapid data processing. This step yielded to 2'509'422 candidate probes.

(4) *Node assignment*. Consequently, the program assigns a node to each candidate probe, i.e. probes matching several SSU rRNA gene sequences were assigned to the highest taxonomic level common to the referred sequences. For example, a probe matching some representatives of both *Citrobacter* (genus) and *Escherichia* (genus) is considered specific of the parental taxonomic level, namely *Enterobacteriaceae* (family).

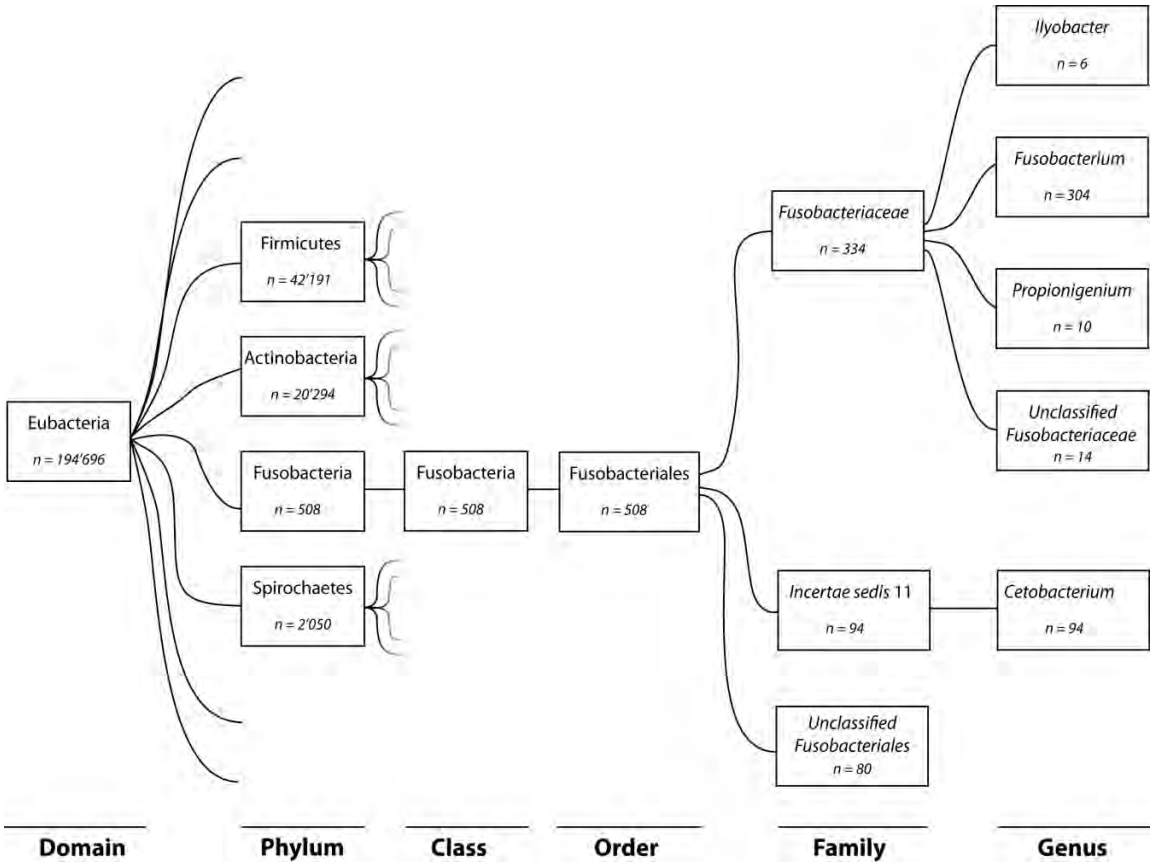


Fig. 12: Schematic representation of the RDP focusing on the Fusobacteria class. This figure demonstrates the complexity of the probe selection process where a minimum of probes must be found in order to attain the highest coverage for each node. *n* represents the total number of sequences assigned to a taxon.

(5) *Probe filtering*. Because the preceding steps produce a large number of candidate probes and because the number of features available on the Agilent platform is limited (10'263 probes per array was the maximum for Agilent custom arrays when the probes were designed), a filtering step is necessary. Consequently, probes maximizing node coverage are favored and probes covering less than 1.2% – determined empirically – of the sequences of a given phylogenetic node are discarded. After this step 8'195 probes describing different phylogenetic levels were selected.

(6) *Dealing with nucleotide degeneracy*. About 2.3% of the selected probes present at least one ambiguous nucleotide. Since most of these probes with degeneracy convey a

significant contribution to the phylogenetic coverage we decided to generate all possible combination of non-degenerated sequences. This step yielded a final probe set of 9’477 probes, resulting in a global coverage of 78.3% of the total number of 16S rRNA gene sequences. Subsequently, a poly-T tail of 35 nucleotides was added to enhance sensitivity (see 1.2.2.1).

Chapter 2

Results

2.1 Novel microarray design strategy to study complex bacterial communities

Article 1

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Mar. 2008, p. 1876–1885 Vol. 74, No. 6
0099-2240/08/$08.00+0 doi:10.1128/AEM.01722-07

Novel Microarray Design Strategy To Study Complex Bacterial Communities[▽]

Antoine Huyghe,[1]* Patrice Francois,[1] Yvan Charbonnier,[1] Manuela Tangomo-Bento,[1] Eve-Julie Bonetti,[1] Bruce J. Paster,[2] Ignacio Bolivar,[3] Denise Baratti-Mayer,[4] Didier Pittet,[5] Jacques Schrenzel,[1] and the Geneva Study Group on Noma (GESNOMA)

Aim: The goal of this article is to introduce and validate our phylogenetic array and introduce its applicability.

Outcome: In this article, we explain how our in-house probe design software selects probe sequences from the 16S rRNA gene sequences downloaded from the Ribosomal Database Project (RDP). The final set of 9'477 phylogenetic 25-mer probes covers 78.3% of rRNA gene sequences of the RDP (release 9.34; about 200'000 sequences). Node coverage at the phylum level, defined as the percentage of sequences covered by probes matching a given node, ranges from 21% to 100%. Using spiked samples, we show that our strategy is efficient in detecting differences that represent as low as 1% of the total

abundance in microbial mixtures. Subsequently, we compare our results to previous cloning-sequencing studies and observe a good concordance between the two techniques. Also, we show that our phylogenetic approach reveals a broader bacterial diversity than sequencing of the 16S rRNA gene. Finally, we explain that our strategy is well suited to compare bacterial flora, and more precisely the oral microbiota.

Author's contribution: AH developed the probe design software, performed the design and analysis of the microarray experiments. AH wrote the manuscript.

2.2 Microarray analysis of the microflora of noma lesions

Article 2

Microarray analysis of the microflora of noma lesions

Huyghe A, Francois P, Tangomo M, Girard M, Baratti-Mayer D, Bolivar I, Pittet D, Schrenzel J and the Geneva Study Group on Noma (GESNOMA).

Article in preparation

Aim: The goal of this second article is to compare the oral flora of noma children against the oral flora of gingivitis and healthy subjects using our two microarray strategies (see chapter 1.3.3).

Outcome: In this article we introduce our low-density 16S rRNA gene microarray designed on the basis of a previous cloning sequencing study (Bolivar et al., 2009). Using our phylogenetic and low-density microarrays, we confirm that *Peptostreptococcus* and Prevotellaceae are more abundant in noma lesions. In addition, we show that other oral pathogens such as *Atopobium* spp., *Streptococcus pyogenes* and *Streptococcus anginosus* are associated with advanced noma lesions. We also propose to reject the old observation that *Fusobacterium* is more abundant in noma lesions, reinforcing observations made in recent studies (Bolivar et al., 2009; Paster et al., 2002).

Finally, we show that noma lesions exhibit less bacterial diversity than acute necrotizing gingivitis (ANG) comforting previous hypothesis stipulating that ANG might precede acute noma.

Author's contribution: AH designed the two microarrays (16S rRNA microarray and phylogenetic microarray). AH analyzed the experiments and wrote the manuscript.

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Mar. 2008, p. 1876–1885 Vol. 74, No. 6
0099-2240/08/$08.00+0 doi:10.1128/AEM.01722-07

Novel Microarray Design Strategy To Study Complex Bacterial Communities▽

Antoine Huyghe,[1]* Patrice Francois,[1] Yvan Charbonnier,[1] Manuela Tangomo-Bento,[1] Eve-Julie Bonetti,[1] Bruce J. Paster,[2] Ignacio Bolivar,[3] Denise Baratti-Mayer,[4] Didier Pittet,[5] Jacques Schrenzel,[1] and the Geneva Study Group on Noma (GESNOMA)

Genomic Research Laboratory, Infectious Diseases Service, University of Geneva Hospitals, Geneva, Switzerland[1]; The Forsyth Institute, Boston, Massachusetts[2]; Institut für Angewandte Immunologie, Zuchwil, Switzerland[3]; Reconstructive and Plastic Surgery, Faculty of Medicine, Geneva, Switzerland[4]; and Infection Control Program, Faculty of Medicine, Geneva, Switzerland[5]

Received 26 July 2007/Accepted 10 January 2008

Assessing bacterial flora composition appears to be of increasing importance to fields as diverse as physiology, development, medicine, epidemiology, the environment, and the food industry. We report here the development and validation of an original microarray strategy that allows analysis of the phylogenic composition of complex bacterial mixtures. The microarray contains ~9,500 feature elements targeting 16S rRNA gene-specific regions. Probe design was performed by selecting oligonucleotide sequences specific to each node of the seven levels of the bacterial phylogenetic tree (domain, phylum, class, order, family, genus, and species). This approach, based on sequence information, allows analysis of the bacterial contents of complex bacterial mixtures to detect both known and unknown microorganisms. The presence of unknown organisms can be suspected and mapped on the phylogenetic tree, indicating where to refine analysis. Initial proof-of-concept experiments were performed on oral bacterial communities. Our results show that this hierarchical approach can reveal minor changes (≤1%) in gingival flora content when samples collected in individuals from similar geographical origins are compared.

Assessing bacterial flora composition is increasingly important in order to analyze progressive gut colonization after birth (18, 34), to unravel bacterial role in the epidemics of obesity in humans (44), or to better understand flora changes upon antibiotic selection pressure or in various disease states. The gingival flora was selected as a proof of concept for the present study. It contains a wide variety of microorganisms, either commensals or potential pathogens, but the vast majority remains simply uncultured or unknown. The complexity of this bacterial flora benefits from very diverse growth environments (O_2, CO_2, pH, accessibility to nutrients and epithelial debris, etc.), creating an array of anatomic niches for bacterial colonization. To illustrate this diversity, one estimates the number of bacterial species found in all oral niches to be approximately 500 to 700 (1, 35). Some of these bacteria are clearly associated with cavities, gingivitis, and periodontitis and might be implicated in less commonly encountered lesions such as noma (13, 14). Furthermore, particular bacterial species commonly found in the oral cavity can contribute to systemic diseases. Han et al. (20) showed that *Fusobacterium nucleatum*, associated with periodontal disease, can be implicated in preterm births. Similarly, the presence of *Porphyromonas gingivalis* has been documented to trigger inflammation in aortic endothelial cells (7, 42). In accordance with these observations, characterization of the bacterial oral flora is essential for the study of poorly understood diseases, such as noma, and for the development of novel diagnostic approaches.

A classical way to characterize members of complex bacterial communities relies on 16S rRNA gene sequence analysis. This target is particularly adapted to phylogenic studies since it contains highly conserved and variable moieties permitting reliable and detailed bacterial classification (12, 25, 37). In this approach, nucleic acids are directly extracted from samples without any prior cultivation; amplification is then performed using universal primers targeting conserved stretches of the 16S rRNA gene, and identification is based on similarity with sequences deposited in public ribosomal gene databases. Since the rate of nucleotide sequence change correlates with the evolutionary distance between organisms (8), sequence relatedness can be used for a phylogenic approach. To analyze complex bacterial flora, large-scale cloning and sequencing of 16S rRNA gene targets can provide a detailed catalogue of bacterial flora from a representative sample. However, despite its accuracy and potential for bacterial quantitation, this approach cannot be applied to compare large groups of samples exhibiting intra- and intervariability due to the very large number of experiments that would have to be performed.

Microarrays are frequently used to monitor gene regulation and expression on a genome-wide scale (6). Still, development of this technology for the comprehensive characterization of the content of complex microbial communities is still under way, and different approaches have been proposed. Functional gene arrays (49) target enzymes involved in a peculiar metabolic process such as methane oxidation (2), nitrogen fixation

* Corresponding author. Mailing address: Genomic Research Laboratory, Infectious Diseases Service, University of Geneva Hospitals, Micheli-du-Crest 24, 1211 Geneva 14, Geneva, Switzerland. Phone: 41(0)223729338. Fax: 41(0)223729830. E-mail: antoine.huyghe@genomic.ch.

▽ Published ahead of print on 18 January 2008.

(32, 43), and sulfur metabolism or iron metabolism (50). While allowing characterization of key bacteria involved in a determined metabolism pathway, this approach is also useful for monitoring the functional state of a microbial community under different environmental conditions. Alternatively, profiling of prokaryotic populations has been achieved by using 16S rRNA gene microarrays with probes targeting bacterial groups such as *Cyanobacteria* (5), *Rhodocyclales* (31), or *Alphaproteobacteria* (40). Development of high-density microarrays allowed extending the scope of phylogenetic oligoarrays to the whole bacterial kingdom (3, 33, 48). Although 16S rRNA microarrays do not appear to be optimal for discovering new taxa, this approach has permitted the detection of a broader bacterial diversity than the use of clone libraries (10).

The utilization of microarrays is appealing for evaluating samples containing complex flora. The design of oligonucleotide probes appears to be mandatory for resolving punctual sequence differences compared to PCR products that have been shown to exhibit poor performance for single-nucleotide polymorphism or punctual mutation analysis (19, 27, 29, 38). Moreover, oligonucleotide probes are more flexible and can be tailored to meet critical criteria such as sequence specificity and physicochemical properties.

To enable large-scale studies of complex bacterial flora composition in collections of samples, we developed an original oligonucleotide microarray design based on a phylogenic approach. Microarray design was performed by selecting ~9,500 25-nucleotide probes recognizing 16S rRNA gene targets that were specific to nodes matching the seven levels of the bacterial phylogenetic tree (domain, phylum, class, order, family, genus, and species). While providing information on the taxonomic composition of microbial communities, this approach should also prove useful for detecting uncharacterized species or to detect over- or under-representation of specific bacterial groups leading to imbalanced flora content. Our study shows that this hierarchical approach can reveal minor changes in microflora composition–as low as 1% of the global composition—when two complex, but related, bacterial populations are compared.

MATERIALS AND METHODS

Microarray design and manufacturing. In-house software was developed to produce the best probe set maximizing node coverage while respecting the number of available array features (unpublished data). The input to this program is a list of the 194,696 bacterial small-subunit (SSU) rRNA gene sequences classified by the Ribosomal Database Project (RDP; release 9.34) in 32 different taxonomic groups (9). The program then proceeds in four stages.

First, each SSU rRNA gene sequence was scanned from its 5′ to 3′ end in order to extract all possible 25-nucleotide subsequences as a pool of candidate probes. Melting temperatures (T_m) were determined for each candidate probe by using thermodynamic parameters based on a nearest-neighbor model (41). Candidate probes with a predicted T_m of 60 ± 5°C were stored in a hash table structure to eliminate duplicate sequences and to permit rapid data processing.

The next stage involved assigning a node to each candidate probe. Probes matching several SSU rRNA gene sequences were assigned to the nearest parent node common to the referred sequences. For example, a probe matching both *Streptococcus* and *Lactococcus* genera was assigned to the *Streptococcaceae* order (see 'probe C' in Fig. 1).

In the third stage we decided, due to the limited number of microarray features (10,263) and also due to the large number of candidate probes (2,509,422), to maximize node coverage by selecting only probes that provided substantial coverage (i.e., ≥1.2% coverage of all sequences of a given node, as determined empirically).

Stages 1 to 3 provided a set of 8,195 probes describing phylogenetic classes from the domain to the species level. About 2.3% of these probes presented one or more ambiguous nucleotides. Since most of these polymorphisms conveyed a significant contribution to phylogenic coverage, we decided to consider as a fourth stage all possible degenerated positions for this subset of targets, yielding a final probe set of 9,477 probes resulting in a global coverage of 78.3%. To minimize steric hindrance, all 25-mer probes were poly(T)-tailed to reach an overall length of 60 nucleotides. Microarrays were manufactured by in situ synthesis (Agilent Technologies, Palo Alto, CA).

Biological samples. Two distinct bacterial mixtures were used to validate our microarray approach. We first generated a defined artificial sample (sample A) using equal amounts of cRNA originating from three different organisms: *Streptococcus pyogenes* ATCC 12344, *Fusobacterium necrogenes* ATCC 25556, and *Chromobacterium violaceum* ATCC 12472. Artificial sample A (200 ng of total cRNA) was compared to the same bacterial mixture previously spiked with 25% (i.e., 50 ng of total RNA) *Escherichia coli* ATCC 25922, yielding to another 200-ng cRNA sample (spiked sample A).

Sterile endodontic paper points were used to collect gingival fluid from the dentogingival sulcus of two healthy European male subjects aged 28 and 41 years (samples B1 and B2, respectively). Samples were stored in RLT buffer (RNeasy Minikit; Qiagen, Basel, Switzerland) at −80°C for subsequent analyses. Then, 2 μg of cRNA of gingival samples B1 and B2 was then compared against their equivalents, but previously spiked with a lower concentration (1%) of *F. necrogenes* ATCC 25556 (spiked sample B1 and spiked sample B2, respectively).

RNA extraction and quantification. To lyse cells, 100 mg of glass beads (diameter, 100 μm; Schieritz & Hauenstein AG, Arlesheim, Switzerland) were added to the samples. Volume was adjusted to 350 μl with RLT buffer, and samples were vortex mixed for 1 min. Total RNA was isolated and purified by using the RNeasy Micro kit (Qiagen) according to the manufacturer's instructions. Samples were lyophilized and dissolved in 5 μl of sterile water. Total RNA quality was assessed by using RNA Picochips on a BioAnalyzer 2100 (Agilent). The RNA quantity was assessed by one-step reverse transcription-quantitative PCR using 0.2 μM concentrations of two primers (forward, GGCAAGCGTTA TCCGGAATT; reverse, GTTTCCAATGACCCTCCACG; Invitrogen, Basel, Switzerland) and a 0.1 μM concentration of probe (CCTACGCGCGCTTTAC GCCCA, 5′-end coupled to FAM and 3′-end coupled to TAMRA; Eurogentec, Seraing, Belgium) designed in a highly conserved region of bacterial 16S rRNA gene, allowing amplification of most of the bacterial 16S rRNA sequence. One-step reverse transcription-quantitative PCR amplification (final volume of 15 μl; Invitrogen) was performed on an SDS 7700 (PE Biosystems, Santa Clara, CA) using the following cycling procedure: t_1, 20 min at 50°C; t_2, 10 min at 94°C; t_3, 15 s at 94°C; and t_4, 1 min at 60°C (t_3 and t_4 were each repeated 40 times). Using this strategy, a positive fluorescent signal was obtained between cycles 21 and 30.

RNA amplification. All of the purified RNA was subjected to in vitro transcription using a MessageAmp II-Bacteria kit (Ambion, Austin, TX) according to the manufacturer's instructions. Amplified RNA was labeled during in vitro transcription in the presence of Cy3 or Cy5 cyanine dyes (Perkin-Elmer, Boston, MA). Quality, quantity, amplification efficiency, and dye incorporation were evaluated using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Inc., Rockland, DE) and a BioAnalyzer 2100 on RNA Nano 6000 chips (Agilent).

Microarray hybridization, scanning, and analysis. All samples were hybridized in duplicate. Cy5- and Cy3-labeled cRNAs were diluted in a total of 250 μl of Agilent hybridization buffer and hybridized at 60°C for 17 h in a dedicated hybridization oven (Robbins Scientific, Sunnyvale, CA). Slides were washed, dried under nitrogen flow, and scanned (Agilent) using 100% PMT power for both wavelengths.

Image analysis and signal quantification were achieved by using Feature Extraction software (version 6; Agilent). Probes exhibiting a nonuniform signal (i.e., pixel noise exceeding an established threshold) or mean signal values inferior to the corresponding background plus 2.6 standard deviations were excluded from subsequent analyses.

For spiking experiments, LOWESS (locally weighted linear regression) transformation was used to correct background subtracted signals for unequal dye incorporation, and a geometric mean was applied to average signals between duplicates. Statistical analysis consisted of a two-tailed Student t test with a P value tailored according to the relative spike abundance ($P < 0.01$ for sample A, 25% spiking; $P < 0.05$ for samples B1 and B2, 1% spiking).

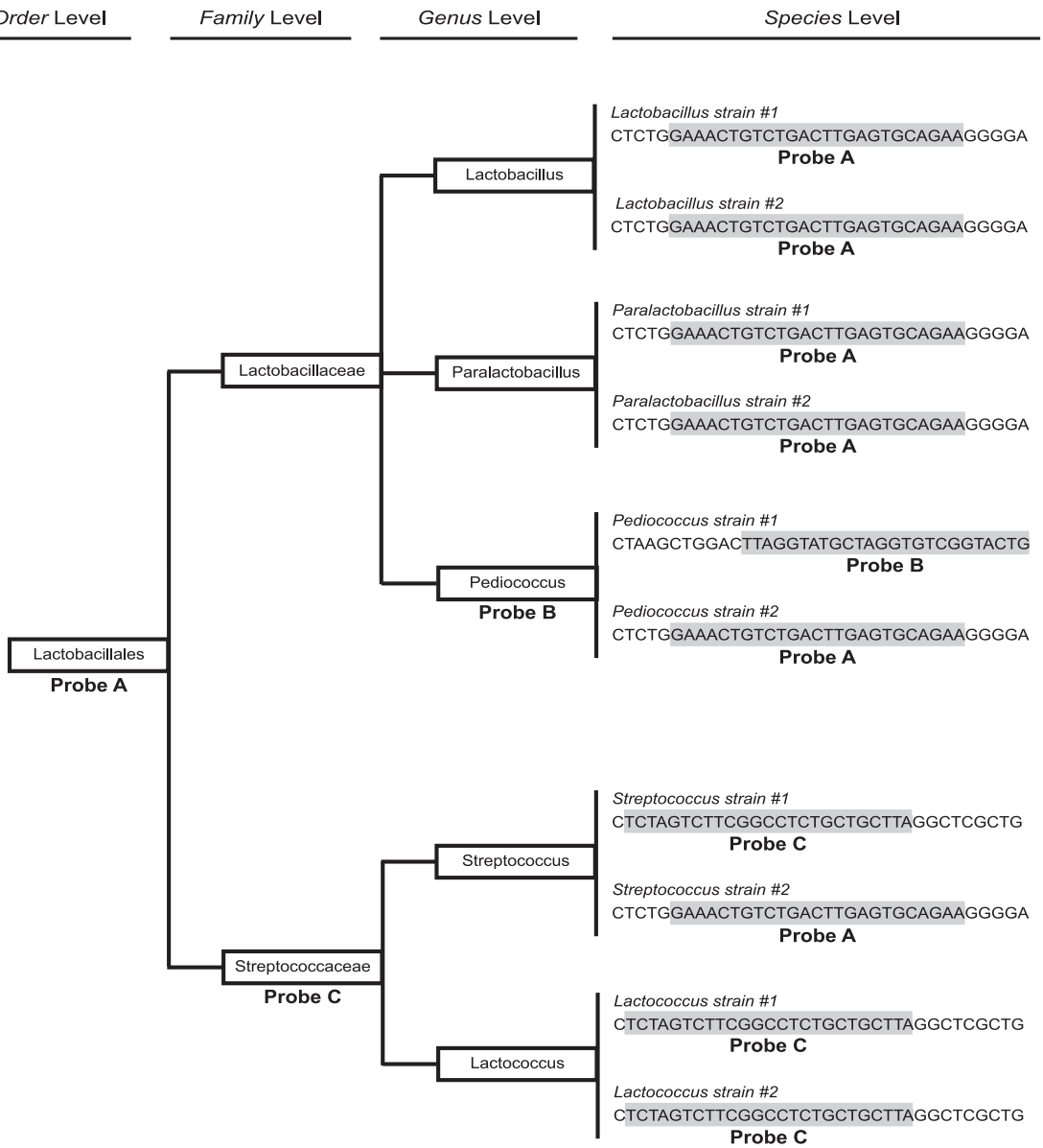


FIG. 1. Schematic representation of the probe selection process on a small subset of 16S rRNA gene sequences. Candidate probes are compared to the library of 16S rRNA gene sequences and assigned to the most distal common node. For example, probe A, which is common to the *Lactobacillus*, *Paralactobacillus*, *Pediococcus*, and *Streptococcus* genera, is assigned to the *Lactobacillales* order level. Probe C is assigned to the family level of the *Streptococcaceae* since it detects both *Streptococcus* and *Lactococcus* spp. In contrast, probe B is specific to some *Pediococcus* species, and it is therefore assigned to the genus level.

For assessing the bacterial subgingival flora on nonspiked samples, the local background was subtracted from the raw mean signal. Subsequently, probe signals were averaged among duplicates by using a geometric mean.

Comparison with 16S rRNA gene sequencing. To compare obtained results with data generated by large-scale cloning and sequencing approaches in different populations of healthy volunteers, we used published data of Kroes et al. (26) and Paster et al. (35). Briefly, Paster et al. used clone library sequencing to investigate the bacterial diversity in the subgingival plaque of healthy subjects and subjects affected by various kinds of periodontitis and acute necrotizing ulcerative gingivitis. From the five libraries matching healthy subjects, we chose three libraries showing the widest diversity in species and phylotypes in order to perform a direct comparison with our data. In the same way we used the data from the study of Kroes et al. describing the bacterial diversity found in a single healthy human volunteer, likewise using clone library sequencing.

RESULTS

Microarray design. The final set of 9,477 phylogenic 25-mer probes covers 78.3% of the 194,696 SSU rRNA gene sequences listed in the RDP database, i.e., 1 single probe matches an average of 16 sequences. Node coverage at the phylum level, defined as the percentage of sequences covered by probes matching a given node, averages to 78.2% and ranges from 21% for the *Chloroflexi* phylum to 100% for *Chrysiogenetes*, *Chlamydiae*, *Dictyoglomus*, and *Incertae sedis* BRC1 phyla (Fig. 2). Nodes or phyla that might be particularly important in the investigation of oral diseases of unknown etiol-

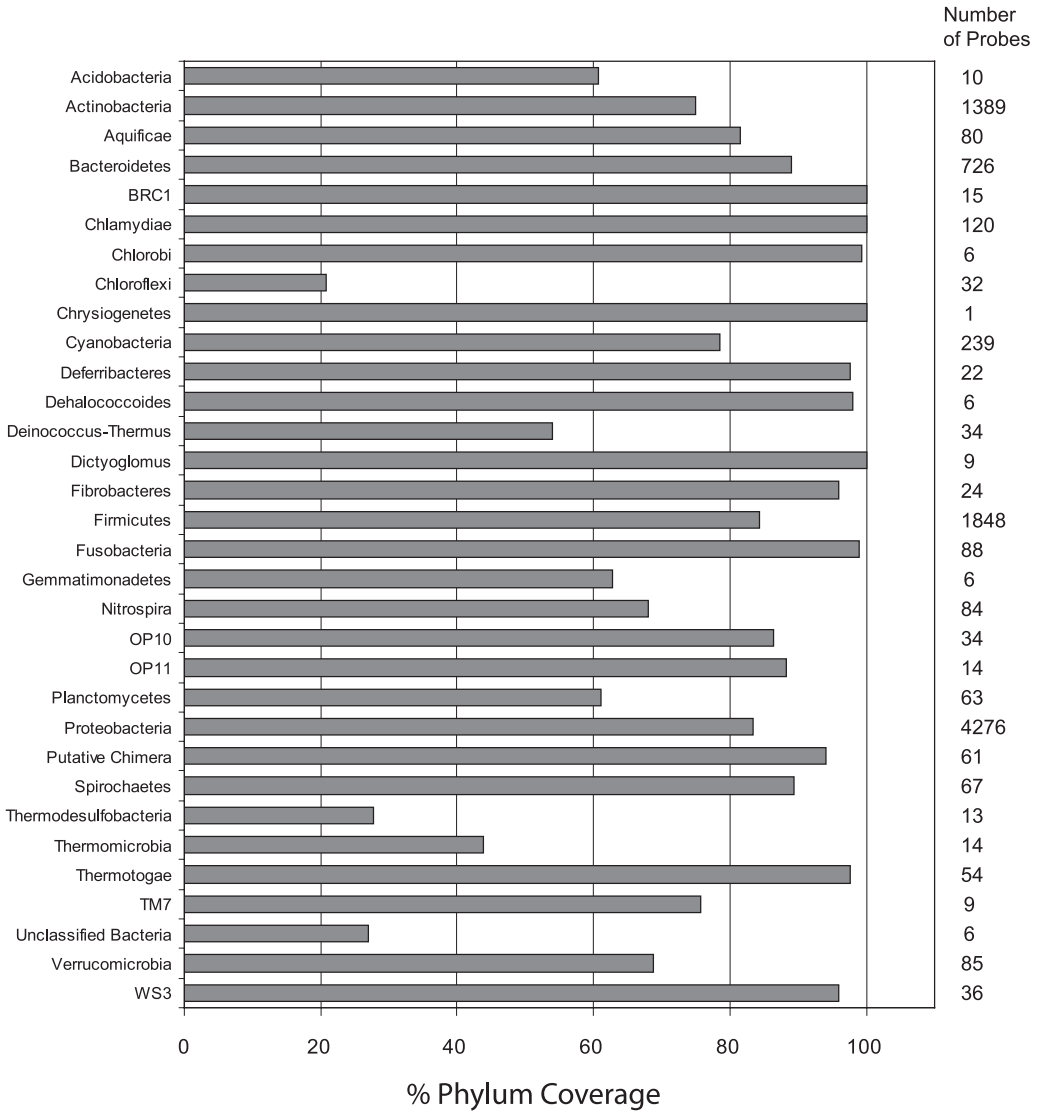
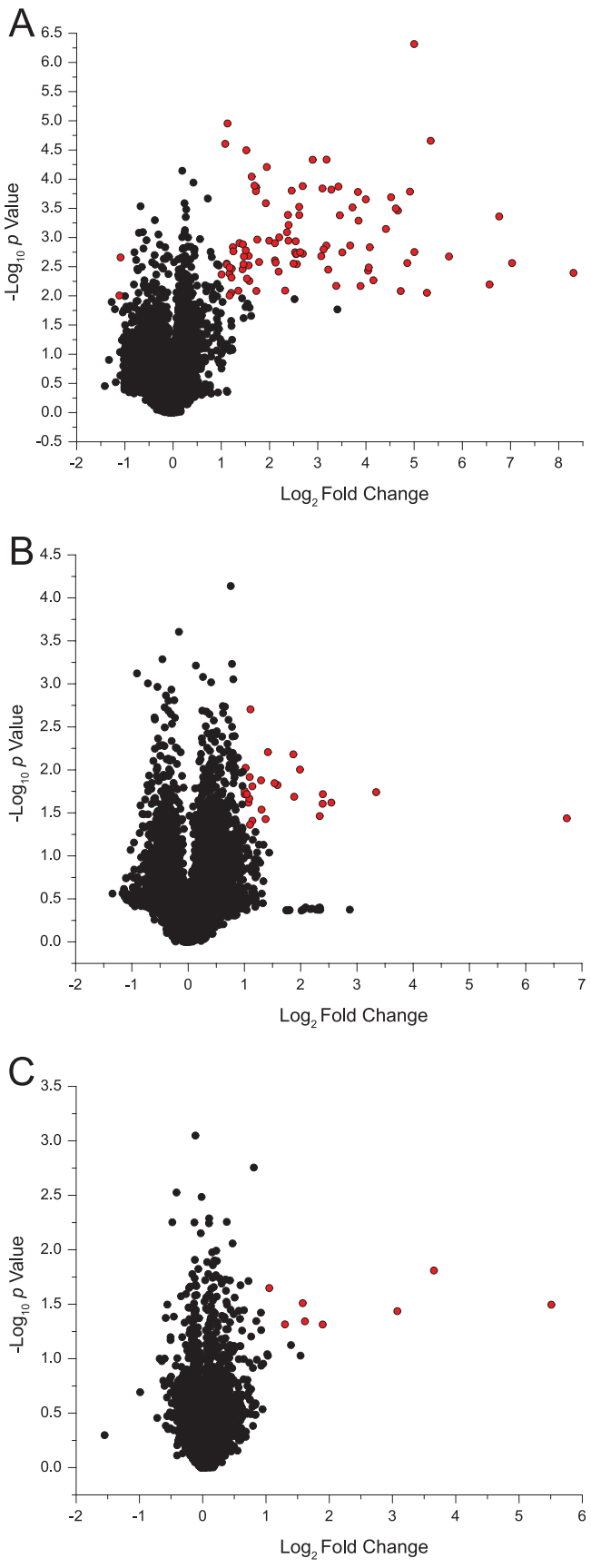


FIG. 2. Microarray coverage for the 32 phyla described in the RDP (see Materials and Methods). Bars represent the percentage of strains detected within each phylum. The number of selected probes for providing this phylum coverage is specified to the right of the figure.

ogy, such as the noma (14, 36), show appreciable coverage yields, e.g., the *Fusobacteria* (99%), *Spirochaetes* (89%), and *Bacteroidetes* (including *Prevotella* [89%]) phyla.

Bacterial detection from a defined mixture. To assess the reliability of our approach, we first attempted to detect differences in bacterial composition by comparing a relatively simple mixture of three distinct bacterial species with its equivalent, but supplemented with another bacterium representing 25% of the total mixture. By performing duplicate experiments, we checked whether the spiked species, *Escherichia coli*, could be reliably detected from this defined sample and whether it would induce unexpected changes in the identification scheme. Probes showing a fold change >2 and a *P* value of <0.01 were considered to represent targets differentially present between compared samples. Among 9,477 probes present on our microarray, 6,991 probes (74%) yielded signals above background values, but only 97 probes (1%) displayed statistical significance in spiked and nonspiked samples. The "volcano plot" depicted in Fig. 3A shows the *P* values of the probes producing a fluorescent signal plotted against their fold change. Analysis reveals that 95 probes yielded statistically significant signals in the spiked mixture (upper right side of panel A), while only 2 probes were detected as significant in the nonspiked sample (upper left side). In the spiked sample, 56 probes out of 95 (59%) matched genus and family nodes that represent the *Gammaproteobacteria* class. The most significant feature is a probe matching the *Escherichia* genus with a fold change of >300 ($P = 0.004$), which is the only probe at this level to match *E. coli* ATCC 25922. The second most relevant feature (fold change > 130, $P = 0.0027$) of this analysis is a probe matching the class of the *Gammaproteobacteria* that also includes the *Escherichia* genus. The third probe matching the *Alcalilimnicola* genus (*Gammaproteobacteria* class) shows a fold change of ~109 ($P = 0.0004$) and would not be expected, based on in silico predictions. Partial cross-hybridization could also explain why probes that do not directly match *E. coli* ATCC 25922, but



other nodes, such as *Shigella*, are revealed by our analysis. Note, however, that these two species displayed strongly homologous ribosomal gene sequences.

Detecting bacterial changes from complex samples. We then assessed the ability of our microarray to detect minor changes in the flora composition of clinical samples from healthy subjects. Nucleic acids were extracted from a complex bacterial flora (i.e., subgingival samples) obtained from two healthy volunteers. Nucleic acids were labeled, hybridized, and compared to the same samples spiked with 1% *F. necrogenes* ATCC 25556. *F. necrogenes* was selected as a bacterium that does not belong to the normal oral flora (data not shown) (i) to maximize the diversity of spiking material and (ii) to assess whether lower detection sensitivity could be achieved. Total amounts of cRNA loaded onto the microarray were adapted by using 2 μg (instead of 200 ng) and a much lower percentage of spiked material (1% instead of 25%). We also adjusted the analysis parameters to compare samples using a twofold change and a *P* value of 0.05 since the sample content is presumably highly similar. A total of 27 probes revealed statistical significance in the spiked sample B1, whereas only 8 probes were detected in the spiked sample B2. In both analyses, the large majority of probes with a fold change of >3 belonged to the *Fusobacterium* genus, the *Fusobacteriales* order, or other closely phylogenetically related nodes (Table 1). In both samples, probes matching the *Fusobacterium* genus showed the most significant fold changes: ~106-fold for sample B1 and ~45-fold factor for sample B2. In contrast, no probes yielded statistically significant signals in the nonspiked samples, as illustrated by the upper left corners of the volcano plots (Fig. 3B and C).

Figure 4 depicts probes that were detected as significant during this spiking experiment when we used sample B2. For better readability, they are plotted on the actual phylogenic tree trimmed from phyla where no significant signal was detected. Four probes (see Table 1, sample B2) adequately refer to the expected nodes in the phylogeny of the spiked bacterium, *F. necrogenes*. One other probe points to a node identified as "unclassified *Fusobacteriales*" and shows a moderate (3.7) increase in the fold change. This observation can be explained by the strong homology of this probe with the sequence of the spiked target which is highly related. Only two other probes appear as unrelated to this phylogeny and map environmental and gut flora SSU sequences belonging to the *Melittangium* and *Saprospira* genera (Table 1, sample B2). Interestingly, these two probes revealed very high homology—position 23 [GGAATCGCTAGTAATCGCAAAT(G/C)AG]

FIG. 3. Volcano plots display a summary of test statistics as a function of fluorescence intensity ratios (i.e., fold change). Each plot compares spiked versus nonspiked samples, using material of increasing bacterial complexity. Red dots in the upper left or upper right corners depict significant targets in the nonspiked sample or in the spiked sample, respectively. (A) Comparison of one defined bacterial mixture containing three bacterial species with or without a spike of 25% of the nucleic acid amounts (significance is defined as a fold change of ≥2 and $P \leq 0.01$). (B and C) Comparisons of gingival flora from two healthy volunteers (B1 [B] and B2 [C]) with or without a spike of 1% of the nucleic acid amounts (significance is defined as a fold change of ≥2 and $P \leq 0.05$).

TABLE 1. Nodes identified as statistically significant in spiked samples B1 and B2

Sample and node	Description	Rank	Fold change	*P*
Sample B1				
1.21.1.1.1.1	*Fusobacterium*	Genus	106.23	0.037
1.21.1.1	*Fusobacteriales*	Order	10.15	0.018
1.21.1.1.1.1	*Fusobacterium*	Genus	5.84	0.024
1.21.1.1.1.1	*Fusobacterium*	Genus	5.26	0.019
1.21.1.1.3	Unclassified *Fusobacteriales*	Unclassified	5.25	0.025
1.21.1.1.3	Unclassified *Fusobacteriales*	Unclassified	5.06	0.035
1.21.1.1.1.2	*Ilyobacter*	Genus	3.97	0.010
1.21.1.1.2.1	*Cetobacterium*	Genus	3.69	0.021
1.13.1	*Clostridia*	Class	3.66	0.007
1.21.1.1.1.3	*Propionigenium*	Genus	3.01	0.015
1.20.3.1.2.4	Unclassified *Saprospiraceae*	Unclassified	2.90	0.014
1.21.1.1.1.1	*Fusobacterium*	Genus	2.67	0.006
1.12.3.4.1.6	*Rhodanobacter*	Genus	2.60	0.037
1.20.2.1.3	Unclassified *Flavobacteriales*	Unclassified	2.47	0.029
1.12.3.8.1.8	Unclassified *Oceanospirillaceae*	Unclassified	2.46	0.013
1.12.3.11.1.8	*Psychromonas*	Genus	2.21	0.039
1.14.1.4.26.1	*Pseudonocardia*	Genus	2.21	0.016
1.21.1.1.3	Unclassified *Fusobacteriales*	Unclassified	2.16	0.002
1.12.2.6.1.8	*Sterolibacterium*	Genus	2.15	0.044
1.20.3.1.5.4	Unclassified *Crenotrichaceae*	Unclassified	2.14	0.012
1.21.1.1.3	Unclassified *Fusobacteriales*	Unclassified	2.13	0.022
1.12.3.4.1.1	*Xanthomonas*	Genus	2.11	0.024
1.12.3.14.1.8	Unclassified *Pasteurellaceae*	Unclassified	2.06	0.019
1.12.1.3.1.4	*Antarctobacter*	Genus	2.04	0.019
1.25.1	BRC1	Genus	2.03	0.009
1.22.1.1.1.3	Unclassified *Verrucomicrobiaceae*	Unclassified	2.01	0.018
1.14.1.2.1.1	*Rubrobacter*	Genus	2.01	0.019
Sample B2				
1.21.1.1.1.1	*Fusobacterium*	Genus	45.6	0.032
1.21.1.1	*Fusobacteriales*	Order	12.6	0.016
1.21.1.1.1.1	*Fusobacterium*	Genus	8.4	0.037
1.21.1.1.3	Unclassified *Fusobacteriales*	Unclassified	3.7	0.049
1.12.4.6.1.3	*Melittangium*	Genus	3.1	0.046
1.20.3.1.2.4	Unclassified *Saprospiraceae*	Unclassified	3.0	0.031
1.21.1.1.2.1	*Cetobacterium*	Genus	2.5	0.048
1.21.1.1.1.1	*Fusobacterium*	Genus	2.1	0.023

and positions 1, 9, 23, and 24 [(G/T)TGCGTCC(T/C)ATTAG CTAGATGG(TA/AG)A], respectively—to the sequence of the spiked organism.

Defining the bacterial gingival flora in healthy subjects using phylogenic microarray. We analyzed the bacterial composition of healthy subgingival samples. Analysis of sample B1 and B2 yielded in the identification of 31 and 114 probes, respectively, with a significant fluorescent level. A total of 19 probes overlapped between the two samples. Although all of these probes matched different phylogenetic levels, a large majority matched at the genus level, such as *Actinomyces*, *Bacteroides*, *Capnocytophaga*, *Gemella*, *Porphyromonas*, *Prevotella*, *Rickettsiella*, *Streptococcus*, and TM7, as reported in previous studies (1, 26, 30, 35). At the phylum level, both samples displayed approximately the same flora profile (Fig. 5) encompassing *Proteobacteria*, *Firmicutes*, *Actinobacteria*, and *Bacteroidetes*. Interestingly, the *Bacteroidetes* phylum appeared to be better represented in sample B1 than in sample B2. Some rare and unexpected phyla were detected with our microarray approach, such as *Aquificae* and *Planctomycetes*. This part of the flora constituted a small proportion of the total bacterial content and was not identified in previous studies as part of the gingival flora but was mainly found in environmental samples (11, 24, 47).

We finally sought to compare the results of our approach to those of previously published studies (26, 35). We should emphasize here that these studies on the gingival flora were markedly different since volunteers were recruited in another continent (i.e., from the United States) and that the flora determination was performed by conventional generic amplification of 16S rRNA genes, followed by a cloning and sequencing strategy. However, despite these caveats, we observed remarkably similar compositions and abundances across phyla (Fig. 5), strongly suggesting a rather common composition in the gingival flora between these subjects at the phylum level. However, our data suggest that additional phyla were part of the normal oral flora.

DISCUSSION

Based on an original hierarchical phylogenic design, the ~9,500 probe set on our oligoarray covers 78.3% of the 194,696 bacterial SSU rRNA gene sequences described in release 9.34 of RDP. Although phyla coverage ranges from 21%

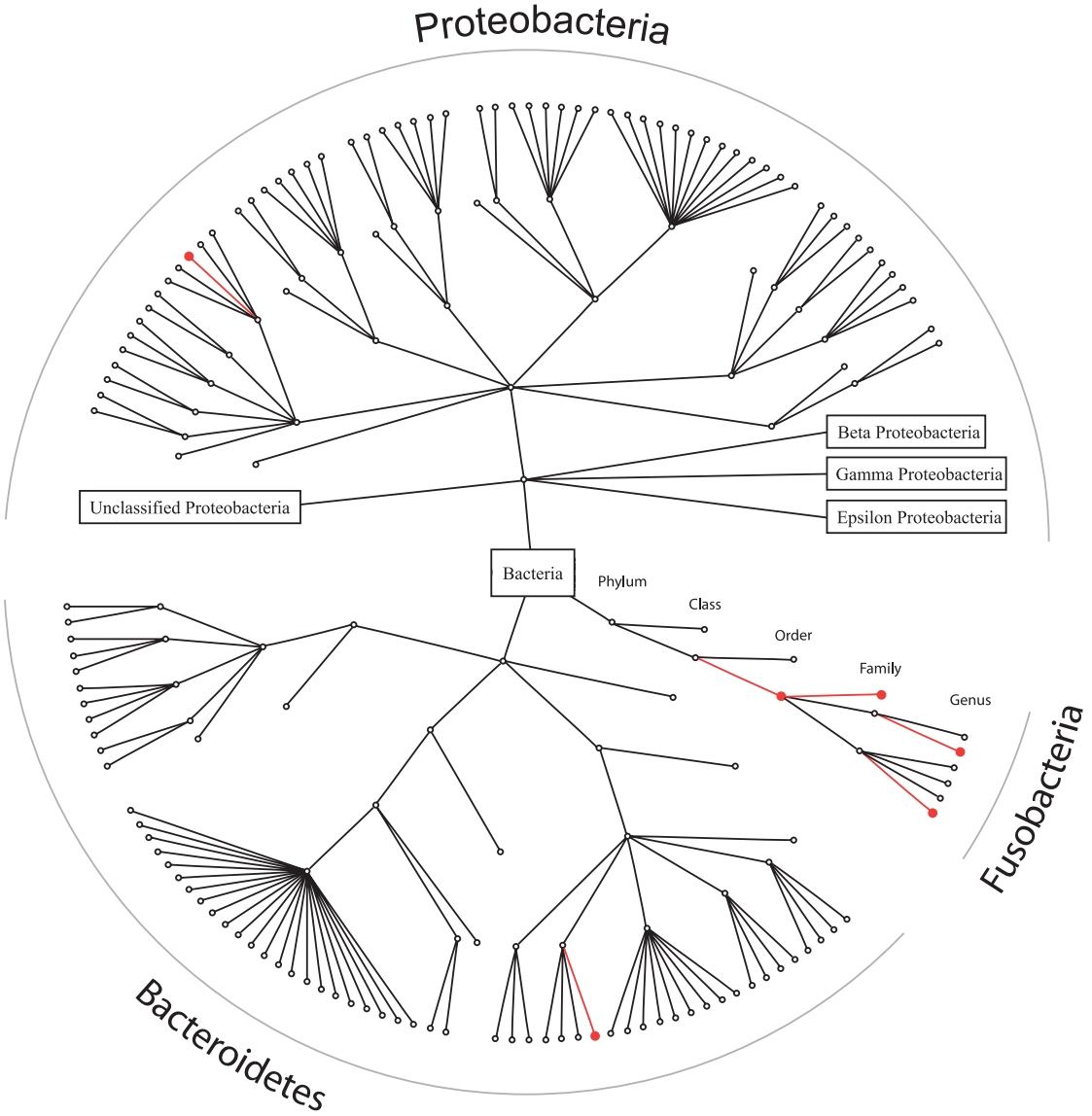


FIG. 4. Phylogenic tree representation of all phyla detected by statistical analysis ($P < 0.05$) comparing a gingival sample (B2) to its replicate spiked with 1% *F. necrogenes* ATCC 25556. Red dots and lines depict nodes and branches where at least 1 probe yielded a statistically significant signal. For better readability, only the *Alphaproteobacteria* class is fully represented among the *Proteobacteria* phylum.

(*Chloroflexi*) to 100% (e.g., *Chlamydiae*), the design process was not intended to be limited to the study of specific phyla and thus should prove useful for studying the whole eubacterial domain, which was the starting point of our strategy. This last point is of crucial importance since no specific phylum should have a priori more weight during the profiling of any altered oral flora.

Potential applications of such a microarray consist of detecting bacterial composition changes over time or across many related samples (e.g., healthy versus diseased or site for the study of noma, for example).

In the present study, the limited sensitivity of microarray techniques required amplification of the starting nucleic acids. In addition, the amplification strategy proposed here appears to be robust and is potentially utilizable for other samples for which the amounts of starting material are strictly limited. Our strategy proved to be efficient in detecting minor differences (as low as 1% of *F. necrogenes*) in flora composition in controlled mixtures and, more importantly, in complex natural samples. Our approach was able to characterize the diversity of the gingival flora down to the genus level in two healthy European subjects. The results showed good congruence with previous studies performed in American subjects (26, 35) using a cloning-sequencing strategy. This observation is noteworthy because this comparison involved volunteers originating from two distinct continents and using two markedly different methods. Our results suggest that different social and dietary habits have limited influence on gingival flora composition, as defined by our microarray strategy. This microarray approach revealed a broader diversity of microorganisms at the genus level than traditional clone libraries methods, a finding that has already been reported by DeSantis et al. (10). The underestimation of

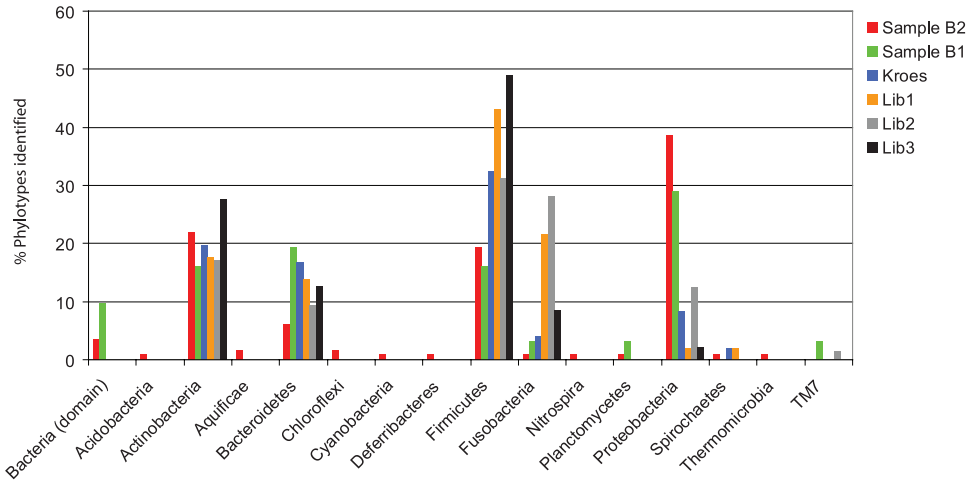


FIG. 5. Cumulative prevalence of phylotypes identified using microarrays (samples B1 and B2) compared to previously published analyses of gingival flora. The figure depicts all phylotypes identified by any of these studies. Lib1, Lib2, and Lib3 describe libraries generated for three healthy subjects as described by Paster et al. (35; B. Paster, unpublished data).

microbial diversity may be explained either by a cloning bias or by the paucity of clones sequenced. Moreover, since our strategy is based on 16S rRNA sequences, we can hypothesize an additional bias due to sequence over-representation in public sequence databases (23). This bias can be expected on important human or animal pathogens, as well as organisms of economic interest.

Nonetheless, this phylogenetic approach not only allows the monitoring of bacterial population dynamics in different ecosystems but also permits the detection of potential pathogenic bacteria or bacterial taxons, as recently illustrated in a study of airborne bacterial composition (4). In addition, the same strategy, as well as functional gene arrays, can be implemented to evaluate temporal evolution of complex bacterial communities such as in agricultural, environmental, or human commensal ecosystems. Such applications include evaluating the impact of ecosystem modification on environmental flora (3) or the identification of a specific bacterial population showing particular metabolic capacities, such as the metabolism of toxic compounds (21, 39). In human medicine, a potential use might be for monitoring flora composition alteration or bias composition, due to chemical, antimicrobial treatments (15) or due to metabolic dysfunction, or for understanding its natural evolution during the life cycle (34). This type of strategy represents serious advantages over culture or even other non-culture-based methodologies (22). The characteristics of sensitivity and specificity, as well as the actual throughput, are now compatible with real-time monitoring or analysis (17).

We have found that the efficiency of our microarray can be improved in several ways. We could rely on novel 16S sequences retrieved from the latest release of RDP, as well as data emerging from various metagenomic projects (16, 45). In addition, it would be useful to add probes targeting the *Archaebacteria* domain, since methanogenic *Archaea* have been recently reported in subgingival sites of patients suffering from periodontitis (28). Obviously, our design is dependent on the quality of the sequences retrieved by RDP. In this regard, short 16S rRNA gene sequences may be prone to misclassification at the genus level due to anomalies in the taxonomy and that lack of data on short sequences (46). In addition, close sequence homology across different genera (e.g., *Escherichia* and *Shigella*) explains why specific probes cannot always be designed. The potential for cross-hybridization by the probes should not be minimized, especially in the highly conserved 16S rRNA gene sequences. Whereas careful assessment of such cross-hybridization potential appears warranted for the "de novo" characterization of a microbial community, this is clearly less of a concern here because our approach was developed and tested to detect differences between related samples.

Finally, the arbitrary twofold change cutoff was selected as an empirical compromise between sensitivity and specificity. Further experimental determinations are now warranted to more precisely define this cutoff value and determine whether it can be applied to the whole range of fluorescence signals. It remains to be proven whether statistical approaches might prove more reliable in the long run.

Preliminary steps to the identification of uncharacterized bacterial species can be performed using our hierarchical approach, providing an alternative to classical sequencing techniques. However, this point remains to be experimentally proven. Initially intended for the study of bacterial diversity in oral samples, our approach may be useful for the study of various bacterial communities associated either with medical (intestinal or skin flora), environmental (soil, sludge, wastewaters, etc.), or food industry samples. In addition, the same approach could be used to monitor the evolution of bacterial communities over time, replacing laborious techniques such as library cloning and sequencing.

This approach can be conveniently implemented to perform large-scale profiling studies of the oral bacterial flora. Experiments are under way to monitor gingival flora in noma lesions from individual patients and compare these samples to matched healthy controls from the same geographical origin. Again, implementing such novel microarray strategies for flora composition analyses require careful validation.

ACKNOWLEDGMENTS

This study was supported by the Geneva Study Group on Noma (GESNOMA), a research group funded by the Hirzel Foundation. P.F., J.S., and B.J.P. are supported by grants from the Swiss National Science Foundation (3100A0-112370/1) and COST (C05.0103 [J.S.] and 3100A0-116075/1 [P.F.]) and from the National Institutes of Health (DE-11443 [B.J.P.]).

REFERENCES

1. **Aas, J. A., B. J. Paster, L. N. Stokes, I. Olsen, and F. E. Dewhirst.** 2005. Defining the normal bacterial flora of the oral cavity. J. Clin. Microbiol. **43:**5721–5732.
2. **Bodrossy, L., N. Stralis-Pavese, J. C. Murrell, S. Radajewski, A. Weilharter, and A. Sessitsch.** 2003. Development and validation of a diagnostic microbial microarray for methanotrophs. Environ. Microbiol. **5:**566–582.
3. **Brodie, E. L., T. Z. DeSantis, D. C. Joyner, S. M. Baek, J. T. Larsen, G. L. Andersen, T. C. Hazen, P. M. Richardson, D. J. Herman, T. K. Tokunaga, J. M. Wan, and M. K. Firestone.** 2006. Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. Appl. Environ. Microbiol. **72:**6288–6298.
4. **Brodie, E. L., T. Z. DeSantis, J. P. M. Parker, I. X. Zubietta, Y. M. Piceno, and G. L. Andersen.** 2007. Urban aerosols harbor diverse and dynamic bacterial populations. Proc. Natl. Acad. Sci. USA **104:**299–304.
5. **Castiglioni, B., E. Rizzi, A. Frosini, K. Sivonen, P. Rajaniemi, A. Rantala, M. A. Mugnai, S. Ventura, A. Wilmotte, C. Boutte, S. Grubisic, P. Balthasart, C. Consolandi, R. Bordoni, A. Mezzelani, C. Battaglia, and G. De Bellis.** 2004. Development of a universal microarray based on the ligation detection reaction and 16S *rrnA* gene polymorphism to target diversity of cyanobacteria. Appl. Environ. Microbiol. **70:**7161–7172.
6. **Charbonnier, Y., B. Gettler, P. Francois, M. Bento, A. Renzoni, P. Vaudaux, W. Schlegel, and J. Schrenzel.** 2005. A generic approach for the design of whole-genome oligoarrays, validated for genomotyping, deletion mapping and gene expression analysis on *Staphylococcus aureus* 3. BMC Genomics **6:**95.
7. **Chou, H. H., H. Yumoto, M. Davey, Y. Takahashi, T. Miyamoto, F. C. Gibson III, and C. A. Genco.** 2005. *Porphyromonas gingivalis* fimbria-dependent activation of inflammatory genes in human aortic endothelial cells. Infect. Immun. **73:**5367–5378.
8. **Clarridge, J. E., III.** 2004. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin. Microbiol. Rev. **17:**840–862.
9. **Cole, J. R., B. Chai, R. J. Farris, Q. Wang, S. A. Kulam, D. M. McGarrell, G. M. Garrity, and J. M. Tiedje.** 2005. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Res. **33:**D294–D296.
10. **DeSantis, T. Z., E. L. Brodie, J. P. Moberg, I. X. Zubieta, Y. M. Piceno, and G. L. Andersen.** 2007. High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. Microb. Ecol. **53:**371–383.
11. **Elshahed, M. S., N. H. Youssef, Q. Luo, F. Z. Najar, B. A. Roe, T. M. Sisk, S. I. Buhring, K. U. Hinrichs, and L. R. Krumholz.** 2007. Phylogenetic and metabolic diversity of *Planctomycetes* from anaerobic, sulfide- and sulfur-rich Zodletone Spring, Oklahoma. Appl. Environ. Microbiol. **73:**4707–4716.
12. **Elsholz, B., R. Worl, L. Blohm, J. Albers, H. Feucht, T. Grunwald, B. Jurgen, T. Schweder, and R. Hintsche.** 2006. Automated detection and quantitation of bacterial RNA by using electrical microarrays. Anal. Chem. **78:**4794–4802.
13. **Enwonwu, C. O., W. A. Falkler, Jr., E. O. Idigbe, B. M. Afolabi, M. Ibrahim, D. Onwujekwe, O. Savage, and V. I. Meeks.** 1999. Pathogenesis of cancrum oris (noma): confounding interactions of malnutrition with infection. Am. J. Trop. Med. Hyg. **60:**223–232.
14. **Falkler, W. A., Jr., C. O. Enwonwu, and E. O. Idigbe.** 1999. Isolation of *Fusobacterium necrophorum* from cancrum oris (noma). Am. J. Trop. Med. Hyg. **60:**150–156.
15. **Flanagan, J. L., E. L. Brodie, L. Weng, S. V. Lynch, O. Garcia, R. Brown, P. Hugenholtz, T. Z. DeSantis, G. L. Andersen, J. P. Wiener-Kronish, and J. Bristow.** 2007. Loss of bacterial diversity during antibiotic treatment of intubated patients colonized with *Pseudomonas aeruginosa*. J. Clin. Microbiol. **45:**1954–1962.
16. **Frank, D. N., A. L. St. Amand, R. A. Feldman, E. C. Boedeker, N. Harpaz, and N. R. Pace.** 2007. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc. Natl. Acad. Sci. USA **104:**13780–13785.
17. **Gentry, T. J., G. S. Wickham, C. W. Schadt, Z. He, and J. Zhou.** 2006. Microarray applications in microbial ecology research. Microb. Ecol. **52:**159–175.
18. **Gill, S. R., M. Pop, R. T. Deboy, P. B. Eckburg, P. J. Turnbaugh, B. S. Samuel, J. I. Gordon, D. A. Relman, C. M. Fraser-Liggett, and K. E. Nelson.** 2006. Metagenomic analysis of the human distal gut microbiome. Science **312:**1355–1359.
19. **Hacia, J. G.** 1999. Resequencing and mutational analysis using oligonucleotide microarrays. Nat. Genet. **21:**42–47.
20. **Han, Y. W., R. W. Redline, M. Li, L. Yin, G. B. Hill, and T. S. McCormick.** 2004. *Fusobacterium nucleatum* induces premature and term stillbirths in pregnant mice: implication of oral bacteria in preterm birth. Infect. Immun. **72:**2272–2279.
21. **He, Z., T. J. Gentry, C. W. Schadt, L. Wu, J. Liebich, S. C. Chong, Z. Huang, W. Wu, B. Gu, P. Jardine, C. Criddle, and J. Zhou.** 2007. GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. ISME J. **1:**67–77.
22. **Hugenholtz, P., B. M. Goebel, and N. R. Pace.** 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol. **180:**4765–4774.
23. **Hugenholtz, P.** 2002. Exploring prokaryotic diversity in the genomic era. Genome Biol. **3:**REVIEWS0003.
24. **Hugler, M., H. Huber, S. J. Molyneaux, C. Vetriani, and S. M. Sievert.** 2007. Autotrophic CO_2 fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum *Aquificae*: evidence for two ways of citrate cleavage. Environ. Microbiol. **9:**81–92.
25. **Jordan, J. A., A. R. Butchko, and M. B. Durso.** 2005. Use of pyrosequencing of 16S rRNA fragments to differentiate between bacteria responsible for neonatal sepsis. J. Mol. Diagn. **7:**105–110.
26. **Kroes, I., P. W. Lepp, and D. A. Relman.** 1999. Bacterial diversity within the human subgingival crevice. Proc. Natl. Acad. Sci. USA **96:**14547–14552.
27. **LaForge, K. S., V. Shick, R. Spangler, D. Proudnikov, V. Yuferov, Y. Lysov, A. Mirzabekov, and M. J. Kreek.** 2000. Detection of single nucleotide polymorphisms of the human mu opioid receptor gene by hybridization or single nucleotide extension on custom oligonucleotide Gelpad microchips: potential in studies of addiction. Am. J. Med. Genet. **96:**604–615.
28. **Lepp, P. W., M. M. Brinig, C. C. Ouverney, K. Palm, G. C. Armitage, and D. A. Relman.** 2004. Methanogenic *Archaea* and human periodontal disease. Proc. Natl. Acad. Sci. USA **101:**6176–6181.
29. **Li, J., M. Pankratz, and J. A. Johnson.** 2002. Differential gene expression patterns revealed by oligonucleotide versus long cDNA arrays. Toxicol. Sci. **69:**383–390.
30. **Lillo, A., F. P. Ashley, R. M. Palmer, M. A. Munson, L. Kyriacou, A. J. Weightman, and W. G. Wade.** 2006. Novel subgingival bacterial phylotypes detected using multiple universal polymerase chain reaction primer sets. Oral Microbiol. Immunol. **21:**61–68.
31. **Loy, A., C. Schulz, S. Lücker, A. Schöpfer-Wendels, K. Stoecker, C. Baranyi, A. Lehner, and M. Wagner.** 2005. 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "*Rhodocyclales*." Appl. Environ. Microbiol. **71:**1373–1386.
32. **Moisander, P. H., L. Shiue, G. F. Steward, B. D. Jenkins, B. M. Bebout, and J. P. Zehr.** 2006. Application of a *nifH* oligonucleotide microarray for profiling diversity of N_2-fixing microorganisms in marine microbial mats. Environ. Microbiol. **8:**1721–1735.
33. **Palmer, C., E. M. Bik, M. B. Eisen, P. B. Eckburg, T. R. Sana, P. K. Wolber, D. A. Relman, and P. O. Brown.** 2006. Rapid quantitative profiling of complex microbial populations. Nucleic Acids Res. **34:**e5.
34. **Palmer, C., E. M. Bik, D. B. Digiulio, D. A. Relman, and P. O. Brown.** 2007. Development of the human infant intestinal microbiota. PLoS Biol. **5:**e177.
35. **Paster, B. J., S. K. Boches, J. L. Galvin, R. E. Ericson, C. N. Lau, V. A. Levanos, A. Sahasrabudhe, and F. E. Dewhirst.** 2001. Bacterial diversity in human subgingival plaque. J. Bacteriol. **183:**3770–3783.
36. **Paster, B. J., J. W. Falkler, Jr., C. O. Enwonwu, E. O. Idigbe, K. O. Savage, V. A. Levanos, M. A. Tamer, R. L. Ericson, C. N. Lau, and F. E. Dewhirst.** 2002. Prevalent bacterial species and novel phylotypes in advanced noma lesions. J. Clin. Microbiol. **40:**2187–2191.
37. **Qian, Q., Y. W. Tang, C. P. Kolbert, C. A. Torgerson, J. G. Hughes, E. A. Vetter, W. S. Harmsen, S. O. Montgomery, F. R. Cockerill III, and D. H. Persing.** 2001. Direct identification of bacteria from positive blood cultures by amplification and sequencing of the 16S rRNA gene: evaluation of BACTEC 9240 instrument true-positive and false-positive results. J. Clin. Microbiol. **39:**3578–3582.
38. **Relogio, A., C. Schwager, A. Richter, W. Ansorge, and J. Valcarcel.** 2002. Optimization of oligonucleotide-based DNA microarrays. Nucleic Acids Res. **30:**e51.
39. **Rhee, S. K., X. Liu, L. Wu, S. C. Chong, X. Wan, and J. Zhou.** 2004. Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. Appl. Environ. Microbiol. **70:**4303–4317.
40. **Sanguin, H., A. Herrera, C. Oger-Desfeux, A. Dechesne, P. Simonet, E. Navarro, T. M. Vogel, Y. Moënne-Loccoz, X. Nesme, and G. L. Grundmann.** 2006. Development and validation of a prototype 16S rRNA-based taxonomic microarray for *Alphaproteobacteria*. Environ. Microbiol. **8:**289–307.
41. **Santalucia, J., Jr.** 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. USA **95:**1460–1465.
42. **Takahashi, Y., M. Davey, H. Yumoto, F. C. Gibson III, and C. A. Genco.** 2006. Fimbria-dependent activation of pro-inflammatory molecules in *Porphyromonas gingivalis*-infected human aortic endothelial cells. Cell. Microbiol. **8:**738–757.
43. **Tiquia, S. M., L. Wu, S. C. Chong, S. Passovets, D. Xu, Y. Xu, and J. Zhou.**

2004. Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. BioTechniques **36:**664–670. 672:674.

44. **Turnbaugh, P. J., R. E. Ley, M. A. Mahowald, V. Magrini, E. R. Mardis, and J. I. Gordon.** 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature **444:**1027–1031.
45. **Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith.** 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304:**66–74.
46. **Wang, Q., G. M. Garrity, J. M. Tiedje, and J. R. Cole.** 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. **73:**5261–5267.
47. **Webster, N. S., and D. Bourne.** 2007. Bacterial community structure associated with the Antarctic soft coral, *Alcyonium antarcticum*. FEMS Microbiol. Ecol. **59:**81–94.
48. **Wilson, K. H., W. J. Wilson, J. L. Radosevich, T. Z. DeSantis, V. S. Viswanathan, T. A. Kuczmarski, and G. L. Andersen.** 2002. High-density microarray of small-subunit ribosomal DNA probes. Appl. Environ. Microbiol. **68:**2535–2541.
49. **Wu, L., D. K. Thompson, G. Li, R. A. Hurt, J. M. Tiedje, and J. Zhou.** 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. Appl. Environ. Microbiol. **67:**5780–5790.
50. **Yin, H., L. Cao, G. Qiu, D. Wang, L. Kellogg, J. Zhou, Z. Dai, and X. Liu.** 2007. Development and evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in acid mine drainages and bioleaching systems. J. Microbiol. Methods **70:**165–178.

Microarray analysis of the microbiota of noma lesions

Antoine Huyghe[1,2], Patrice François[1], Andrea Mombelli[3], Manuela Tangomo[1], Myriam Girard[1], Denise Baratti-Mayer[4], Ignacio Bolivar[5], Didier Pittet[6], Jacques Schrenzel[1] and the Geneva Study Group on Noma (GESNOMA).

[1] Genomic Research Laboratory. Infectious Diseases Service, University of Geneva Hospitals, Switzerland; [2] Swiss Institute of Bioinformatics, Michel-Servet 1, 1211 Geneva 4, Geneva, Switzerland; [3] Department of Periodontology, School of Dental Medicine, University of Geneva, Geneva, Switzerland; [4] Reconstructive and Plastic Surgery, Faculty of Medicine, Geneva, Switzerland; [5] Institut für Angewandte Immunologie, Geneva, Switzerland; [6] Infection Control Program, Faculty of Medicine, University of Geneva, Geneva, Switzerland.

Running title: Microarray analysis of the microbiota of noma lesions

Keywords: microarray, gingival samples, noma, 16S rRNA gene, microflora, metagenomics.

Corresponding author: Patrice François PhD, Genomic Research Laboratory. Infectious Diseases Service, University of Geneva Hospitals, Gabrielle-Perret-Gentil 4, 1211 Geneva 14, Geneva, Switzerland. Ph:+41(0)223729815, Fax:+41(0)223729830.

email:patrice.francois@genomic.ch

Abstract

Noma (cancrum oris) is a gangrenous disease of unknown etiology affecting the maxillo-facial region of young children in less developed countries. In an attempt to better understand the microbiology of this disease, we used phylogenetic and low-density 16S rRNA gene microarrays to characterize the gingival flora of acute noma lesions, acute necrotizing gingivitis (ANG) and healthy control subjects. Our observations allow dispelling doubts on *Fusobacterium necrophorum*, the previously suspected causative agent of noma, as this species was definitely not associated with noma lesions. Various oral pathogens were more abundant in noma lesions, notably *Atopobium* spp., *Prevotella intermedia*, *Peptostreptococcus* spp., *Streptococcus pyogenes* and *Streptococcus anginosus.* But on the other hand, periodontal pathogens such as *Aggregatibacter actinomycetemcomitans*, *Capnocytophaga spp.*, *Porphyromonas* spp. and *Fusobacteriales* were more abundant in healthy controls. Moreover, the overall loss of bacterial diversity observed in noma samples as well as its homology to that of ANG supports a previous hypothesis that ANG might be a step towards noma.

Introduction

Noma, also known as c*ancrum oris*, is a gangrenous disease that typically affects soft as well as hard tissues of the maxillo-facial region. The disorder occurs in young children of poor or less developed areas of the world, mainly in sub-Saharan countries, but also in Latin America (Barmes et al., 1997) and Asia (Bourgeois and Leclercq, 1999; Srour et al., 2008). The high mortality rate associated with the disease during the acute stage (90% lethality when children are not treated) can be explained because of starvation, sepsis or aspiration pneumonia (Baratti-Mayer et al., 2003). In 1998, the WHO estimated a global yearly incidence of 140,000 cases (Bourgeois and Leclercq, 1999), but the exact prevalence of the disease is not known. Although the exact etiology of noma is still unknown, risk factors such as malaria, measles, tuberculosis, malnutrition or poor oral hygiene are supposed to play a role in the emergence of the disease (Enwonwu et al., 1999). In children living in these less developed countries, acute necrotizing gingivitis (ANG) is suspected to be a precursor of noma (Enwonwu, 1972; Enwonwu et al., 2006; Horning, 1996), but only a small proportion of ANG cases progress to noma. A proposed theory (Falkler et al., 1999) suggests that multiple factors such as malnutrition, weakened immune functions and prior viral infections (measles, *Herpesviridae*), all worsened by poor oral hygiene, could play in unison to lower the host resistance and promote the development of oral ulcers. These lesions would then serve as sites of entrance for a trigger microorganism. Potential bacterial candidates include *Fusobacterium necrophorum* and *Prevotella intermedia* (Falkler et al., 1999; Falkler et al., 2000).

The understanding of the role of specific microorganisms, or consortia of multiple organisms, in the pathogenesis of noma remains incomplete due to fundamental shortcomings of currently available data. First, when culture techniques were used, the diversity of the microbiota was inevitably underestimated as a large range of fastidious organisms is

uncultivable (Paster 2001). Second, because the disease develops rapidly and strikes remote areas with poor access to medical facilities primarily, microbiological findings from early cases of noma are sparse. Results reported from advanced lesions may however reflect changes in local oral ecological conditions enabling the development of the disease, rather than its direct cause. Third, this disease prevails in populations of which even the normal oral flora remains poorly investigated. It therefore remains to be determined if seemingly unusual microbiological findings truly reflect the presence of disease or just geographic location, a particular lifestyle, or socioeconomic status.

To advance the knowledge on the etiology of noma, it appeared essential to us to obtain access to early acute cases of noma, to study in parallel the microbiota of related conditions and healthy controls, and to apply advanced microbiological techniques for the study of the diversity of the sampled microflora. The first step consisted in inventorying the bacterial diversity in oral samples of children in Niger with acute noma, acute necrotizing gingivitis, and healthy controls by using culture-independent phylogenetic techniques (Bolivar et al., 2009). The specific aim of the next step, reported here, was to validate and extend these observations by developing and using two distinct microarray approaches on an extensive collection of microbial samples obtained in the clinical campaign.

MATERIALS AND METHODS

Biological samples

A large epidemiological and sampling campaign was conducted as part of the GESNOMA project from September 2001 to October 2006 in the Zinder area, Niger. Gingival fluids from children, aged between 6 months and 12 years, affected either by ANG or acute noma were included in the present study. Exclusionary criteria were set to select children who had no

antibiotic, no dental cleaning and did not receive fortified food during the 3 previous months. Subjects with old lesions were also excluded. In order to establish the clinical diagnosis, every subject underwent general, facial and oral examination with special attention given to the presence of bleeding, gingival pain, ulceration, pseudo-membranes and halitosis. Gingival fluid from the dento-gingival sulcus of children was collected from both lesion and non-lesion site of the subjects using sterile endodontic paper points. For each noma case, four healthy control children originating from the same village were included. Samples were split into 5 distinct groups: healthy controls (C), noma lesion site (N), noma non-lesion site (Nn), gingivitis lesion site (G), and gingivitis non-lesion site (Gn). Samples were subsequently stored in chaotrophic medium (RLT buffer, Qiagen) at -80°C until processing.

High-density phylogenetic microarrays

Microarray design and manufacturing

The design of the phylogenetic microarray used for this study is described in (Huyghe et al., 2008). Briefly, the 25-mer oligonucleotide probes of the array match the 7 different levels (domain, phylum, class, order, family, genus and species) of the whole bacterial 16S rDNA phylogenetic tree. Probes were selected from the release 9.34 of Ribosomal Database Project (Cole et al., 2005) (RDP; 194,696 small-subunit rRNA gene sequences) in order to respect a homogenous *Tm* of 60 ± 5°C. The final probe set, composed of 9,477 probes, covered 78.3% of the RDP sequences. To minimize steric hindrance during hybridization, all 25-mer probes were poly(T)-tailed to reach an overall length of 60 nucleotides. Microarrays were manufactured by *in situ* synthesis (Agilent Technologies, Palo Alto, CA).

RNA extraction and quantification

Bacterial lysis was performed using 100 mg of glass beads (diameter: 100µm; Schieritz & Hauenstein, Switzerland) added to the samples. Volume was adjusted to 350 µl with RLT buffer (Qiagen) and samples were vortexed for 1 minute. Total RNA was isolated and purified using the RNeasy Micro Kit (Qiagen) following manufacturer's instructions. Samples were lyophilized and dissolved in 5 µl sterile water. Total RNA quality was assessed using RNA Picochips on a BioAnalyzer 2100 (Agilent). RNA quantity was assessed by one-step RT-qPCR using 0.2 µM of primers forward: 5'-GGCAAGCGTTATCCGGAATT-3', reverse: 5'-GTTTCCAATGACCCTCCACG-3' (Invitrogen, Basel, Switzerland) and 0.1 µM of probe (probe: 5'-CCTACGCGCGCTTTACGCCCA-3', 5'-end coupled to FAM and 3'-end coupled to TAMRA, Eurogentec, Seraing, Belgium). This assay is designed in a highly conserved region of the bacterial 16S rRNA gene allowing the amplification of most bacterial sequences. One-step RT-qPCR amplification (Invitrogen, final volume of 15 µl) was performed on a SDS 7700 (PE Biosystems, Santa Clara, CA, USA) using the following cycling procedure: t1, 20 min at 50°C; t2, 10 min at 94°C; t3, 15 sec at 94°C; t4, 1 min at 60°C (t3 and t4 were repeated 40 times). Using this strategy, a positive fluorescent signal was consistently obtained between cycles 21-30.

RNA amplification

The totality of purified RNA was subjected to *in vitro* transcription using the MessageAmp II-Bacteria kit (Ambion, Austin, TX, USA) according to the manufacturer's instructions. Amplified RNA was labeled during *in vitro* transcription in the presence of Cy-3 or Cy-5 cyanine dyes (Perkin-Elmer, Boston, MA, USA). Quality, quantity, amplification efficiency and dye incorporation were evaluated using the NanoDrop® ND-1000 Spectrophotometer (NanoDrop Technologies, Inc. Rockland, DE, USA) and the BioAnalyzer 2100 on RNA Nano 6000 chips (Agilent).

Microarrays hybridization, scanning and analysis

For each village, labeled cRNA from the four healthy control samples were pooled together and co-hybridized with their cognate case. Cy5- and Cy3-labelled cRNAs were diluted in a total of 250 µl Agilent hybridization buffer, and hybridized at 60°C for 17 hours in a dedicated hybridization oven (Robbins Scientific, Sunnyvale, CA, USA). Slides were washed, dried under nitrogen flow, and scanned (Agilent) using 100% PMT power for both wavelengths.

Image analysis and signal quantification were achieved using Feature Extraction software (version 6, Agilent). LOWESS (locally weighted linear regression) transformation was used to correct background subtracted signals for unequal dye incorporation. Data were quantile normalized and Log2 transformed. Samples were averaged in 5 groups: noma lesion (N) or non-lesion site (Nn), gingivitis lesion (G) or non-lesion site (Gn), and healthy controls (C). Statistical significance test consisted in variance analysis (one-way ANOVA) combined with the Benjamini and Hochberg false discovery rate correction. Differences between groups were considered statistically significant when corrected p-value ≤ 0.05 and fold change ≥ 2.

In parallel, the 517 probes showing a statistical difference between groups (p-value ≤ 0.05) were subjected to hierarchical clustering using the Ward's minimum variance method implemented in hclust from R (http://www.r-project.org/).

Low-density microarrays

Microarray design and manufacturing

Based on the most relevant sequences inventoried by the previously described cloning sequencing study by Bolivar *et al.* (Bolivar et al., 2009), we designed a low-density 16S rDNA array. Bolivar and colleagues reported a dataset containing 1'237 partial 16S rRNA gene sequences representing 339 different phylotypes. We used an arbitrary cutoff of 1% of overall abundance to select from this dataset the most abundant sequences and to design

one or more probes from each such sequence. Using this cutoff, the 132 most abundant 16S rRNA gene sequences were scanned for probes respecting defined physico-chemical properties (*Tm* = 65±5°C; probe length = 23–50 nt; < -5.0 kcal/mol for hairpins; < -8.0 kcal/mol for self-dimers; and dinucleotide repeats shorter than 5 bp) using a commercial software (Array DesignerTM 2.0 by Premier Biosoft). Consequently, the resulting candidate probes were tested for specificity using an improved version of Olicheck (Charbonnier et al., 2005) using default parameters, leading to a final set of 271 probes. Additionally, probes specific to various Archaebacteriae (downloaded from probeBase (Loy et al., 2007)) and 8 negative control probes (artificial sequences with uniform *Tm*, no secondary structure and containing more than 5 mismatches to known sequences in the public domain) were added to the final probe set. The 335 oligonucleotide probes were synthesized with a C6-linker with free primary amine (Sigma-Aldrich) and spotted on ArrayStrips microarrays (Clondiag GmbH, Jena, Germany).

DNA extraction and biotinylation PCR

Extraction of total genomic DNA (gDNA) was performed with glass beads (Sigma-Aldrich, diameter: 212-300 µm) and DNeasy kit (Qiagen) following the manufacturer's protocol. Asymmetrical PCR was chosen in order to favor the amplification of the negative strand. Total gDNA was amplified and biotinylated by duplex PCR using eubacterial universal primers fD1 (0.4 µM) and rD1 (4 µM) as published in (Weisburg et al., 1991). Each 25 µl reaction contained 15 µl of ReadyMix kit (Sigma-Aldrich, ref. P4600 withMgCl2), 1nM of biotin-16-dUTP and 10 ng of template DNA. Quality control and quantification of PCR products were assessed by using a NanoDrop® ND-1000 Spectrophotometer (NanoDrop Technologies Inc., USA).

Microarrays hybridization and scanning

Using agitation, array strips were washed 5 min with bi-distilled water and 5 min with hybridization buffer at room temperature. 10 µl of labeled amplified DNA was mixed with 90 µl of hybridization buffer, then denatured at 95°C during 2 min and subsequently kept on ice. The denatured sample was incubated in the array strip at 50°C for 1 hour. After hybridization, samples were carefully removed from the array strips, then washed 3-times with proprietary buffers (100 µl Washing Buffer I 5 min at 30°C, 100 µl Washing Buffer II 5 min at 20°C and 100 µl Washing Buffer III 5 min at 20°C). To block the reaction, 100 µl of blocking buffer solution (2% of blocking reagent in 6X SSPE/0.005% Triton) was added in each array and incubated 15 min at 30°C with shaking. Then, the blocking solution was removed and 60 µl of HRP streptavidin (100 pg/µl; Poly-HRP streptavidin by Thermo Scientific in 6X SSPE) was added, and the array strips were incubated for 15 min at 30°C. After removal of the conjugation solution, array strips were washed 3-times during 5 min with proprietary buffers. After removal of the washing solution, 50 µl of staining solution (SeramunGrün by Seramun Diagnostica GmbH) was added and incubated for 5 min at 25°C. After removal of the staining solution, scanning of the microarrays was performed with the Clondiag Array Mate device.

Data analysis

Signal intensities were extracted using IconoClust software (Clondiag GmbH). For each spot, extinction signal of local background was subtracted from the intensity value. Normalization and analysis were performed using the Partek Genomic Suite 6.4 (Partek, USA). Firstly, array signals were adjusted in order to normalize the quantity of DNA. Subsequently, each array was normalized against the 8 negative control probes, and then quantile-normalized. Following calibration experiments (not shown), we defined spots with a signal value $\geq$ 0.35 and a confidence index $\geq$ 0.75 as positive spots. Differences among groups were sought using ANOVA test, where probes with a fold-change $\geq$ 2 and a *P value* $\leq$ 0.05 were considered as statistically significant.

RESULTS

High density phylogenetic microarrays

Eighty-four samples were hybridized on the high-density phylogenetic arrays (Table 1). In order to highlight statistically significant differences, among these probes; reflecting imbalance in flora composition across the 5 defined conditions, ANOVA (Analysis Of VAriance) analysis was performed on this dataset. The analysis identified 517 probes showing a p-value $\leq$ 0.05, of which 123 probes showed a fold-change $\geq$ 2 in a last one the four pair-wise comparisons (Figure 1). The comparison between lesion-sites of noma samples (N) and healthy controls (C) clearly showed the highest number of probes with a significant difference (n = 82).

Compared to the healthy controls, a lower bacterial diversity was found in ANG and even a lower one was recorded in noma samples. In both gingivitis and noma lesion-sites, *Peptostreptococcus* spp., *Prevotella* spp. and *Nocardioidaceae* were significantly more represented in lesion sites as compared to their cognate non-lesion sites. When comparing diseased sites to healthy controls, the largest difference was found in the groups of the *Firmicutes*, the *Bacteroidetes* and the *Fusobacteriales*. Indeed, less *Porphyromonadaceae*, *Tannerella spp.*, *Capnocytophaga spp.* (up to 10-fold changes), *Fusobacteriales* (probe matching Leptotrichia and *Streptobacillus moniliformis*) and *Cetobacterium spp.* (on the basis of the latest RDP release - version 10 - the corresponding probes would be more precisely redefined as matching the *Leptotrichia* genus) were found in noma samples. However, *Syntrophomonadaceae* and saprophyte phylotypes such as *Nitrospina*, *Haliangium*, *Saprospiraceae*, and *Hymenobacter* appeared over-represented in gingivitis samples only. When comparing diseased sites to healthy controls, one single taxon (described as *Unnamed* in the RDP and included in the *Cyanobacteria* phylum) revealed more abundant in noma lesions only. This “taxon” includes various environmental *Cyanobacteria* such as

Phormidium spp., *Synechococcus* spp., *Scytonema crustaceum*, *Oscillatoria rosea* and *Calothrix contarenii*.

Clearly, lesion-sites of gingivitis samples exhibit a higher bacterial diversity than lesion-sites of noma samples (Figure 1: column N/G), notably within the *Bacteroidetes* (including *Porphyromonadaceae* and *Capnocytophaga* spp.) and *Firmicutes* (particularly *Peptococcaceae* and *Syntrophomonadaceae*) phyla. Given that noma lesions stay localized (Baratti-Mayer et al., 2003), we compared bacterial content of lesion sites to non-lesion sites in an attempt to describe intra-patient differences in micro-flora composition (Figure 1: column N/Nn). Noticeably, a broader bacterial diversity is found in non-lesion sites of noma samples, particularly with groups such as *Proteobacteria* (including *Lautropia*, an oral taxon of the *Burkholderiaceae* family), *Bacteroidetes* (including *Capnocytophaga* spp.), *Firmicutes* (notably with *Veillonella* spp. and *Streptococcus* spp.) and *Fusobacteriales*. Conversely, the genus *Peptostreptococcus* is 2-fold more abundant in the lesional sites of our noma samples.

ANG and noma samples, although extremely similar, exhibit some differences, notably in terms of phylotypes classically associated to various environmental milieus: *Cyanobacteria* (including “unclassified Family 1.1”), *Streptomycetaceae* and *Oceanimonas* are more abundant in noma samples. On the contrary, probes matching phylotypes such as *Microcoleus*, *Pasteurellaceae*, *Peptococcaceae*, *Bacteroidetes* (matching essentially *Alistipes* spp. and *Porphyromonas* spp.), *Porphyromonadaceae* (including *Porphyromonas* spp.), *Capnocytophaga*, *Sphingobacterium* and *Psychroflexus* show limited hybridization signals with noma samples.

Hierarchical clustering performed on the probes exhibiting significant p-value between conditions (n = 517), allows grouping our samples in 4 subclusters (Figure 2). In clusters I and IV, lesion and non-lesion sites of the same patient co-clustered at a limited distance, with a predominance of gingivitis samples in cluster I and a majority of noma samples in cluster IV. Cluster II regroups noma lesion site samples, with the exception of one non-lesion site

sample co-clustering with its lesional counterpart (subject #24). Compared to other profiles, samples of cluster II show a lower bacterial diversity, except for some *Cyanobacteria*. On the other hand, cluster III regroups a majority of non-lesion noma samples.

Low-density microarray

We designed a 16S rRNA array based on the most relevant sequences inventoried by a cloning sequencing study (Bolivar *et al.*, submitted) and obtained from our high-density microarray analysis, in noma, ANG and healthy controls. The purpose of this array is to extrapolate the results obtained by this previous study on a larger scale.

We successfully labeled and amplified 673 samples (Table 1) from the five different conditions. ANOVA analysis was performed on our dataset, resulting in 187 probes with a significant group p-value (P <= 0.05). The microbiologic profiles of four groups (noma lesion and non-lesion site, gingivitis lesion site and healthy controls) are depicted in Figure 3. Two genera are present in 100% of all four groups: *Fusobacterium* sp. and *Prevotella* sp. Overall, some bacteria are found in high abundance in the lesion sites of gingivitis subjects when compared to the other groups. These genera/families include: some *Lachnospiraceae* (including *Catonella* sp. and *Oribacterium* sp.), *Lautropia* sp., *Peptostreptococcus* sp., *Spirochaetaceae* (more particularly *Treponema* sp.) and some *Prevotellaceae*. Conversely, *Capnocytophaga* sp. is found in higher prevalence in healthy samples compared to the other groups. Several genera are found in high abundance in each 4 groups such as: *Abiotrophia* sp., *Clostridiales*, *Fusobacterium* sp., *Gemella morbillorum*, *Neisseria cinerea*, some *Prevotellaceae*, *Streptococcus thermophilus* and *S. gordonii*. Moreover, *P. intermedia* strain 6 is found in higher frequency in lesion sites of ANG and noma (93.5% and 62.6% of the subjects respectively) in comparison to healthy subjects (11.3%); same observation for *P. melaninogenica* with a prevalence of 100% and 94% in the lesion sites of gingivitis and noma respectively, and only 29.3% in healthy subjects. Strikingly, except for some

Lachnospiraceae, no taxon appears more prevalent in the lesion sites of noma subjects compared to the 3 other groups.

ANOVA analysis reveals 135 probes statistically significant (P<=0.05 and |fold-change| >= 2) between gingivitis, noma and healthy groups (figure 4). Remarkably, three genera/species are more abundant in healthy sites/subjects when compared to the lesion sites of both noma and gingivitis: *Aggregatibacter actinomycetemcomitans*, *Lautropia* sp., *Neisseria* sp., *Streptococcus sanguinis* and *Capnocytophaga* sp.; this last genus exhibiting the highest fold changes in the control conditions (up to 20.45-fold in gingivitis and 13.49-fold in noma). Conversely, lesion sites present a higher abundance in *Bacteroidetes*, *Dialister pneumosintes* (11-fold in noma), *Filifactor alocis* (*Peptostreptococcaceae*), *Lachnospiraceae*, *Porphyromonas endodontalis*, *Prevotellaceae* (particularly *Prevotella nigrescens* and *P. intermedia*), *Spirochaetaceae* (including *Treponema* sp.) and *Streptococcus oligofermentans*. Note also that the potential pathogens *Leptotrichia* and some *Porphyromonadaceae* are less abundant in noma lesions compared to gingivitis. When compared to gingivitis and controls, lesion sites of noma samples clearly show a higher abundance in *Actinobacteria* (probe matching *Atopobium* spp.), *Prevotellaceae*, *Streptococcus pyogenes*; and *Staphylococcaceae* and *Streptococcus anginosus* when compared to gingivitis only. Interestingly, *Fusobacterium* sp. appears under-represented in the lesion site of noma samples.

In order to evaluate the relatedness between conditions we performed Principal Component Analysis (PCA) on the dataset resulting from the ANOVA analysis (de Haan et al., 2007). The first principal component (figure 5) contains 53.5% of the explained variance and lesion samples of both gingivitis and noma appear closely related, while non-lesions sites appear more related to healthy controls. However, on the second component (23.6% of explained variance) a clear distinction can be made between gingivitis (lower part of the graph) and noma samples.

DISCUSSION

Our study is to date, the first and largest effort to elucidate the role of microorganisms in the etiology of noma. A total of 757 samples were incorporated in our study, representing 413 different subjects from the Zinder area in Niger. Following clinical examinations of patients and control, samples were stabilized in chaotrophic salt and send to our laboratory, benefiting from the implication of the GESNOMA group (Baratti-Mayer et al., 2004). Overall, the vast majority of samples were of sufficient quality for RNA extraction or PCR amplification. Approximately 5% of samples only were not analyzable. 84 samples were hybridized on our phylogenetic array in order to compare global bacterial profiles between conditions. In parallel, 673 samples were hybridized on the low-density microarray in an attempt to strengthen the results obtained by Bolivar et al. (Bolivar et al., 2009) on a larger scale.

Results obtained with both arrays are in good agreement and, reassuringly, observations made with the previous cloning sequencing study of Bolivar et al. (Bolivar et al., 2009) and our low-density array go in the same direction. However, a few taxa were only characterized by either one of the two approaches. Several factors can explain this slight disparity. First, the low-density array was designed in order to detect the most abundant phylotypes (≥1%) characterized with the cloning sequencing study of Bolivar et al. When cloning sequencing allows absolute quantification of bacterial components, our phylogenetic approach allows only relative quantification (fold-change), meaning that we only measure differences between conditions. Secondly, the probe set of our phylogenetic array is not comprehensive as we cover 78.3% of the sequences of the RDP (release 9.34), thus, 16S sequences not represented on our array are simply missed. Finally, cloning sequencing for bacterial identification can suffer from PCR primers bias. Indeed, the specificity of the so-called

“universal” primers can be limited for certain taxa (Baker et al., 2003; Blackwood et al., 2005; Isenbarger et al., 2008) and 16S rRNA libraries are not necessarily representative of true prokaryotic diversity.

We observed that potential oral pathogens such as are more abundant in healthy samples compared to diseased conditions such as *Aggregatibacter actinomycetemcomitans*, *Capnocytophaga* (up to 20-fold), *Porphyromonas* and *Fusobacteriales*. *Fusobacterium necrophorum* is an opportunistic pathogen associated with necrobacillosis in wallabies, a disease similar to noma in humans (Enwonwu, 1995). This genus was recovered from seven of eight advanced noma lesions (Falkler et al., 1999) and was long considered as playing a key role in the development of the disease. In our study, members of the genus *Fusobacteria* appear neither more prevalent nor more abundant in noma lesions, and other representatives of the *Fusobacteriales* order (more precisely *Cetobacterium*, *Leptotrichia* and *Streptobacillus moniliformis*) are associated with healthy sites. Our observations comfort results obtained during previous cloning sequencing studies performed by Bolivar (Bolivar et al., 2009) and Paster (Paster et al., 2002), and could finally dispel doubts on this genus.

In the present study, Prevotellaceae and more precisely *Prevotella intermedia* are clearly associated with noma. The presence of *P. intermedia* in noma lesions was already documented in previous studies (Bolivar et al., 2009; Falkler et al., 1999; Phillips et al., 1995), but was not detected in the 4 noma samples processed in the cloning sequencing study of Paster and colleagues (Paster et al., 2002). This pathogen is encountered in adult periodontitis (Slots, 1986; Wennström et al., 1987) and is frequently isolated in endodontic infections (Gharbia et al., 1994; Tomazinho and vila-Campos, 2007; van Winkelhoff et al., 1985). It could take part in etiology of noma by promoting tissue destruction by its ability to degrade lipids and produce proteolytic enzymes (Enwonwu et al., 2000). Moreover, *P. intermedia* produces immunoglobulin A1 (IgA1) proteases which could play a critical role in decreasing the mouth mucosal immunity (Frandsen et al., 1995), hence promoting development of other pathogens in oral lesions.

In noma lesions, we observed a notable abundance of other phylotypes commonly associated with oral infections, such as dental caries (*Atopobium* (Preza et al., 2008)), periodontitis (*Prevotella* spp. (Colombo et al., 2009) and *Peptostreptococcus* spp. (Rôças and Siqueira, 2008)), dentoalveolar infections (*Prevotella* spp. (Parahitiyawa et al., 2009)) or palatal abscess (*S. anginosus* (Whiley et al., 1990)). Additionally, *S. pyogenes* was recovered in noma lesions. This strictly human pathogen can cause a variety of infections such as skin infections (impetigo), throat infections including pharyngitis and tonsillitis (Aziz and Kotb, 2008). *S. pyogenes* is also involved in necrotizing fasciitis, a subcutaneous infection with high mortality rates (Bisno and Stevens, 1996); and could potentially participate in the development of noma lesions by the release of various virulence factors such as streptolysin, proteases and exotoxins.

An interesting feature of our study is the presence of certain phylotypes never recovered in the oral cavity, such as Cyanobacteria. The same observation was made in the previous cloning sequencing study of Paster (Paster et al., 2002). These environment-associated taxa likely colonized the host as advanced noma lesions are open and subject to external contamination.

Compared to gingivitis we observed a lesser overall bacterial diversity in lesion sites of noma and tends to comfort previous studies stipulating that gingivitis can progress to noma if untreated (Enwonwu, 1972; Horning, 1996).

Although no bacterial species was identified as the causative agent of noma, our study gives better insight on the bacteriology of noma. Our results will be integrated in a large epidemiological study and the resulting observations will add to the understanding of the pathogenesis of noma. The observed loss of bacterial diversity and apparition of oral pathogens in noma lesions is interesting for the understanding of the disease. However, we cannot confirm that the alteration of the oral microbiota in noma lesions explains the disease by itself, or if our observations reflect a secondary infection. Our study gives better insight on

the bacteriology of noma, but additional studies are necessary to decipher the etiology of the disease and answer this question.

FIGURES

1- Sample Collection

	Site	High-density arrays	Low-density arrays
Gingivitis	lesion (l)	7	31
	non-lesion (nl)	7	34
Noma	lesion (l)	19	83
	non-lesion (nl)	19	82
Controls	-	32	443
	Total	84	673

Figure 1

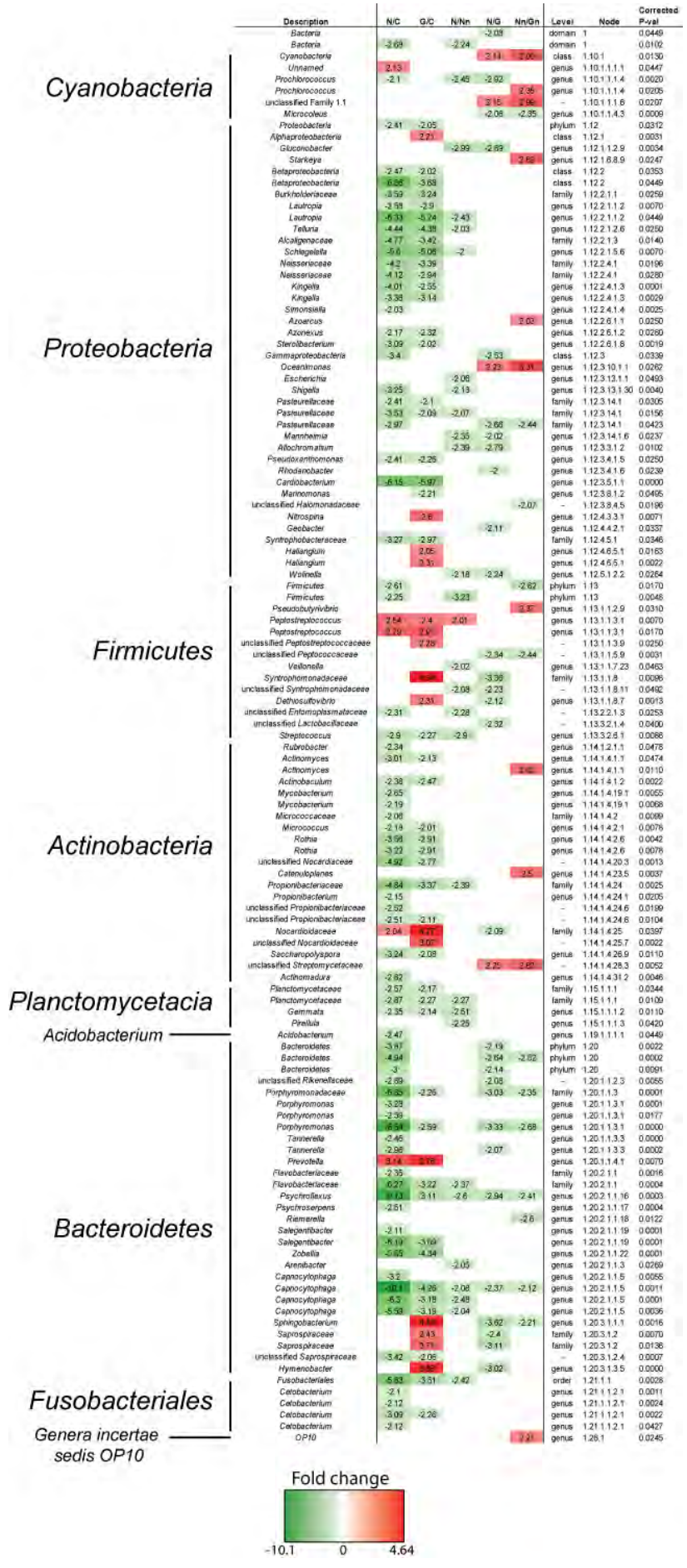

Description
N/C
G/C
N/Nn
N/G
Nn/Gn
Level
Node
Corrected P-val
Cyanobacteria
Proteobacteria
Firmicutes
Actinobacteria
Planctomycetacia
Acidobacterium
Bacteroidetes
Fusobacteriales
Genera incertae sedis OP10
Fold change
-10.1
0
4.64

Figure 2

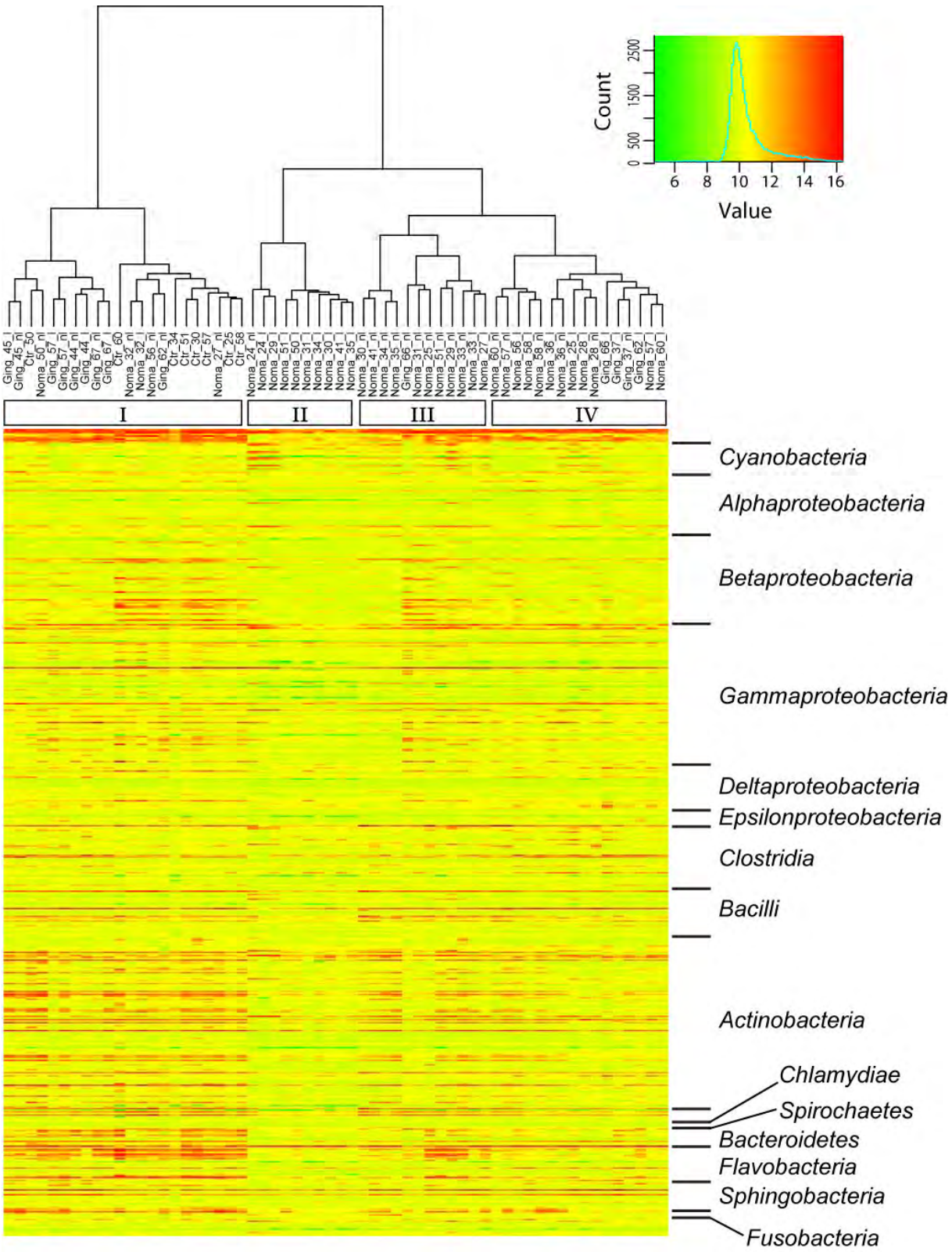

Count
0
500
1500
2500
6
8
10
12
14
16
Value
I
II
III
IV
Cyanobacteria
Alphaproteobacteria
Betaproteobacteria
Gammaproteobacteria
Deltaproteobacteria
Epsilonproteobacteria
Clostridia
Bacilli
Actinobacteria
Chlamydiae
Spirochaetes
Bacteroidetes
Flavobacteria
Sphingobacteria
Fusobacteria

Figure 3

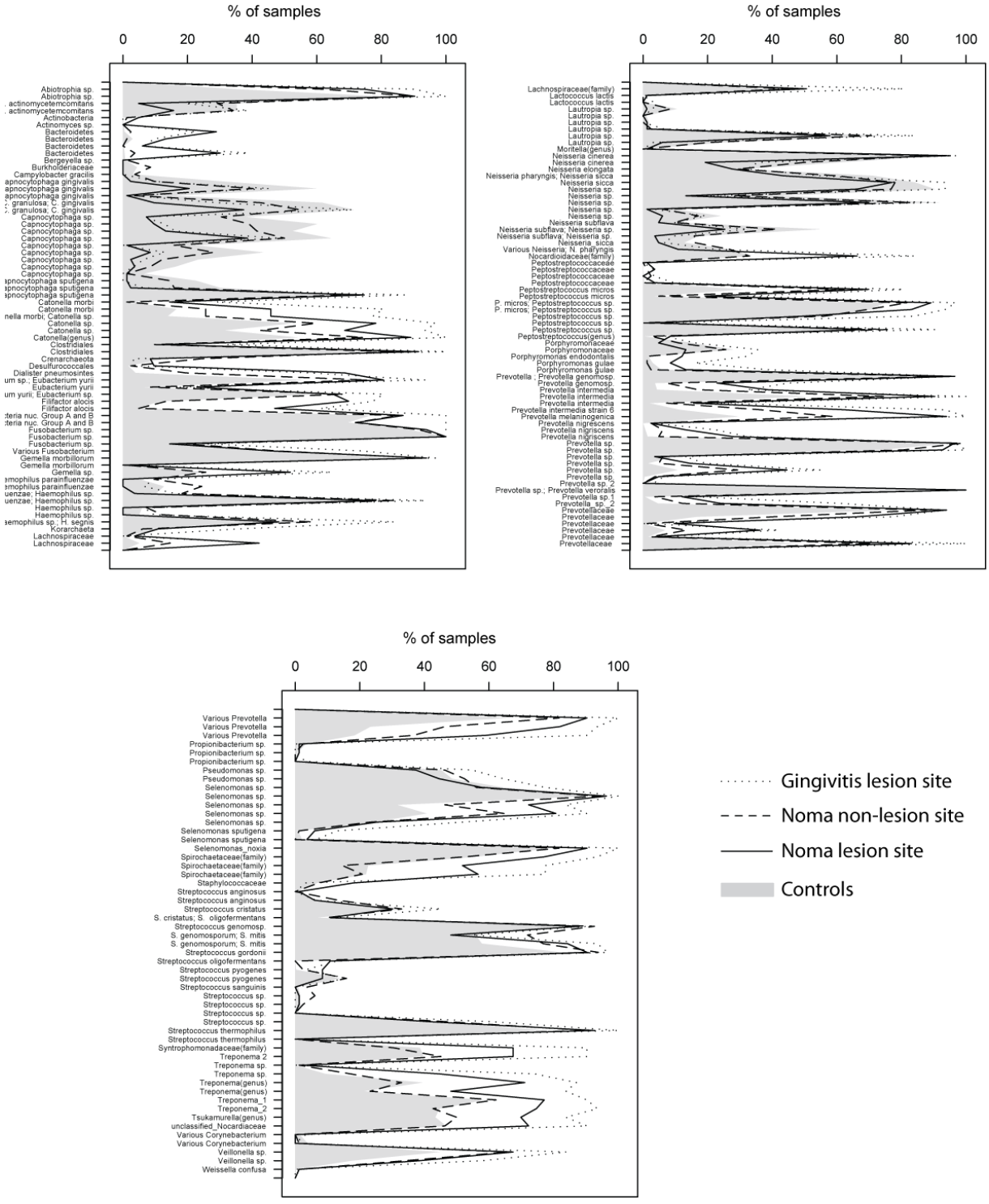

% of samples
0
20
40
60
80
100
Gingivitis lesion site
Noma non-lesion site
Noma lesion site
Controls

Figure 4

Description	N/C	G/C	N/Nn	N/G	Nn/Gn
Aberrophio sp.		2.17			
Aggregatibacter actinomycetemcomitans	-5.13	-1.57	-7.51		2.32
Aggregatibacter actinomycetemcomitans	-4.15	-2.36	-3.85		
Actinobacteria	3.11		5.39	7.58	
Actinomyces naeslundii			-2.94		
Bacteroidetes	5.01	3.35	3.76		
Bacteroidetes	4.80		4.06		
Bacteroidetes	6.07	3.69	2.47		
Bacteroidetes	2.93		3.80		
Bacteroidetes	2.25				
Beripeyella sp.	-2.36	-3.01			
Burkholderiaceae		-3.79	-2.04		3.30
Campylobacter gracilis		-6.29			
Capnocytophaga gingivalis	-4.29	-3.23	-2.58		
Capnocytophaga gingivalis	-2.91	-8.82			
Capnocytophaga gingivalis	-5.34	-9.74			
Capnocytophaga granulosa; C. gingivalis	-7.44	-5.96	-3.86		
Capnocytophaga granulosa; C. gingivalis	-3.18	-2.87			
Capnocytophaga sp.	-10.34	-12.00	-5.35		
Capnocytophaga sp.	-3.99	-3.00	-2.14		
Capnocytophaga sp.	-7.00	-7.82	-4.20		
Capnocytophaga sp.	-7.34	-6.26	-3.70		
Capnocytophaga sp.	-5.30	-5.48	-3.44		
Capnocytophaga sp.	-4.21	-8.31	-3.06		
Capnocytophaga sp.	-2.99	-2.91	-2.04		
Capnocytophaga sp.	-8.90	-7.76			
Capnocytophaga sp.	-3.70				
Capnocytophaga sp.	-9.12	-10.56	-5.51		
Capnocytophaga sp.	-3.89				
Capnocytophaga sputigena	-10.48	-7.67	-4.99		
Capnocytophaga sputigena	-13.49	-20.45	-6.17		
Catonella morbi	2.59	2.60			
Catonella morbi	2.05	2.36			
Catonella morbi; Catonella sp.	2.61	3.57			
Catonella sp.		2.04			
Catonella sp.	2.41	2.50			
Catonella sp.		2.09			
Clostridiales	2.15				
Clostridiales					-2.42
Various *Corynebacterium; C. matruchotii*	-4.74	-7.18			
Various *Corynebacterium; C. matruchotii*	-3.07				
Desulfurococcales			2.28		
Dialister pneumosintes	-11.01	-8.17	3.29		
Eubacterium yurii	2.07				
Filifactor alocis	5.37	3.32	4.48		
Filifactor alocis	5.90	3.69	5.41		-2.21
Fusobacterium sp.	-2.60		-2.10		
Gemella morbillorum	-2.48				
Gemella sp.		2.02			
Haemophilus parainfluenzae	-3.58		-4.56		
Haemophilus parainfluenzae; Haemophilus sp.		-2.50			
Haemophilus sp.	-3.69		-3.69		
Haemophilus sp.	-2.78		-3.51		
Lachnospiraceae	2.27				
Lachnospiraceae	4.56	3.13	2.32		
Lactococcus lactis					2.89
Lautropia sp.	-4.29				
Lautropia sp.	-5.29	-5.96	-3.78		
Lautropia sp.	-2.29	-4.31			
Lautropia sp.	-3.82	-10.54	-2.40		
Lautropia sp.	-2.97	-12.57			
Leptotrichia sp.	-3.44				
Leptotrichia sp.	-3.29				
Montella sp.	2.39		2.35		
Neisseria cinerea	-3.30	-3.20	-2.20		
Neisseria pharyngis; Neisseria sicca	-2.26	-2.27			
Neisseria sp.	-3.73	-3.77	-2.85		
Neisseria sp.	-2.08				
Neisseria sp.	-6.79	-8.31	-4.19		
Neisseria sp.	-2.33	-6.79			
Neisseria subflava		-3.53			
Neisseria subflava; Neisseria sp.	-2.39	-3.23			
Neisseria subflava; Neisseria sp.	-3.30	-4.80	-2.67		
Neisseria subflava; Neisseria sp.; *N.cinerea*		-2.06			
Neisseria sicca	-3.05	-4.24	-2.85		
Various *Neisseria; N. pharyngis*	-2.36	-2.59	-2.11		
Nocardioidaceae		2.41	2.06		
Peptostreptococcaceae	2.69		3.84		
Peptostreptococcus micros	3.49	3.00			
Peptostreptococcus sp.		2.23			
Peptostreptococcus sp.	5.19	3.06			
Porphyromonadaceae	-2.34				
Porphyromonadaceae	-2.46				
Porphyromonas endodontalis	2.45	3.24	3.64		
Porphyromonas endodontalis			-3.03		
Porphyromonas gulae		3.05	3.03		
Porphyromonas gulae		2.57	2.76		
Prevotella genomosp.	2.36	3.17			
Prevotella intermedia		2.03			
Prevotella intermedia	5.13	3.37	2.06		
Prevotella intermedia	5.16	3.02	2.29		
Prevotella intermedia strain 6	4.04	4.79			
Prevotella melaninogenica	3.58	2.88			
Prevotella nigrescens	3.52				
Prevotella nigrescens	7.47	1.62	2.62		
Prevotella nigrescens	4.77	4.86			
Prevotella sp.	2.24				
Prevotella sp. 1	4.80	2.67			
Prevotella sp. 2	2.72	4.17			
Prevotella sp. 2	2.17				
Prevotellaceae	4.21	2.85			
Prevotellaceae	4.07		5.58	7.38	
Prevotellaceae	2.76	2.24			
Prevotellaceae	2.80	2.95			
various *prevotella*	4.10	2.41			
various *prevotella; P. intermedia*	2.96	3.05			
Various *Prevotella; P. intermedia* strain 2	4.77	3.69	3.17		
Various *Prevotella; P. intermedia* strain 5	3.25	4.31			
Propionibacterium sp.	-3.62	-10.46			
Propionibacterium sp.			4.46		
Selenomonas sp.; *Selenomonas* sp.	2.08				
Spirochaetaceae		2.73	2.00		
Spirochaetaceae	2.01	3.40	2.89		-2.34
Staphylococcaceae				2.71	2.30
Staphylococcaceae	-2.01	-3.02			
Streptococcus anginosus		-2.25			
Streptococcus anginosus				3.61	
Streptococcus cristatus; S. oligofermentans					2.06
Streptococcus genomosporum; S. mitis			-2.11		
Streptococcus oligofermentans	4.83	4.02	4.94		
Streptococcus oligofermentans	2.56		3.69		
Streptococcus pyogenes	5.22		2.05	3.69	
Streptococcus pyogenes		-2.58			
Streptococcus sanguinis	-2.85	-6.42	-3.04		
Streptococcus sanguinis	-2.41	-5.75	-2.24		
Streptococcus sp.		-2.04			
Streptococcus sp.		-2.90			
Streptococcus sp.		-2.67			
Tannerella forsythensis	-2.44				
Treponema sp.	3.76	3.68	2.01		
Treponema sp.	2.70				
Treponema sp.	2.45	2.26	2.02		
Treponema sp.		2.07	2.14		-2.09
Treponema 2		2.47			-2.27
Tsukamurella sp.		2.24			

Figure 5

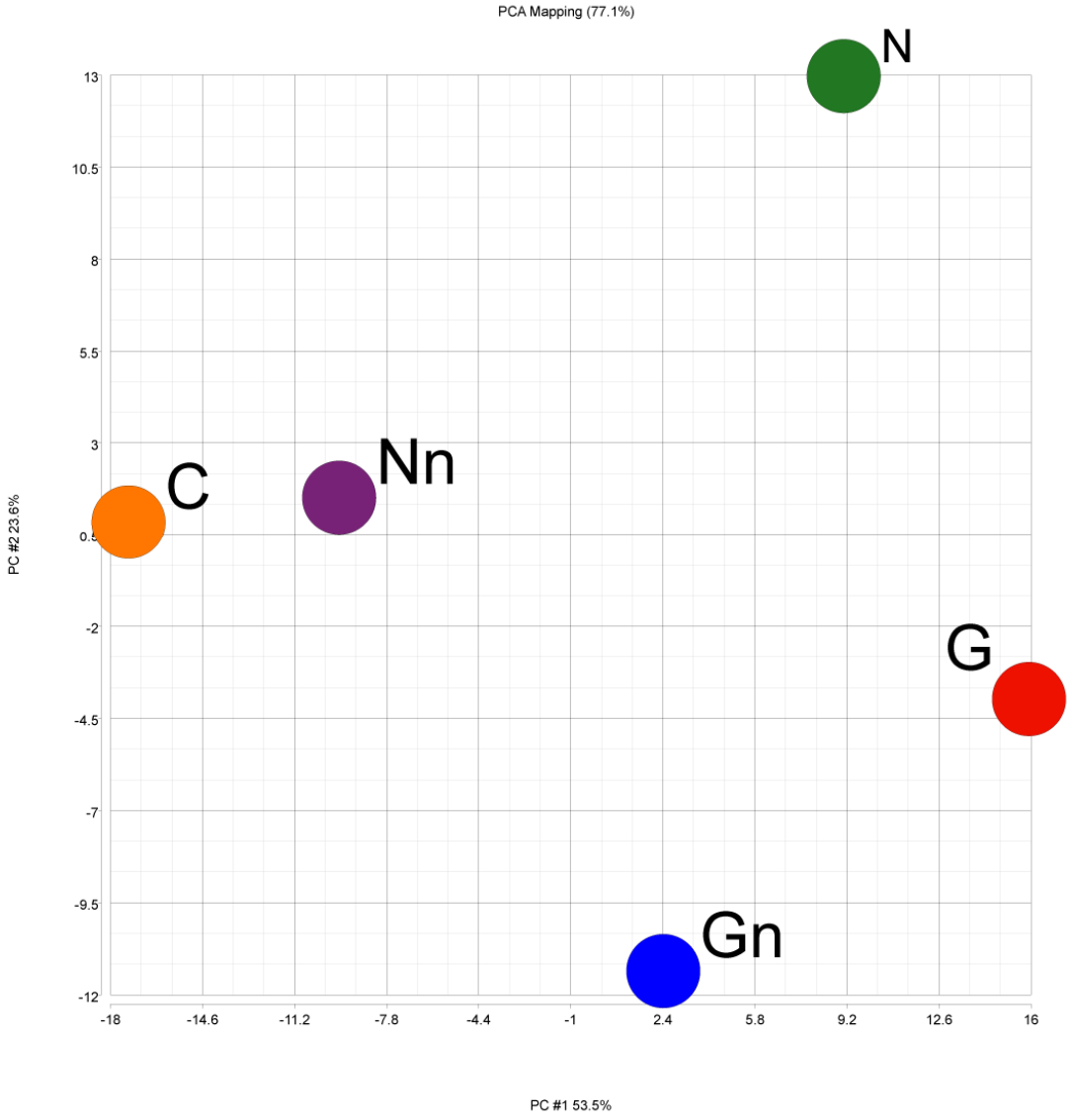

PCA Mapping (77.1%)
N
Nn
C
G
Gn
PC #2 23.6%
PC #1 53.5%
13
10.5
8
5.5
3
-2
-4.5
-7
-9.5
-12
-18
-14.6
-11.2
-7.8
-4.4
-1
2.4
5.8
9.2
12.6
16

Figure Legends

Table 1

Sample collection.

Figure 1

Heatmap of samples hybridized on phylogenetic microarrays. Only probes showing a statistically significant signal between groups (|fold-change| >= 2 and p-value <=0.05; n = 123) are depicted here. Figures in boxes represent fold-changes. C: Healthy controls; N: Noma lesion site; Nn: Noma non-lesion site; G: Gingivitis lesion site; Gn: Gingivitis non-lesion site.

Figure 2

Hierarchical clustering of samples hybridized on phylogenetic microarrays. Hierarchical clustering was performed on the 517 probes with a p-value <= 0.05. Samples are grouped in 5 distinct clusters, I to V. Samples with the "_l" prefix stand for lesion sites, while "_nl" stands for non-lesion sites.

Figure 3

Prevalence of bacteria that are significantly different among subjects.

Figure 4

Heatmap of samples hybridized on the low-density microarrays. Figures represent averaged signal intensities fold-changes in the different group comparisons. C: Healthy controls; N: Noma lesion site; Nn: Noma non-lesion site; G: Gingivitis lesion site; Gn: Gingivitis non-lesion site.

Figure 5

PCA performed on the dataset resulting from the ANOVA analysis. C: Healthy controls; N: Noma lesion site; Nn: Noma non-lesion site; G: Gingivitis lesion site; Gn: Gingivitis non-lesion site.

Reference List

Aziz,R.K. and Kotb,M. (2008) Rise and persistence of global M1T1 clone of Streptococcus pyogenes. *Emerg Infect Dis* 14: 1511-1517.

Baker,G.C., Smith,J.J., and Cowan,D.A. (2003) Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* 55: 541-555.

Baratti-Mayer,D., Pittet,B., and Montandon,D. (2004) [GESNOMA (Geneva Study Group on Noma): state-of-the-art medical research for humanitarian purposes]. *Ann Chir Plast Esthet* 49: 302-305.

Baratti-Mayer,D., Pittet,B., Montandon,D., Bolivar,I., Bornand,J.E., Hugonnet,S. et al. (2003) Noma: an "infectious" disease of unknown aetiology. *The Lancet Infectious Diseases* 3: 419-431.

Barmes,D.E., Enwonwu,C.O., Leclercq,M.H., Bourgeois,D., and Falkler,W.A. (1997) The need for action against oro-facial gangrene (noma). *Trop Med Int Health* 2: 1111-1114.

Bisno,A.L. and Stevens,D.L. (1996) Streptococcal Infections of Skin and Soft Tissues. *N Engl J Med* 334: 240-246.

Blackwood,C.B., Oaks,A., and Buyer,J.S. (2005) Phylum- and class-specific PCR primers for general microbial community analysis. *Appl Environ Microbiol* 71: 6193-6198.

Bolivar,I., Stadelmann,B., Baratti-Mayer,D., Mombelli,A., and And the Gesnoma study group . Bacterial diversity in oral samples of children in Niger with acute Noma, acute necrotizing gingivitis, and healthy controls. 2009. *Submitted*

Bourgeois,D.M. and Leclercq,M.H. (1999) The World Health Organization initiative on noma. *Oral Dis* 5: 172-174.

Charbonnier,Y., Gettler,B., François,P., Bento,M., Renzoni,A., Vaudaux,P. et al. (2005) A generic approach for the design of whole-genome oligoarrays, validated for genomotyping, deletion mapping and gene expression analysis on Staphylococcus aureus. *BMC Genomics* 6: 95.

Cole,J.R., Chai,B., Farris,R.J., Wang,Q., Kulam,S.A., McGarrell,D.M. et al. (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res* 33: D294-D296.

Colombo,A.P., Boches,S.K., Cotton,S.L., Goodson,J.M., Kent,R., Haffajee,A.D. et al. (2009) Comparisons of subgingival microbial profiles of refractory periodontitis, severe periodontitis, and periodontal health using the human oral microbe identification microarray. *J Periodontol* 80: 1421-1432.

de Haan,J.R., Wehrens,R., Bauerschmidt,S., Piek,E., van Schaik,R.C., and Buydens,L.M.C. (2007) Interpretation of ANOVA models for microarray data using PCA. *Bioinformatics (Oxford, England)* 23: 184-190.

Enwonwu,C.O. (1972) Epidemiological and biochemical studies of necrotizing ulcerative gingivitis and noma (cancrum oris) in Nigerian children. *Arch Oral Biol* 17: 1357-1371.

Enwonwu,C.O. (1995) Noma: a neglected scourge of children in sub-Saharan Africa. *Bull World Health Organ* 73: 541-545.

Enwonwu,C.O., Falkler,W.A., and Idigbe,E.O. (2000) Oro-facial gangrene (noma/cancrum oris): pathogenetic mechanisms. *Crit Rev Oral Biol Med* 11: 159-171.

Enwonwu,C.O., Falkler,W.A., Idigbe,E.O., and Savage,K.O. (1999) Noma (cancrum oris): questions and answers. *Oral Dis* 5: 144-149.

Enwonwu,C.O., Falkler,W.A., and Phillips,R.S. (2006) Noma (cancrum oris). *Lancet* 368: 147-156.

Falkler,W.A., Enwonwu,C.O., Ewell,A.J., and Idigbe,E.O. (2000) Isolation of fusobacteria from the oral cavities of malnourished Nigerian children living in agricultural and herding villages. *Oral Dis* 6: 103-105.

Falkler,W.A., Enwonwu,C.O., and Idigbe,E.O. (1999) Isolation of Fusobacterium necrophorum from cancrum oris (noma). *The American Journal of Tropical Medicine and Hygiene* 60: 150-156.

Frandsen,E.V., Reinholdt,J., Kjeldsen,M., and Kilian,M. (1995) In vivo cleavage of immunoglobulin A1 by immunoglobulin A1 proteases from Prevotella and Capnocytophaga species. *Oral Microbiol Immunol* 10: 291-296.

Gharbia,S.E., Haapasalo,M., Shah,H.N., Kotiranta,A., Lounatmaa,K., Pearce,M.A., and Devine,D.A. (1994) Characterization of Prevotella intermedia and Prevotella nigrescens isolates from periodontic and endodontic infections. *J Periodontol* 65: 56-61.

Horning,G.M. (1996) Necotizing gingivostomatitis: NUG to noma. *Compend Contin Educ Dent* 17: 951-4, 956, 957.

Huyghe,A., François,P., Charbonnier,Y., Tangomo-Bento,M., Bonetti,E.J., Paster,B.J. et al. (2008) Novel microarray design strategy to study complex bacterial communities. *Appl Environ Microbiol* 74: 1876-1885.

Isenbarger,T.A., Finney,M., Rios-Velazquez,C., Handelsman,J., and Ruvkun,G. (2008) Miniprimer PCR, a New Lens for Viewing the Microbial World. *Appl Environ Microbiol* 74: 840–849.

Loy,A., Maixner,F., Wagner,M., and Horn,M. (2007) probeBase--an online resource for rRNA-targeted oligonucleotide probes: new features 2007. *Nucleic Acids Res* 35: D800-D804.

Parahitiyawa,N.B., Scully,C., Leung,W.K., Yam,W.C., Jin,L.J., and Samaranayake,L.P. (2009) Exploring the oral bacterial flora: current status and future directions. *Oral Dis*.

Paster,B.J., Falkler Jr,W.A., Enwonwu,C.O., Idigbe,E.O., Savage,K.O., Levanos,V.A. et al. (2002) Prevalent bacterial species and novel phylotypes in advanced noma lesions. *J Clin Microbiol* 40: 2187-2191.

Phillips,R.S., Enwonwu,C.O., and Falkler,W.A. (1995) Pro- versus anti-inflammatory cytokine profile in African children with acute oro-facial noma (cancrum oris, noma). *Eur Cytokine Netw* 16: 70-77.

Preza,D., Olsen,I., Aas,J.r.A., Willumsen,T., Grinde,B.r., and Paster,B.J. (2008) Bacterial profiles of root caries in elderly patients. *J Clin Microbiol* 46: 2015-2021.

Rôças,I.N. and Siqueira,J.F. (2008) Root canal microbiota of teeth with chronic apical periodontitis. *J Clin Microbiol* 46: 3599-3606.

Slots,J. (1986) Bacterial specificity in adult periodontitis. A summary of recent work. *J Clin Periodontol* 13: 912-917.

Srour,M.L., Watt,B., Phengdy,B., Khansoulivong,K., Harris,J., Bennett,C. et al. (2008) Noma in Laos: Stigma of Severe Poverty in Rural Asia. *Am J Trop Med Hyg* 78: 539-542.

Tomazinho,L.F. and vila-Campos,M.J. (2007) Detection of Porphyromonas gingivalis, Porphyromonas endodontalis, Prevotella intermedia, and Prevotella nigrescens in chronic endodontic infection. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 103: 285-288.

van Winkelhoff,A.J., Carlee,A.W., and de Graaff,J. (1985) Bacteroides endodontalis and other black-pigmented Bacteroides species in odontogenic abscesses. *Infect Immun* 49: 494-497.

Weisburg,W.G., Barns,S.M., Pelletier,D.A., and Lane,D.J. (1991) 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* 173: 697–703.

Wennström,J.L., Dahlén,G., Svensson,J., and Nyman,S. (1987) Actinobacillus actinomycetemcomitans, Bacteroides gingivalis and Bacteroides intermedius: predictors of attachment loss? *Oral Microbiol Immunol* 2: 158-162.

Whiley,R.A., Fraser,H., Hardie,J.M., and Beighton,D. (1990) Phenotypic differentiation of Streptococcus intermedius, Streptococcus constellatus, and Streptococcus anginosus strains within the "Streptococcus milleri group". *J Clin Microbiol* 28: 1497-1501.

Chapter 3

Conclusion & perspectives

The objective of this thesis was to develop microarray solutions adapted to study and describe complex bacterial flora recovered from biological samples. My work consisted in designing, implementing and using a variety of software dedicated to the design and analysis of microarray experiments. Two distinct strategies relying on the ribosomal gene were developed: a ‘classical’ 16S rRNA gene array and a more complex phylogenetic microarray.

The development of a phylogenetic microarray is not a trivial task. This particular project required implementing a memory-optimized software solution able to deal with a large amount of input sequences, and to select probes respecting particular physico-chemical properties.

The efficiency of our design could be improved in several ways. Firstly, we could update on demand our design with the latest release of RDP, as well as use data emerging from various metagenomic projects. Secondly, probes targeting the Archaea superkingdom could be added to our probe set. Indeed, this taxon is often neglected in modern microbiology while its presence in biological samples is real. Consequently, the impact of Archaea presence is unknown. For example, it was speculated that syntrophic interactions (syntrophy: relationship between different two bacterial species growing in the same environment, in which one or both benefit nutritionally from the presence of the other) between Archaea and Bacteria may be an important feature of polymicrobial infections (Eckburg et al., 2003; Lepp et al., 2004). Additionally, we must keep in mind that our design is dependent on the quality of the sequences retrieved by the RDP. In this regard, short 16S rRNA gene sequences may be prone to misclassification at the genus level due to errors in the taxonomy or due to poor sequence quality. To bypass this issue,

we could simply consider filtering the 16S rRNA gene sequences retrieved from the RDP based on their length.

Nonetheless, our observations revealed that this approach is well-suited to perform relative quantitative analysis of bacteria flora, and is an appropriate alternative to the classical cloning-sequencing approach. Our phylogenetic microarray allows analysis of the bacterial contents of complex bacterial mixtures to detect both known and unknown microorganisms. The presence of unknown organisms can be suspected and mapped on the phylogenetic tree, indicating where to refine analysis. A variety of applications can be considered with this strategy, such as monitoring bacterial population dynamics or detecting taxa of peculiar interest (pathogenic taxa for example) in environmental or biological samples.

The development of the low-density 16S rRNA gene array required to design probes for 16S sequences that often exhibit strong similarities in particular for strain-specific probes. This required the fine-tuning of probe selection parameters while keeping good sensitivity and specificity. Also, a variety of computational tools were developed for this platform in order to deal with the large number of processed samples (n = 673). These tools included simple file parsers but also more complex analytical programs.

Our approach proved particularly well-adapted to monitor the gingival microbiota of a rather large collection of Nigerien children. Reassuringly, our results are in good congruence with the observations made with our phylogenetic array and with the cloning-sequencing study previously performed by Bolivar and colleagues (Bolivar et al., 2009).

With this array, we managed relative quantification of bacterial phylotypes in oral samples, thus observing differences in bacterial composition between the studied conditions. Absolute quantification of bacterial species is evidently possible, but preliminary calibration of every single probe would be a prerequisite as probe behavior is idiosyncratic.

To my knowledge, this low-density 16S rRNA gene microarray is the first one developed specifically to study the flora of a particular disease. Paster and colleagues (Colombo et al., 2009) developed the HOMIM (Human Oral Microbe Identification Microarray) platform but this 16S rRNA gene array is somewhat more global as it aims at simultaneous detecting the most prevalent oral bacterial species, without taking into account any particular disease. Initially intended for the study of bacterial diversity in oral samples, our approach can be extrapolated to the study of various microbiota associated

either with medical (intestinal or skin flora), environmental (soil, sludge, wastewaters, etc.), or food industry samples; and can be used as a faster and cheaper alternative to cloning sequencing of the ribosomal gene.

The development of low-density and phylogenetic arrays permitted studying the oral flora of Nigerien children in an effort to better understand the underlying microbiology of noma. Several noteworthy observations stand out of our study. Firstly, the bacterial diversity is clearly unbalanced between healthy and diseased conditions; lesional sites of noma showing a reduced bacterial diversity than gingival lesional sites. Secondly, as observed with principal component analysis and hierarchical clustering, there is no important difference between the bacterial profiles of noma and acute necrotizing gingivitis (ANG). These two observations comfort the consensus theory stipulating that ANG is the precursor of noma, and that an unidentified factor could act as a trigger for the disease. Another important observation is that *Fusobacterium necrophorum*, the long suspected causative agent of noma, was not characterized with noma lesions and should therefore be crossed-out from the list of potential causative agents. Another important feature of our results is that *Prevotella* spp., Spirochaetaceae, *Dialister pneumosintes* and *Peptostreptococcus* are more abundant on the lesional sites of both noma and gingivitis. In addition, a lower abundance of *Capnocytophaga*, Syntrophomonadaceae and Proteobacteria was monitored in noma lesions compared to gingivitis. Inversely, *Streptococcus pyogenes* and Cyanobacteria are more abundant in noma lesions. Cyanobacteria are typically associated with environmental milieus but since advanced noma lesions are open to the environment, it was not surprising to detect this phylotype. Concerning *S. pyogenes*, this strictly human pathogen can cause a variety of infections such as skin infections (impetigo), throat infections including pharyngitis and tonsillitis (Aziz and Kotb, 2008). *S. pyogenes* is also involved in necrotizing fasciitis, a subcutaneous infection with high mortality rates (Bisno and Stevens, 1996); and could potentially participate in the development of noma lesions by the release of various virulence factors such as streptolysin, proteases and exotoxins. Prevotellaceae, and more precisely *Prevotella intermedia*, are clearly associated with noma. The presence of *P. intermedia* in noma lesions was already documented in previous studies (Bolivar et al., 2009; Falkler et al., 1999a; Phillips et al., 1995), but was not detected in the 4 noma samples processed in the cloning sequencing study of Paster and colleagues (Paster et al., 2002). This pathogen is encountered in adult periodontitis (Slots, 1986; Wennström et al.,

1987) and is frequently isolated in endodontic infections (Gharbia et al., 1994; Tomazinho and vila-Campos, 2007; van Winkelhoff et al., 1985). It could take part in the etiology of noma by promoting tissue destruction by its ability to degrade lipids and produce proteolytic enzymes (Enwonwu et al., 2000; Slots et al., 1986). Moreover, *P. intermedia* produces immunoglobulin A1 (IgA1) proteases which could play a critical role in decreasing the mouth mucosal immunity (Frandsen et al., 1995), hence promoting development of other pathogens in oral lesions.

In conclusion, although no bacterial species was identified as the causative agent of noma, our study gives better insight on the bacteriology of this disease. Additionally, our results will be integrated in a large epidemiological study and the resulting observations will add to the understanding of the pathogenesis of noma.

The development of these two microarray required dealing with a large amount of sequence data during the design and analysis steps. Clearly, such a study would not have been possible without the extensive use of bioinformatics.

The two strategies reported here were applied in order to characterize the bacterial content of oral samples from a diseased condition, namely noma. Intuitively, these approaches could be conveniently used to study other complex bacterial samples from medical or environmental origins.

Microarray technology has been used in a large variety of applications such as gene expression, pathogen detection, SNP discovery, resequencing or comparative genome application. However, this technology suffers from several limitations: knowledge of the target sequence is the prerequisite for probe design; design of species-discriminating probes can be tricky; and finally, probe behavior remains only partly understood which can result in cross-hybridizations, requiring careful validation of microarray design by wet-lab experiments.

Over the past year, high throughput DNA sequencing technology has become more accessible to the scientific community and could, in the very near future, become an alternative to microarrays. However, microarrays will not necessarily be put aside as they could, for example, be exploited as a preparative platform to capture nucleic acids (Porreca et al., 2007; Wang et al., 2003). Microbiology is rapidly evolving and molecular methods are now extensively challenging conventional microbiology techniques. This field will benefit from this evolution as microarray technology appears to be the ideal tool

to assess the diversity of the microbial world as well as to improve our knowledge of host–pathogen interactions.

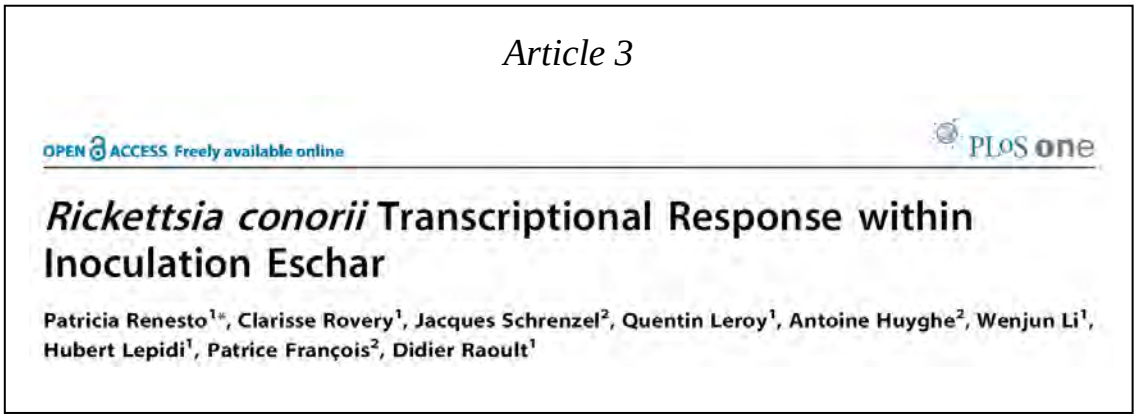
Chapter 4

Contributions

This chapter introduces other projects where I was directly involved in the design and/or analysis of whole-genome microarray experiments.

4.1 *Rickettsia conorii* transcriptional response within inoculation eschar.

Article 3



Background: *Rickettsia conorii*, the causative agent of the Mediterranean spotted fever, is transmitted to humans by the bite of infected ticks *Rhipicephalus sanguineus*. The skin thus constitutes an important barrier for the entry and propagation of *R. conorii*. Given this, analysis of the survival strategies used by the bacterium within infected skin is critical for our understanding of rickettsiosis.

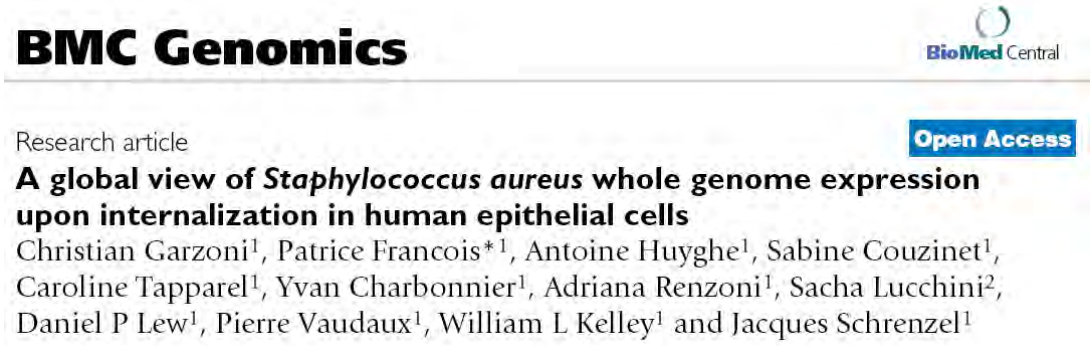
Findings: Here, we report the first genome-wide analysis of *R. conorii* gene expression from infected human skin biopsies. Our data showed that *R. conorii* exhibited a striking transcript signature that is remarkably conserved across patients, regardless of genotype. The expression profiles obtained using custom Agilent microarrays were validated by quantitative RT-PCR. Within eschars, the amount of detected *R. conorii* transcripts was of 55%, this value being of 74% for bacteria grown in Vero cells. In such infected host tissues, approximately 15% (n = 211) of the total predicted *R. conorii* ORFs appeared differentially expressed compared to bacteria grown in standard laboratory conditions. These genes are mostly down-regulated and encode proteins essential for bacterial replication. Some of the strategies displayed by rickettsiae to overcome the host defense barriers, thus avoiding killing, were also pointed out. The observed up-regulation of rickettsial genes associated with DNA repair is likely to correspond to a DNA-damaging agent enriched environment generated by the host cells to eradicate the pathogens. Survival of *R. conorii* within eschars also involves adaptation to osmotic stress, changes in cell surface proteins and up-regulation of some virulence factors. Interestingly, in contrast to down-regulated transcripts, we noticed that up-regulated ones rather exhibit a small nucleotide size, most of them being exclusive for the spotted fever group rickettsiae.

Conclusion: Because eschar is a site for rickettsial introduction, the pattern of rickettsial gene expression observed here may define how rickettsiae counteract the host defense.

Author's contribution: AH designed the microarray probe set, took part in the analysis of the gene expression experiments and participated in the writing of the manuscript.

4.2 A global view of Staphylococcus aureus whole genome expression upon internalization in human epithelial cells.

Article 4



Background: *Staphylococcus aureus*, a leading cause of chronic or acute infections, is traditionally considered an extracellular pathogen despite repeated reports of *S. aureus* internalization by a variety of non-myeloid cells in vitro. This property potentially contributes to bacterial persistence, protection from antibiotics and evasion of immune defenses. Mechanisms contributing to internalization have been partly elucidated, but bacterial processes triggered intracellularly are largely unknown.

Results: We have developed an in vitro model using human lung epithelial cells that shows intracellular bacterial persistence for up to 2 weeks. Using an original approach we successfully collected and amplified low amounts of bacterial RNA recovered from infected eukaryotic cells. Transcriptomic analysis using an oligoarray covering the whole *S. aureus* genome was performed at two post-internalization times and compared to gene expression of non-internalized bacteria. No signs of cellular death were observed after prolonged internalization of *Staphylococcus aureus* 6850 in epithelial cells. Following internalization, extensive alterations of bacterial gene expression were observed. Whereas major metabolic pathways including cell division, nutrient transport and regulatory processes were drastically down-regulated, numerous genes involved in iron scavenging and virulence were up-regulated. This initial adaptation was followed by a transcriptional increase in several metabolic functions. However, expression of several toxin genes known to affect host cell integrity appeared strictly limited.

Conclusion: These molecular insights correlated with phenotypic observations and demonstrated that S. *aureus* modulates gene expression at early times post infection to promote survival. *Staphylococcus aureus* appears adapted to intracellular survival in non-phagocytic cells.

Author's contribution: AH took part in the design and analysis of the microarray.

4.3 Exploring glycopeptide-resistance in *Staphylococcus aureus*: a combined proteomics and transcriptomics approach for the identification of resistance-related markers.

Article 5

BMC Genomics

BioMed Central

Research article

Open Access

Exploring glycopeptide-resistance in *Staphylococcus aureus*: a combined proteomics and transcriptomics approach for the identification of resistance-related markers

Alexander Scherl[†1], Patrice François[*†2], Yvan Charbonnier[2], Jacques M Deshusses[1], Thibaud Koessler[2], Antoine Huyghe[2], Manuela Bento[2], Jianru Stahl-Zeng[3], Adrien Fischer[2], Alexandre Masselot[4], Alireza Vaezzadeh[1], Francesca Gallé[2], Adriana Renzoni[2], Pierre Vaudaux[2], Daniel Lew[2], Catherine G Zimmermann-Ivol[1,5], Pierre-Alain Binz[4], Jean-Charles Sanchez[1], Denis F Hochstrasser[1,5,6] and Jacques Schrenzel[2]

BMC Genomics. 2006 Nov 22;7(1):296

Background: To unravel molecular targets involved in glycopeptide resistance, three isogenic strains of *Staphylococcus aureus* with different susceptibility levels to vancomycin or teicoplanin were subjected to whole-genome microarray-based transcription and quantitative proteomic profiling. Quantitative proteomics performed on membrane extracts showed exquisite inter-experimental reproducibility permitting the identification and relative quantification of >30% of the predicted *S. aureus* proteome.

Results: In the absence of antibiotic selection pressure, comparison of stable resistant and susceptible strains revealed 94 differentially expressed genes and 178 proteins. As expected, only partial correlation was obtained between transcriptomic and proteomic results during stationary-phase. Application of massively parallel methods identified one third of the complete proteome, a majority of which was only predicted based on genome sequencing, but never identified to date. Several over-expressed genes represent previously reported targets, while series of genes and proteins possibly involved in the glycopeptide resistance mechanism were discovered here, including regulators, global regulator attenuator, hyper-mutability factor or hypothetical proteins. Gene expression of these markers was confirmed in a collection of genetically unrelated strains showing altered susceptibility to glycopeptides.

Conclusion: Our proteome and transcriptome analyses have been performed during stationary-phase of growth on isogenic strains showing susceptibility or intermediate level of resistance against glycopeptides. Altered susceptibility had emerged spontaneously after infection with a sensitive parental strain, thus not selected in vitro. This combined analysis allows the identification of hundreds of proteins considered, so far as hypothetical protein. In addition, this study provides not only a global picture of transcription and expression adaptations during a complex antibiotic resistance mechanism but also unravels potential drug targets or markers that are constitutively expressed by resistant strains regardless of their genetic background, amenable to be used as diagnostic targets.

Author's contribution: AH took part in the design and analysis of the microarray.

4.4 Use of oligoarrays for characterization of community-onset methicillin-resistant Staphylococcus aureus.

Article 6

JOURNAL OF CLINICAL MICROBIOLOGY, Mar. 2006, p. 1040–1048 Vol. 44,
0095-1137/06/$08.00+0 doi:10.1128/JCM.44.3.1040–1048.2006

Use of Oligoarrays for Characterization of Community-Onset Methicillin-Resistant *Staphylococcus aureus*

Thibaud Koessler,[1] Patrice Francois,[1]* Yvan Charbonnier,[1] Antoine Huyghe,[1] Manuela Bento,[1] Sasi Dharan,[3] Gesuele Renzi,[2] Daniel Lew,[1,2] Stephan Harbarth,[3] Didier Pittet,[3] and Jacques Schrenzel[1,2]

Until recently, methicillin-resistant *Staphylococcus aureus* (MRSA) was considered the prototype of a hospital-acquired bacterial pathogen. However, recent reports have shown that MRSA has now emerged in the community. Characterization of specific markers for distinguishing the origin of isolates could contribute to improved knowledge of MRSA epidemiology. The release of whole-genome sequences of hospital- and community-acquired *S. aureus* strains allowed the development of whole-genome content analysis techniques, including microarrays. We developed a microarray composed of 8,191 open reading frame-specific oligonucleotides covering >99% of the four sequenced *S. aureus* genomes (N315, Mu50, MW2, and COL) to evaluate gene contents of hospital- and community-onset *S. aureus* strains. In parallel, pulsed-field gel electrophoresis, variable number of tandem repeats, antibiogram, staphylococcal cassette chromosome-mec element typing, and presence of the Panton-Valentine leukocidin gene were evaluated in a collection of 15 clinical isolates. Clusters obtained with microarrays showed a high degree of similarity with those obtained by pulsed-field gel electrophoresis or variable number of tandem repeats. Clusters clearly segregated hospital-onset strains from community-onset strains. Moreover, the microarray approach allowed definition of novel marker genes and chromosomal regions specific for given groups of isolates, thus providing better discrimination and additional information compared to pulsed-field gel electrophoresis and variable number of tandem repeats. Finally, the comparative genome hybridization approach unraveled the occurrence of multiple horizontal transfer events leading to

community-onset MRSA as well as the need for a specific genetic background in recipient strains for both the acquisition and the stability of the mec element.

Author's contribution: AH took part in the design and analysis of the microarray.

4.5 Correlation of proteomic and transcriptomic profiles of *Staphylococcus aureus* during the post-exponential phase of growth.

Article 7

Journal of Microbiological Methods 60 (2005) 247–257

Correlation of proteomic and transcriptomic profiles of *Staphylococcus aureus* during the post-exponential phase of growth

Alexander Scherl[a,1], Patrice François[b,*,1], Manuela Bento[b], Jacques M. Deshusses[a], Yvan Charbonnier[b], Véronique Converset[a], Antoine Huyghe[b], Nadia Walter[a], Christine Hoogland[c], Ron D. Appel[a,c,d], Jean-Charles Sanchez[a], Catherine G. Zimmermann-Ivol[a], Garry L. Corthals[a], Denis F. Hochstrasser[a,e], Jacques Schrenzel[b]

Abstract: A combined proteomic and transcriptomic analysis of *Staphylococcus aureus* strain N315 was performed to study a sequenced strain at the system level. Total protein and membrane protein extracts were prepared and analyzed using various proteomic workflows including: 2-DE, SDS-PAGE combined with microcapillary LC-MALDI-MS/MS, and multidimensional liquid chromatography. The presence of a protein was then correlated with its respective transcript level from *S. aureus* cells grown under the same conditions. Gene-expression data revealed that 97% of the 2'596 ORFs were detected during the post-exponential phase. At the protein level, 23% of these ORFs (591 proteins) were identified. Correlation of the two datasets revealed that 42% of the identified proteins (248 proteins) were amongst the top 25% of genes with highest mRNA signal intensities, and 69% of the identified proteins (406 proteins) were amongst the top 50% with the highest mRNA signal intensities. The fact that the remaining 31% of proteins were not strongly expressed at the RNA level indicates either that some low-abundance proteins were identified or that some transcripts or proteins showed extended half-lives. The most abundant classes identified with the combined proteomic and

transcriptomic approach involved energy production, translational activities and nucleotide transport, reflecting an active metabolism. The simultaneous large-scale analysis of transcriptomes and proteomes enables a global and holistic view of the *S. aureus* biology, allowing the parallel study of multiple active events in an organism.

Author's contribution: AH took part in the design and analysis of the microarray.

OPEN ACCESS Freely available online

PLoS one

Rickettsia conorii Transcriptional Response within Inoculation Eschar

Patricia Renesto[1]*, Clarisse Rovery[1], Jacques Schrenzel[2], Quentin Leroy[1], Antoine Huyghe[2], Wenjun Li[1], Hubert Lepidi[1], Patrice François[2], Didier Raoult[1]

1 Unité des Rickettsies, IRD-CNRS UMR 6236, Faculté de Médecine, Marseille, France, **2** Service of Infectious Diseases/Genomic Research Laboratory, University of Geneva Hospitals, Geneva, Switzerland

Abstract

Background: *Rickettsia conorii*, the causative agent of the Mediterranean spotted fever, is transmitted to humans by the bite of infected ticks *Rhipicephalus sanguineus*. The skin thus constitutes an important barrier for the entry and propagation of *R. conorii*. Given this, analysis of the survival strategies used by the bacterium within infected skin is critical for our understanding of rickettsiosis.

Methodology/Principal Findings: Here, we report the first genome-wide analysis of *R. conorii* gene expression from infected human skin biopsies. Our data showed that *R. conorii* exhibited a striking transcript signature that is remarkably conserved across patients, regardless of genotype. The expression profiles obtained using custom Agilent microarrays were validated by quantitative RT-PCR. Within eschars, the amount of detected *R. conorii* transcripts was of 55%, this value being of 74% for bacteria grown in Vero cells. In such infected host tissues, approximately 15% ($n = 211$) of the total predicted *R. conorii* ORFs appeared differentially expressed compared to bacteria grown in standard laboratory conditions. These genes are mostly down-regulated and encode proteins essential for bacterial replication. Some of the strategies displayed by rickettsiae to overcome the host defense barriers, thus avoiding killing, were also pointed out. The observed up-regulation of rickettsial genes associated with DNA repair is likely to correspond to a DNA-damaging agent enriched environment generated by the host cells to eradicate the pathogens. Survival of *R. conorii* within eschars also involves adaptation to osmotic stress, changes in cell surface proteins and up-regulation of some virulence factors. Interestingly, in contrast to down-regulated transcripts, we noticed that up-regulated ones rather exhibit a small nucleotide size, most of them being exclusive for the spotted fever group rickettsiae.

Conclusion/Significance: Because eschar is a site for rickettsial introduction, the pattern of rickettsial gene expression observed here may define how rickettsiae counteract the host defense.

Citation: Renesto P, Rovery C, Schrenzel J, Leroy Q, Huyghe A, et al. (2008) *Rickettsia conorii* Transcriptional Response within Inoculation Eschar. PLoS ONE 3(11): e3681. doi:10.1371/journal.pone.0003681

Editor: Aric Gregson, University of California Los Angeles, United States of America

Received July 23, 2008; **Accepted** October 20, 2008; **Published** November 10, 2008

Funding: This project was granted by a research program "Hospital Clinical Research Program" from the French Health Ministry. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: patricia.renesto@medecine.univ-mrs.fr

Introduction

Rickettsia conorii is a Gram-negative bacterium responsible for the Mediterranean spotted fever (MSF), a disease transmitted to humans by the brown dog tick *Rhipicephalus sanguineus* [1]. Inoculation of rickettsiae to human beings leads to vasculitis and lesions at the site of tick bite [2]. The cutaneous necrosis that results from severe injury to many small vessels and otherwise called the "tâche noire" is the hallmark of many spotted fever group rickettsioses [3]. The histological examination of eschars collected from patients suffering from boutonneuse fever indicated that the alterations were mainly located in the dermis and subcutaneous tissues and evidenced the presence of rickettsiae in blood vessels [3]. Thus, indirect immunofluorescent detection of *R. conorii* on cryostat sections of skin biopsy specimens from patients was found to improve the early diagnosis of severe and atypical forms of MSF [4,5].

Inoculation eschars correspond to the portal of entry of the infectious agent into the host and the first site of challenge between the infected human being and the bacterium. Within the first 24 hours after the tick attachment, the rickettsiae are already bloodborne and the observed rickettsiae in the tick feeding site and in particular within the eschar are left over rickettsiae undergoing clearance [2]. In this respect, the "tâche noire" was depicted as being an excellent, accessible model for the study of the human-*Rickettsia* interaction [6]. The intralesional expression of local mediators of inflammation and of immune response that could contribute both to anti-rickettsial immunity and the pathogenesis of the MSF, has recently been depicted in infected human tissues [7]. Analysis of the complementary picture, namely the survival strategies used by *R. conorii* within the inoculation site should provide a better understanding of rickettsial pathogenesis. While reports of global gene expression profiling in human tissue or non-invasive patient samples suffering from bacterial diseases are limited [8,9], we explored the RNA profiles of *R. conorii* from eschars collected on MSF patients. This study was made possible by applying a strategy combining removal of eukaryotic contaminants

Table 1. Clinical characteristics of patients included in this study.

sample	age	sex	number of eschars	year of eschar sampling	severe form	geographical site of bite	MST genotype	strain isolated
E	62	F	1	2005	No	France (13)	Nd	no
G	42	F	1	2005	No	Marocco	C	no
C	72	F	1	2004	Yes	France (84)	A	yes
I	67	M	1	2005	Yes	France (13)	A	no
F,D	61	F	2	2004	Yes	France (13)	A	yes
A,B	45	M	1	2006	No	France (13)	A	yes
H	30	M	1	2005	No	Algeria	B	yes

(nd) not determined.
doi:10.1371/journal.pone.0003681.t001

with subsequent random amplification of prokaryotic cDNA [10] that was found convenient for microarray-based transcriptome analysis of obligate intracellular rickettsiae [11]. To identify genes differentially regulated within eschars, rickettsial microarrays hybridized with cDNA obtained from *R. conorii* grown in Vero cells monolayers at 32°C were used as control. The results obtained offer new insights into *R. conorii* survival within the eschar site.

Results

The transcriptome profile of *R. conorii* is highly conserved among different eschars

The bacterial RNA samples purified from 8 eschars derived from 7 individual patients (Table 1) were subjected to whole-genome-wide transcript expression profile analysis and compared to the transcriptome of *R. conorii* grown *in vitro* and used as reference. As illustrated by Figure 1, human skin biopsies of MSF patients are histologically dominated by severe cutaneous necrosis with coagulative necrosis of the epidermis. Such specimens contain *R. conorii*, the bacteria being mostly found in necrotic areas associated with inflammatory cells. Given the scarce amount of available material, only one sample was hybridized twice. Hierarchical clustering of the signal intensities of the individual transcripts in both groups evidenced a high similarity of transcript expression patterns among eschars or infected cell monolayers, respectively. Measurements derived from similarity matrix indicated that largest distance among Vero cells profiles was 0.056 whereas a distance of 0.188 was found for the different eschars (Fig. 2). To accurately assess variations in gene expression within the group ⟨⟨eschar⟩⟩, the phylogenic analysis of the different clinical isolates was achieved with the multispacer typing (MST) based on the sequences of 3 variable intergenic spacers, namely *dksA-xerC*, *mppA-purC*, *rpmE-tRNA(fMet)* [12]. By combining the genotypes obtained from these three intergenic spacers, only 3 MST genotypes were obtained (Table 1). It is interesting to note that the two *R. conorii* isolates for which the transcriptome profiles appeared divergent from the main cluster (samples G and H) exhibited distinct MST genotypes.

The microarray results were confirmed by qRT-PCR for a subset of 16 targets. When comparing both methods, a high correlation coefficient ($R^2 = 0.934$) was observed (Fig. 3).

General overview

When compared with bacteria grown in Vero cell monolayers, the *R. conorii* gene transcripts in eschars were mainly found to be down-regulated. Of the 5,098 probes represented on our microarrays, 211 transcripts representing 15.4% (211/1374) of the *R. conorii* ORFs were differentially expressed ($P<0.05$). Of these, 180 genes were down-regulated two-fold or greater *in vivo* (supplementary Table S2) while only 31 genes exhibited an increased expression (Table 2). In the *R. conorii* genome, the size of the 1,374 annotated ORFs is ranging from 120 to 6,063 nucleotides (nt) with a mean value of 804 nt [13]. We noticed that 87% (27/31) of up-regulated genes within eschars have a size ranging from 153 nt to 723 nt, being thus significantly smaller than the median value.

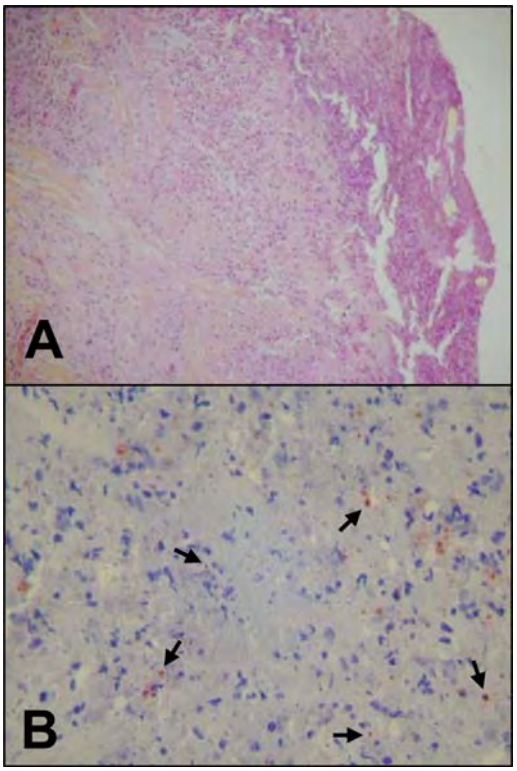


Figure 1. Inoculation eschar from a patient with MSF. (A) Histologic examination of inoculation eschar from a patient with MSF showing numerous dermal inflammatory infiltrates mainly composed of polymorphonuclear leukocytes with large necrotic areas and coagulative necrosis of the epidermis (hematoxylin-eosin-saffron, original magnification ×100). (B) Immunohistochemical detection of *R. conorii* in the inoculation eschar. Anti-*R. conorii* antibodies were detected using biotinylated secondary antibody, followed by avidin-peroxidase color development. The bacteria thus stained in reddish/brown, and indicated by the arrows, appear located between the necrotic inflammatory cells present in the dermis (original magnification ×400).
doi:10.1371/journal.pone.0003681.g001

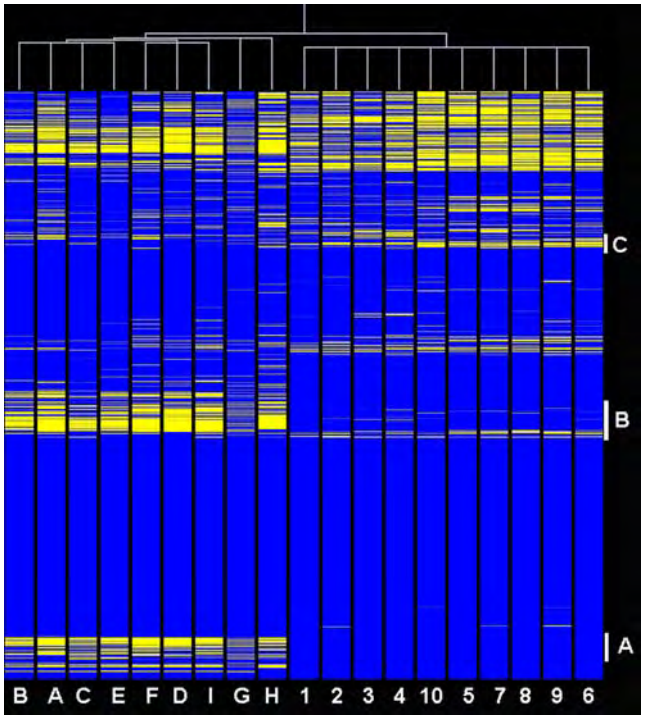


Figure 2. Two-way clustering analysis of *R. conorii* transcriptomic patterns. The figure shows a clear separation of two clusters containing 8 eschars recovered from different infectious episodes and 10 slides corresponding to bacteria grown in Vero cells and used as control, respectively. Each probe set is represented by a single row of colored boxes and each sample (eschar or control) correspond to a single column. The blue areas correspond to genes showing high or medium expression whereas yellow bars indicates genes poorly or not expressed. The dendrogram (white lines) on the top of the figure represents the similarity matrix of probe sets. Among clusters of genes allowing discriminating between the 2 tested conditions, A contains 70 genes involved mainly in different metabolic functions (transporters, DNA repair enzyme *mutL*, metabolic enzymes and numerous ribosomal proteins), cluster B contains an important number of genes involved in energy production (ATP synthesis) as well as genes involved in stress-response (*uvrA* and C and *htpG*), cell division and some virulence factors (*virB4* and *virB10*) and antibiotic resistance determinants. Cluster C contains several hem factors.
doi:10.1371/journal.pone.0003681.g002

We also observed that 21 of these genes, i.e. 67.7% are lacking or highly degraded in the typhus group (TG) rickettsiae. In contrast, the mean size of down-regulated genes was of 1,323 nt and 75% of them belong to the core gene of rickettsiae [14].

Evidence for rickettsial growth arrest within eschars

The ORFs of *R. conorii* whose expression was significantly altered were classified into functional categories according to the Cluster of Orthologous Gene (COG) classification, as defined by the COGs database [15]. As illustrated Figure 4, the most down-regulated COGs were translation (J), cell wall and membrane biogenesis (M), intracellular trafficking and secretion (U) as well as energy production and conversion (C). The reduced expression of 24 genes encoding ribosomal proteins indicates that, within eschars, rickettsiae are reducing their translational capacity (supplementary Table S2). Eighteen of these genes (from RC0981 to RC1007) are grouped together on the rickettsial chromosome and are organized as a ribosomal protein gene cluster. Consistent with a bacterial growth arrest, we also noticed a dramatic decrease in the expression of cell wall components including several Sca family proteins (Sca0, Sca1, Sca4, Sca5 and Sca10). Following 23SrRNA (105.5-fold decrease), the most down-regulated genes are those encoding the rickettsial outer membrane protein (rOmpB otherwise called Sca5: 98.5-fold decrease) and the rickettsial adhesin Adr2 (79.7-fold decrease). While down-regulated in a lower extend (fold-change mean of 5.4), many of the variable genes from the U category, including VirB, SecE, SecF and SecY subunits, encode proteins associated with the cell membrane. We also noticed the down-regulation of genes involved in energy production that should also contribute to bacterial replication arrest. Finally, 18% of down-regulated genes have no clearly defined function (S and R) or are not assigned to any functional categories.

R. conorii response to the host attack

Among the 31 rickettsial genes up-regulated in eschars, several are plausibly involved to escape host response. They include genes coding for proteins involved in DNA repair and modification as RC0550 (RecB exonuclease), RC1204 (tmRNA-binding protein), RC1050 (transposase) and RC1125 (helicase). Bacteria also face to oxidative stress, as indicated by the increase expression of *phbC* (poly-beta-hydroxybutyrate polymerase), *trxB2* (thioredoxin) and *grxC1* (glutaredoxin). An enhanced tolerance to osmotic stress could be provided through the increased expression of *proP9* (proline-betaine transporter) and *nuoL3* (NADH dehydrogenase chain I). Adaptation of *R. conorii* within eschars promoted changes in cell-wall related proteins as glycosyltransferase and cephalosporin hydroxylase (*cmcl*) and was accompanied with the up-regulation of several ATP binding cassette (ABC) transporters (RC0500, *mdlB*, *atm1*) and of two proteins that belong to the KAP NTPases, namely the NACHT NTPase and the putative AAA+ family ATPase, a new family characterized by the presence of transmembrane segments inserted into the P-loop NTPase domain [16]. We also noticed the over-expression of two genes encoding tetratricopeptide repeat (TPR)-containing proteins, suggesting the importance of this motif for protein-protein interactions between *R. conorii* and its host during infection, as previously evoked [17]. Apart from these, we also observed an increased expression of RC1370 and RC1298 genes that encode a prophage antirepressor and lysozyme, respectively. Finally, 11 out of the 31 up-regulated genes encode proteins with unknown function.

Discussion

Infection with *R. conorii* usually occurs following infected tick bite [2] and the first step of this host-pathogen interaction takes place at the inoculation site where anti-rickettsial immunity is enhanced [7]. The presence of an eschar in rickettsial disease is generally linked with a milder disease as more severe rickettsiosis, namely the Rocky Mountain Spotted Fever (RMSF) and epidemic typhus do not exhibit inoculation eschars [1,2,18]. This suggests that eschar corresponds to the front line of human host defense against rickettsia diffusion, a point consistent with the fact that the highest bacteremia were detected in patients suffering from RMSF and typhus [19]. Accordingly, the factors involved in the bacterial survival strategy can be considered as crucial actors of *R. conorii* pathogenesis. We thus examined the expression patterns of *R. conorii* in such environment, using an approach successfully applied for the analysis of transcriptional profile of bacteria within infected epithelial monolayers [11].

The human skin biopsies being rare and precious, the optimal conditions for total RNA extraction were first assayed on eschars experimentally induced through the intradermal injection of *R. conorii* in rabbits (not shown). Starting material used in this study

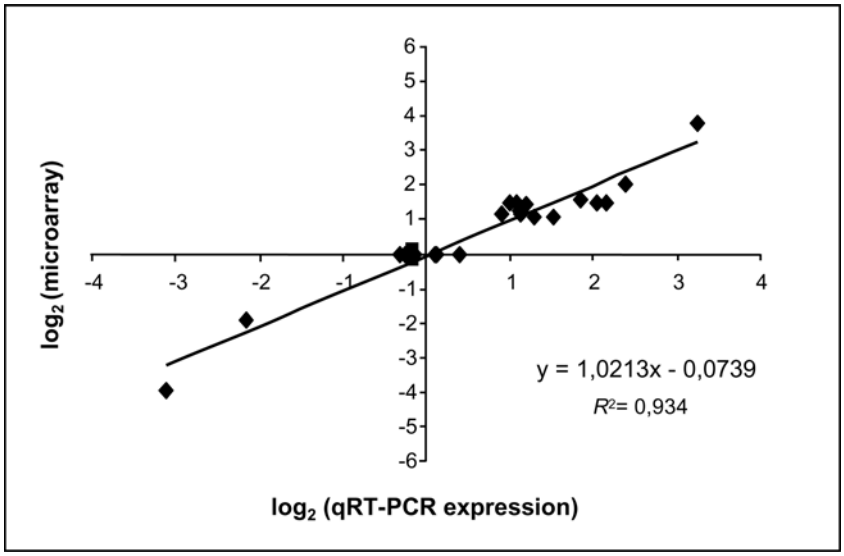


Figure 3. Validation of microarray-based expression profiles by qRT-PCR. The relative transcriptional levels for 16 genes were determined by real-time qRT-PCR using cDNA as template. Following normalization of data based on values measured with non regulated transcripts, the real-time qPCR $\log_2$ values were plotted against the microarray $\log_2$ values. The correlation coefficient (R^2) for comparison of the two datasets is of 0.934.
doi:10.1371/journal.pone.0003681.g003

was collected at different time points after the tick bite and stored at -80°C for variable periods of time (up to 3 years) before processing. We predicted that this factor would interfere with the identification of differences in gene expression because of biological variations. In addition, the pattern of gene expression may also vary depending on the genetic background of infecting strains. Indeed, in microarray-based experiments, a lower hybridization can result either from a reduced amount of transcripts or from divergence in the sequence of the gene [20]. Thus, insertion or deletion events in the genome of clinical strains tested could affect the apparent transcript abundance measured by using a microarray designed from the genome sequence of the Malish strain (seven) of *R. conorii* [13]. Here, three *R. conorii* strains corresponding to cases issued from three different geographical sites were identified. However, analysis of rickettsial transcript expression patterns from the 8 human eschars included in this study yielded reproducible results and only minor variations were observed between biopsy specimens. Finally, results obtained by qRT-PCR indicated that reliable microarray hybridizations can be achieved with rickettsial RNA extracted from multiple and independently obtained eschars.

This work showed that within eschars, the *R. conorii* transcripts were mainly down-regulated compared to bacteria internalized in Vero cells. The most significantly repressed genes are those of the translation machinery. This observation is consistent with previous analysis of the transcriptional changes displayed by 19 different bacterial pathogens upon eukaryotic cell infection [8]. Within these conditions, a general decline of genes involved in general metabolism associated with bacterial growth (translation, transcription, cell wall biogenesis, energy production, transport of carbohydrates, amino acids and nucleotides) was observed. Such a global expression decrease resembles that depicted for *Bacillus subtilis* [21], *Escherichia coli* [22], *Corynebacterium glutamicum* [23], and *Staphylococcus aureus* [24] after inducing the stringent response [25]. The recent analysis of inflammatory and immune mediators present in skin-biopsy samples of patients suffering from MSF evidenced the production of some enzymes including inducible nitric oxide synthase and indoleamine-2,3-dioxygenase, that should contribute to the bacterial growth arrest [7]. However, we observed that in some cases, the immune and inflammatory host response was therefore not strong enough to eradicate all infecting *R. conorii* as the bacterium was cultivable from 5 (62.5%) of the eschar biopsy specimens.

Analysis of obtained data highlighted some of the mechanisms displayed by the rickettsiae to counter the damage and survive within the host cells. Genes encoding proteins involved in genome repair and allowing bacteria to cope with DNA-damaging agents are probably critical in this situation and were up-regulated. DNA repair is a fundamental process used by pathogenic bacteria as one of the defense mechanisms that allow them to survive in their hosts [26]. As described for *Helicobacter pylori* [27], the DNA lesions could result from a cellular oxidative stress, a point consistent with the increased rickettsial defenses against reactive oxygen species. Noteworthy is that production of bactericidal reactive oxygen species is one of the key methods by which mammalian infected cells efficiently kill bacteria [28]. Within eschars, *R. conorii* also deal with osmotic stress as indicated by the up-regulation of proline-betaine transporter and of NADH dehydrogenase I. Proline and betaine are two osmoprotectants accumulated through enhanced uptake rather than synthesis by Gram negative bacteria to overcome the inhibitory effects of hyperosmolarity [29]. While the *nuo* genes of rickettsiae are more closely related to mitochondria than to any other studied microbe [30], enhancement of NADH dehydrogenase expression could correspond to another strategy of osmoadaptation evolved to achieve salt tolerance. *R. conorii* also displayed another hallmark feature of pathogens interacting with host cells, namely the phenotypic changes in the composition of several membrane proteins [8]. It is well established that antigenic variation is an important mechanism that allows pathogens to escape for immunity. In this respect, we noticed an important down-regulation of genes encoding for the Sca family proteins among which Sca0 (rOmpA) and Sca5 (rOmpB) which are two major rickettsial antigenic determinants [31]. Antigenic variation could also be related to the post-translational modifications of proteins. These changes could be afforded by Cmcl, a protein recently classified as a methyltrans-

Table 2. Thirty one genes of *R. conorii* up-reguled within eschars.

ORFs	Genes	Description	Fold-Ch	Size (nt)
RC0550		RecB family exonuclease	13,685	2,517
RC0828**		**Unknown**	**6,763**	**183**
RC0500		**ABC transporter ATP-binding protein**	**4,019**	**723**
RC1370*		**Prophage antirepressor (SPLIT GENE)**	**3,790**	**207**
RC0769**		**Unknown**	**3,661**	**156**
RC1137	*phbC*	**Poly-beta-hydroxybutyrate polymerase (SPLIT GENE)**	**3,628**	**264**
RC1298**		**Lysozyme (FRAGMENT)**	**3,405**	**204**
RC0350**		**Unknown**	**3,225**	**240**
RC1204	*smpA*	**tmRNA-binding protein**	**3,068**	**474**
RC0299*	*mdlB*	**ABC-type multidrug transport system, ATPase and permease components (SPLIT GENE)**	**2,986**	**273**
RC1299**		**Unknown**	**2,972**	**165**
RC0921**		**Unknown**	**2,825**	**153**
RC1349**		**Unknown**	**2,814**	**201**
RC0637	*trxB2*	**Thioredoxin reductase [EC:1.6.4.5]**	**2,785**	**1,023**
RC0267	*grxC1*	**Glutaredoxin**	**2,780**	**309**
RC0461*		**Glycosyltransferase [EC:2.4.1.-], two domains**	**2,708**	**1,815**
RC1050**		**Transposase (FRAGMENT)**	**2,655**	**171**
RC1125*		**Superfamily I DNA and RNA helicases (SPLIT GENE)**	**2,617**	**192**
RC0378	*nuoL3*	**NADH dehydrogenase I chain L [EC:1.6.5.3]**	**2,507**	**1,518**
RC1251		**Unknown**	**2,390**	**504**
RC0221**	*cmcI*	**Cephalosporin hydroxylase (FRAGMENT)**	**2,342**	**321**
RC0914		**Tetratricopeptide repeat-containing protein**	**2,271**	**264**
RC0957*		**Tetratricopeptide repeat-containing protein (FRAGMENT)**	**2,263**	**228**
RC1155*		**Unknown (SPLIT GENE)**	**2,256**	**348**
RC0209*		**Unknown**	**2,225**	**282**
RC0795**		**NACHT family NTPase (FRAGMENT)**	**2,146**	**213**
RC0890*	*pro*P9	Proline/betaine transporter (FRAGMENT)	2,102	515
RC0269	*atm1*	Multidrug resistance protein Atm1 (SPLIT GENE)	2,085	201
RC0060**		**Unknown**	**2,067**	**492**
RC1144**		**Unknown**	**2,066**	**231**
RC1192**		**Putative AAA+ superfamily ATPase (SPLIT GENE)**	**2,061**	**183**

Genes absent (**) or degraded (*) in the TG.
doi:10.1371/journal.pone.0003681.t002

ferase based on structural evidences [32], or by a glycosyltransferase. Because most glycoproteins were associated with virulence factors in bacterial pathogens [33], these events could contribute to differences in both virulence and antigenicity of *R. conorii in vivo*.

Clearly, besides evasion of host defense, rickettsiae also exhibited virulence determinants *in vivo*. Thus, the export of virulence proteins in the host cell cytoplasm could be achieved through the increased expression of ABC transporters that function as type I secretion system in Gram negative bacteria [34]. The way in which rickettsiae use the KAP NTPases in the intracellular host cell environments has not yet been investigated but several members of this family, namely the AAA+ ATPases were described to promote virulence of other bacterial pathogens [35]. A role for the α-superhelical structure domain of KAP NTPases has been evoked in protein-protein interactions [16]. Interestingly, such interactions can also be mediated by the TPR [36], another structural motif present within two *R. conorii* proteins up-regulated *in vivo*. Here again, the functional role of these ubiquitous domains was not deciphered. However, TPR repeat regions have been implicated in the ability of *L. pneumophila* to efficiently establish infection and/or to manipulate host cell trafficking event [37]. The whole picture of this host-pathogen interaction within the first step of infection is summarized Figure 5.

A more general view of obtained results showed that, to ensure microbial fitness and survival in the lethal host environment, *R. conorii* mainly promoted the transcription of small size genes. While a high proportion of them codes for proteins with unknown functions, those with functional attributes including *phbC*, *mdlB* and *proP9* have already been associated with pathogenicity in other bacteria. Another interesting feature is that up-regulated genes are mostly exclusive for SFG rickettsiae, a finding consistent with the fact that the inoculation site corresponds to environmental conditions not encountered by TG bacteria that are transmitted to human by exposure to feces of infected lice or fleas [1].

Recent studies showed that an understanding of the basic mechanisms of adaptation of rickettsiae under various environ-

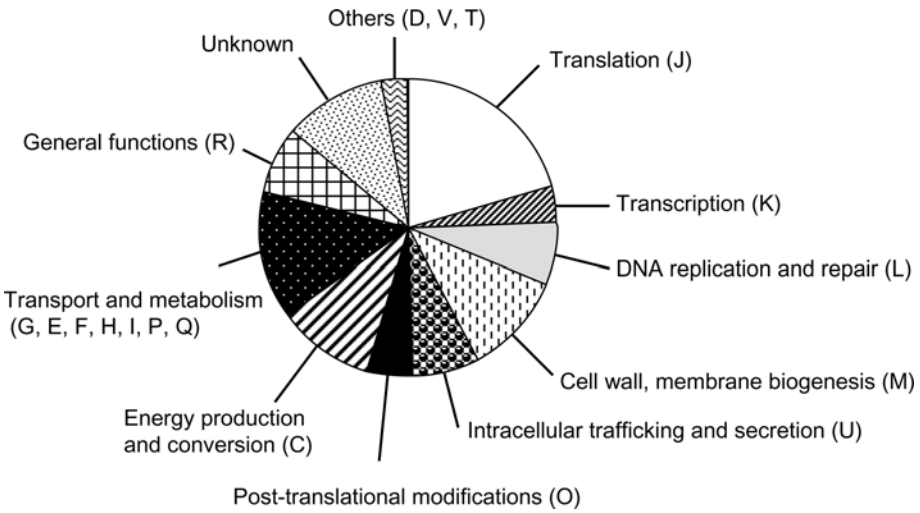


Figure 4. Distribution of *R. conorii* ORFs down-regulated during infection. Pie chart showing the percentages of transcripts down-regulated in eschars and classified according to their respective functional categories. COG category legend is as follow: Cell cycle control, mitosis and meiosis (D); Defense mechanisms (V); Signal transduction mechanisms (T); Transport and metabolism of carbohydrates (G), amino acids (E), nucleotides (F), coenzymes (H), lipids (I), inorganic ions (P), and of secondary metabolites (Q).
doi:10.1371/journal.pone.0003681.g004

ments, including nutrient deprived medium [11] and temperature changes [38] can be gained by determining their global transcriptome profile by microarrays. As shown here, this approach also offers the opportunity to characterize strategies displayed by the bacteria for *in vivo* survival. A new picture of *Rickettsia* pathogenicity, still poorly known, has thus emerged. A better knowledge of such an host-pathogen interaction would offer the opportunity to identify future therapeutic targets usefull for the prevention or the treatment of rickettsiosis.

Materials and Methods

Human subjects and sample collection

From June 2004 through August 2006, cutaneous biopsy specimens were obtained from seven patients in Marseille with eschar-associated illnesses, and who were suspected of suffering from MSF on the basis of initial clinical findings or laboratory analysis. These biopsy specimens, excised by using a scalpel, were dissected in two pieces. For RNA extraction purpose, the tissues

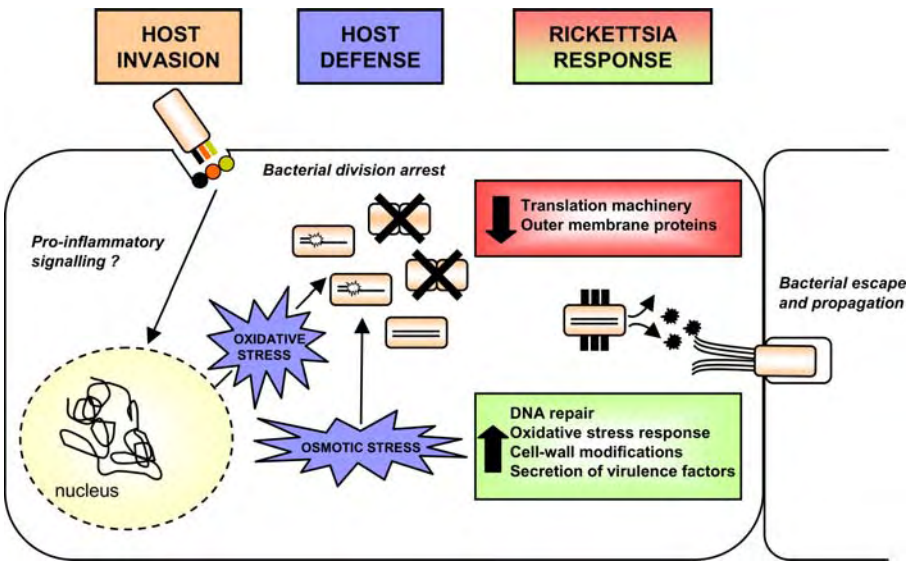


Figure 5. Schematic representation of the interaction between *R. conorii* and its host cell at the site of inoculation. Intracellular uptake of *R. conorii* by induced phagocytosis [46] may activate signalling pathways in the host cells which in turn display several mechanisms to eliminate invading bacteria among which oxidative and osmotic stress. As shown by the work presented in this paper, these events induce a strong down-regulation of *R. conorii* transcripts, mainly of those involved in bacterial replication and classified within translation and cell wall membrane COGs. These pathogenic bacteria therefore evade the host defense through the up-regulation of several factors counteracting DNA damages and through variations of dominant surface antigens allowing to avoid host recognition. The secretion of virulence determinants is also likely to favour survival and colonization of the host.
doi:10.1371/journal.pone.0003681.g005

were collected in sterile Eppendorf tubes containing RNA*later* Stabilization Reagent (Qiagen, Courtaboeuf, France) refrigerated at 4°C overnight before storage at −80°C. The second half of specimens was processed for bacterial culture and PCR-based molecular diagnosis, histopathology as well as immunohistochemical staining after fixation in ethanol. Selected clinical variables of the individual patients are shown in Table 1. These experiments were carried out with both the approval from the local ethics committee (IFR 48 ethics committee) and the written consent of informed patients.

Histologic analysis and immunohistochemical detection of *R. conorii*

Formalin-fixed, paraffin-embedded skin biopsy specimens of the inoculation eschars were cut (3 µm thickness) and stained with hematoxylin-eosin-saffron by routine staining methods. Serial sections of each tissue specimen were also obtained for immunohistochemical investigations. The immunohistochemical analysis was performed by the indirect immunoperoxidase method has described elsewhere [39] and using a polyclonal rabbit antibody against *R. conorii* as primary antibody. Briefly, after deparaffinization, each tissue section was incubated with the polyclonal anti-*R. conorii* antibody (1:2000) and immunodetection was performed with biotinylated immunoglobulins, followed by peroxidase-labeled streptavidin (HistoStain plus kit, Zymed, Montrouge, France) with amino-ethyl-carbazole as substrate. The slides were counterstained with Mayer hematoxylin for 10 min. Images were acquired with a Zeiss Axioskop microscope coupled with a Nikon Coolpix 4500 digital camera.

Multispacer genotyping (MST) of *R. conorii* clinical isolates

Total genomic DNA was extracted from skin-biopsy specimens using the QIAamp Tissue kit (Qiagen), as described by the manufacturer. Amplification of the three highly variable intergenic spacers, *dksA-xerC*, *mppA-purC*, and *rpmE-tRNA^{fMet}* was achieved using previously described primers [12] and HotStarTaq DNA polymerase (Qiagen). All primers were obtained from Eurogentec (Seraing, Belgium). PCR was carried out in a PTC-200 automated thermal cycler (MJ Research, Waltham, Mass.) under the following conditions: an initial 15 min-step at 95°C was followed by 39 cycles of denaturation for 30 s at 94°C, annealing for 30 s at 54°C, and extension for 1 min at 72°C. Final amplification was completed by holding the reaction mixture for 5 min at 72°C to allow complete extension of the PCR products. Following purification (MultiScreen PCR filter plate, Millipore, Saint-Quentin en Yvelines, France), amplicons were sequenced in both directions using the BigDye 1.1 chemistry (Applied Biosystems) on an ABI 3130XL automated sequencer (Applied Biosystems) as described by the manufacturer. To avoid contamination, no positive control was used. Sterile water was used as a negative control in each PCR assay. Sequences from each DNA sample were checked twice in both directions to ensure the reliability of the MST method, then assembled and edited with the Sequencher 4.7 program (GeneCode, Ann Arbor, Mich.). For the phylogenetic analysis, the sequences of three spacers were concatenated. Multiple alignments of the concatenated spacer sequences were carried out using the CLUSTALW software [40]. Phylogenetic relationships were obtained using the neighbor-joining and maximum parsimony methods within the MEGA 4.1 software [41].

Microarray design

ArrayDesigner^{TM} (Premier Biosoft) was used to generate an initial set of probes covering the whole genome of *R. conorii* strain Malish 7 (NC_003103), showing specific physico-chemical properties: (i) 71±12° C target Tm, (ii) 40–60 bp probe length, (iii) <−5.0 kcal/mol for hairpins, (iv) <−8.0 kcal/mol for self-dimers, and (v) dinucleotide repeats shorter than 5 bp. Candidate probes were tested for specificity against the aforementioned *R. conorii* genome using Olicheck [42]. To avoid cross-hybridization with host cell nucleic acids, the resulting set of probes was subjected to Blast analysis against the *Homo sapiens* and *Pan troglodytes* (by default of the African green monkey genome, Vero cells). Probes with >18 consecutive nt matches were excluded. Preceding steps yielded a final oligonucleotide set of 5,098 probes resulting to a final coverage of 97% for ORFs and 63% for inter-ORFs (considering fragment with length >149 bp which is the median size of inter-ORFs). To minimize steric hindrance, all probes <60 nt in length were poly(T)-tailed to reach an overall length of 60 nt, following Agilent's recommendations. Microarrays (2×11K format) were manufactured by *in situ* synthesis SurePrint technology (Agilent Technologies, Palo Alto, CA). All specific oligonucleotides as well as Agilent's control spots were printed in duplicate.

RNA isolation and purification from eschars

Total RNA extraction and purification from the eschar biopsies was carried out using the RNeasy kit (Qiagen) with some modifications. Briefly, the tissue samples removed from RNA*later* were rapidly decontaminated with iodated alcohol. After 2 washings in RNase-free water, 20 mg of tissue excised in small pieces were homogenized in RLT solution with tungsten beads and using the Mixer Mill MM300 (Qiagen). The resulting homogenate was then incubated at 55°C for 10 min with proteinase K (200 µg/ml) and centrifuged for 3 min at 10,000×*g*. Supernatant containing the total RNA fraction was supplemented with ethanol and purified onto RNeasy columns according to the manufacturer's instructions. The amount and quality of obtained RNA were determined with the microfluidic-based platform (Agilent 2100 Bioanalyzer) and using the RNA 6000 Nano Labchip kit (Agilent). In the electropherograms obtained with total RNA, the prokaryotic fraction was not always detected because its low abundance and the profiles of eukaryotic peaks were used as indicators for the integrity of both RNA populations. Estimated amount of total RNA extracted from 20 mg eschar was generally around 10 µg, including 90% of eukaryotic RNAs. These contaminants were removed using the MICROBEnrich procedure (Ambion, Applied Biosystems, Courtaboeuf, France) and prokaryotic cDNA was synthesized as described [10] using the M-MLV Reverse Transcriptase (Invitrogen, Cergy-Pontoise, France).

RNA isolation and purification from infected Vero cells

R. conorii strain Malish 7 (ATCC, VR613) grown in Vero cells for 48 h at 32°C were lysed in TE buffer (10 mM Tris/HCl, 1 mM EDTA pH 7.0) supplemented with lysozyme (10 mg/ml) for 10 minutes at room temperature. Total RNA was then extracted and purified using RNeasy-Midi columns (Qiagen), as previously described [11]. Two batches bacteria grown separately (biological replicates) were collected for this study. We then applied the same procedure as for eschar samples for purification, amplification, labelling and hybridization of *R. conorii* cDNA.

Expression microarrays and analysis

Twenty nanograms of cDNA were amplified using the GenomiPhi DNA amplification kit (Amersham Biosciences, Uppsala, Sweden) and labeled with Cy3-dCTP or Cy5-dCTP dyes (Amersham Biosciences) using the BioPrime DNA labeling System (Invitrogen, Cergy-Pontoise, France). Following purifica-

tion with GFX columns (Amersham Biosciences), the levels of Cy3-dCTP and Cy5-dCTP incorporation were quantified by absorbance measurement at 550 nm and 650 nm, respectively. Hybridizations were performed for 17 h at 60°C in dedicated micro-chambers with 75 pmol of both control or eschar samples. Stringent washings were then performed according to manufacturer's instructions. Slides dried by 30 sec washing with acetonitrile were scanned using 100% Photon Multiplier Tube power for both wavelengths using Agilent scanner (Agilent Technologies, CA, USA) and extracted using Feature Extraction™ software (version 8, Agilent). Local background subtracted signals were corrected for unequal dye incorporation or unequal load of the labeled product and normalized using GeneSpring (Agilent). The algorithm consisted of a rank consistency filter and a curve fit using the default LOWESS (locally weighted linear regression) method. Data consisting of 10 independent control conditions hybridized against 10 independent patient eschar experiments were expressed as Log10 ratios and analyzed using GeneSpring 7.3 (Agilent). Statistical significance of differentially expressed genes was identified by variance analysis (ANOVA) [43,44] performed using GeneSpring, including the Benjamini and Hochberg false discovery rate correction (5%). Expression microarrays (normalized data) were clustered by a hierarchical clustering algorithm by using an average linkage method in GeneSpring. The expression values for a gene across all samples were standardized to have mean of 0 and standard deviation of 1 by linear transformation. To determine the amount of detectable genes, the expression values were averaged for transcripts mapped by 2 or more probes. A cut-off value defined as 2×standard deviation obtained for background intensities was then applied [45].

Microarray data accession numbers

The data have been deposited in NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/). The GEO accession numbers are GPL7040 for the platform of the microarray and GSE12130 for the experimental data set.

Real-time quantitative PCR

Validation of microarray data was achieved using cDNA synthesized from eschar specimens A/B, C and E and from *R. conorii* grown in Vero cells used as reference. Real-time quantitative RT-PCR (qRT-PCR) was performed on the SmartCycler system (Cepheid) together with the QuantiTec Probe PCR kit (Qiagen) or SybrGreen DNA Fast Start kit (Roche Diagnostics, Basel, Switzerland), as indicated. Gene-specific primers are listed in the Supplementary Table S1. The values obtained for the 3 non-differentially expressed genes (*trxB1*, *glyQ* and *dnaK*) were used to normalize all data. The fold change (FC) in expression of the target genes relative to the 3 unregulated genes was determined as follows: $FC = 2^{-\Delta\Delta Ct}$ where $\Delta\Delta Ct = (\text{Mean-}Ct_{target} - \text{Mean-}Ct_{control})_{eschars} - (\text{Mean-}Ct_{target} - \text{Mean-}Ct_{control})_{reference}$. Ct values were defined as the cycle numbers at which the fluorescence signals were detected.

Supporting Information

Table S1 List of primers and conditions for qRT-PCR assays
Found at: doi:10.1371/journal.pone.0003681.s001 (0.04 MB XLS)

Table S2 R. conorii genes down-regulated within eschars
Found at: doi:10.1371/journal.pone.0003681.s002 (0.08 MB XLS)

Author Contributions

Conceived and designed the experiments: PR CR JS DR. Performed the experiments: CR QL WL HL. Analyzed the data: PR QL AH WL HL PF. Contributed reagents/materials/analysis tools: JS AH DR. Wrote the paper: PR PF DR.

References

1. Raoult D, Roux V (1997) Rickettsioses as paradigms of new or emerging infectious diseases. Clin Microbiol Rev 10: 694–719.
2. Parola P, Paddock CD, Raoult D (2005) Tick-borne rickettsioses around the world: emerging diseases challenging old concepts. Clin Microbiol Rev 18: 719–756.
3. Montenegro MR, Mansueto S, Hegarty BC, Walker DH (1983) The histology of "taches noires" of boutonneuse fever and demonstration of *Rickettsia conorii* in them by immunofluorescence. Virchows Arch A Pathol Anat Histopathol 400: 309–317.
4. Raoult D, de Micco C, Gallais H, Toga M (1984) Laboratory diagnosis of Mediterranean spotted fever by immunofluorescent demonstration of *Rickettsia conorii* in cutaneous lesions. J Infect Dis 150: 145–148.
5. Dujella J, Morovic M, Dzelalija B, Gveric M, Novakovic S (1991) Histopathology and immunopathology of skin biopsy specimens in Mediterranean spotted fever. Acta Virol 35: 566–572.
6. Walker DH, Occhino C, Tringali GR, Di Rosa S, Mansueto S (1988) Pathogenesis of rickettsial eschars: the tache noire of boutonneuse fever. Hum Pathol 19: 1449–1454.
7. de Sousa R, Ismail N, Nobrega SD, França A, Amaro M, et al. (2007) Intralesional expression of mRNA of interferon- gamma , tumor necrosis factor-alpha , interleukin-10, nitric oxide synthase, indoleamine-2,3-dioxygenase, and RANTES is a major immune effector in Mediterranean spotted fever rickettsiosis. J Infect Dis 196: 770–781.
8. La MV, Raoult D, Renesto P (2008) Regulation of whole bacterial pathogen transcription within infected hosts. FEMS Microbiol Rev 32: 440–460.
9. Waddell SJ, Butcher PD, Stoker NG (2007) RNA profiling in host-pathogen interactions. Curr Opin Microbiol 10: 297–302.
10. Francois P, Garzoni C, Bento M, Schrenzel J (2007) Comparison of amplification methods for transcriptomic analyses of low abundance prokaryotic RNA sources. J Microbiol Methods 68: 385–391.
11. La MV, Francois P, Rovery C, Robineau S, Barbry P, et al. (2007) Development of a method for recovering rickettsial RNA from infected cells to analyze gene expression profiling of obligate intracellular bacteria. J Microbiol Methods 71: 292–297.
12. Fournier PE, Zhu Y, Ogata H, Raoult D (2004) Use of highly variable intergenic spacer sequences for multispacer typing of *Rickettsia conorii* strains. J Clin Microbiol 42: 5757–5766.
13. Ogata H, Audic S, Renesto-Audiffren P, Fournier PE, Barbe V, et al. (2001) Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. Science 293: 2093–2098.
14. Blanc G, Ogata H, Robert C, Audic S, Suhre K, et al. (2007) Reductive genome evolution from the mother of *Rickettsia*. PloS Genet 3: e14.
15. Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution 225. Nucleic Acids Res 28: 33–36.
16. Aravind L, Iyer LM, Leipe DD, Koonin EV (2004) A novel family of P-loop NTPases with an unusual phyletic distribution and transmembrane segments inserted within the NTPase domain. Genome Biology 5: R30.
17. Ogata H, Robert C, Audic S, Robineau S, Blanc G, et al. (2005) *Rickettsia felis*, from culture to genome sequencing. Ann NY Acad Sci 1063: 26–34.
18. Thorner AR, Walker DH, Petri WAJ (1998) Rocky mountain spotted fever. Clin Infect Dis 27: 1353–1359.
19. Tzianabos T, Anderson BE, McDade JE (1989) Detection of *Rickettsia rickettsii* DNA in clinical specimens by using polymerase chain reaction technology. J Clin Microbiol 27: 2866–2868.
20. Le Gall T, Darlu P, Escobar-Páramo P, Picard B, Denamur E (2005) Selection-driven transcriptome polymorphism in *Escherichia coli/Shigella* species. Genome Res 15: 260–268.
21. Eymann C, Homuth G, Scharf C, Hecker M (2002) *Bacillus subtilis* functional genomics: global characterization of the stringent response by proteome and transcriptome analysis. J Bacteriol 184: 2500–2520.
22. Chang D-E, Smalley DJ, Conway T (2002) Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. Mol Microbiol 45: 289–306.
23. Brockmann-Gretza O, Kalinowski J (2006) Global gene expression during stringent response in Corynebacterium glutamicum in presence and absence of the rel gene encoding (p)ppGpp synthase. BMC Genomics 7: 230.
24. Garzoni C, Francois P, Huyghe A, Couzinet S, Tapparel C, et al. (2007) A global view of *Staphylococcus aureus* whole genome expression upon internalization in human epithelial cells. BMC Genomics 8: 171.
25. Cashel M, Gentry DR, Hernandez VJ, Vinella D (1996) The stringent response. In: Neidhardt FC, Curtis III R, Ingraham JL, Lin ECC, Low KB, et al., eds.

Escherichia coli and *Salmonella*: Cellular and Molecular Biology. Washington, D.C.: ASM Press. pp 1458–1496.
26. Heithoff DM, Conner CP, Mahan MJ (1997) Dissecting the biology of a pathogen during infection. Trends Microbiol 5: 509–513.
27. O'Rourke EJ, Chevalier C, Pinto AV, Thiberge JM, Ielpi L, et al. (2003) Pathogen DNA as target for host-generated oxidative stress: role for repair of bacterial DNA damage in *Helicobacter pylori* colonization. Proc Natl Acad Sci U S A 100: 2789–2794.
28. Radtke AL, O'Riordan MX (2006) Intracellular innate resistance to bacterial pathogens. Cell Microbiol 8: 1720–1729.
29. Csonka LN (1989) Physiological and genetic responses of bacteria to osmotic stress. Microbiol Rev 53: 121–147.
30. Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, et al. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396: 133–140.
31. Teysseire N, Raoult D (1992) Comparison of Western immunoblotting and microimmunofluorescence for diagnosis of Mediterranean spotted fever. J Clin Microbiol 30: 455–460.
32. Oster LM, Lester DR, Terwisscha van Scheltinga A, Svenda M, van Lun M, G, et al. (2006) Insights into cephamycin biosynthesis: the crystal structure of CmcI from *Streptomyces clavuligerus*. J Biol Mol 358: 546–558.
33. Schmidt MA, Riley LW, Benz I (2003) Sweet new world: glycoproteins in bacterial pathogens. Trends Microbiol 11: 564–561.
34. Davidson AL, Maloney PC (2007) ABC transporters: how small machines do a big job. Trends Microbiol 15: 448–455.
35. Henderson B, Allan E, Coates ARM (2006) Stress wars: the direct role of host and bacterial molecular chaperones in bacterial infection. Infect Immun 74: 3693–3706.
36. D'Andrea LD, Reagn L (2003) TPR proteins: the versatile helix. Trends Biochem Sci 28: 655–662.
37. Newton HJ, Sansom FM, Bennett-Wood V, Hartland EL (2006) Identification of *Legionella pneumophila*-specific genes by genomic subtractive hybridization with Legionella micdadei and identification of lpnE, a gene required for efficient host cell entry. Infect Immun 74: 1683–1691.
38. Dreher-Lesnick SM, Ceraul SM, Rahman MS, Azad AF (2008) Genome-wide screen for temperature-regulated genes of the obligate intracellular bacterium, *Rickettsia typhi*. BMC Microbiol 8: 61.
39. Lepidi H, Fournier PE, Raoult D (2000) Quantitative analysis of valvular lesions during *Bartonella* endocarditis. Am J Clin Pathol 114: 880–889.
40. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice 200. Nucleic Acids Res 22: 4673–4680.
41. Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17: 1244–1245.
42. Charbonnier Y, Gettler B, Francois P, Bento M, Renzoni A, et al. (2005) A generic approach for the design of whole-genome oligoarrays, validated for genomotyping, deletion mapping and gene expression analysis on *Staphylococcus aureus*. BMC Genomics 6: 95.
43. Churchill GA (2004) Using ANOVA to analyze microarray data. Biotechniques 37: 173–175.
44. Scherl A, Francois P, Charbonnier Y, Deshusses JM, Koessler T, et al. (2006) Exploring Glycopeptide Resistance in *Staphylococcus aureus*: A combined proteomics and transcriptomics approach for the identification of resistance related markers. BMC Genomics 7: 296.
45. Scherl A, Francois P, Bento M, Deshusses JM, Charbonnier Y, et al. (2005) Correlation of proteomic and transcriptomic profiles of *Staphylococcus aureus* during the post-exponential phase of growth. J Microbiol Methods 60: 247–257.
46. Teysseire N, Boudier JA, Raoult D (1995) *Rickettsia conorii* entry into Vero cells. Infect Immun 63: 366–374.

BMC Genomics

BioMed Central

Research article

Open Access

A global view of *Staphylococcus aureus* whole genome expression upon internalization in human epithelial cells

Christian Garzoni[1], Patrice Francois*[1], Antoine Huyghe[1], Sabine Couzinet[1], Caroline Tapparel[1], Yvan Charbonnier[1], Adriana Renzoni[1], Sacha Lucchini[2], Daniel P Lew[1], Pierre Vaudaux[1], William L Kelley[1] and Jacques Schrenzel[1]

Address: [1]Service of Infectious Diseases, University Hospital of Geneva, Department of Internal Medicine, 24 rue Micheli-du-Crest, CH-1211 Geneva 14, Switzerland and [2]Molecular Microbiology Group, Institute of Food Research, Norwich Research Park, Colney, Norwich, NR4 7UA, UK

Email: Christian Garzoni - christian.garzoni@hcuge.ch; Patrice Francois* - patrice.francois@genomic.ch; Antoine Huyghe - antoine.huyghe@genomic.ch; Sabine Couzinet - scouzinet@hotmail.com; Caroline Tapparel - caroline.tapparel@hcuge.ch; Yvan Charbonnier - yvan.charbonnier@genomic.ch; Adriana Renzoni - Adriana.renzoni@hcuge.ch; Sacha Lucchini - sacha.lucchini@bbsrc.ac.uk; Daniel P Lew - Daniel.lew@hcuge.ch; Pierre Vaudaux - Pierre.vaudaux@hcuge.ch; William L Kelley - William.kelley@hcuge.ch; Jacques Schrenzel - Jacques.schrenzel@genomic.ch

* Corresponding author

Published: 14 June 2007

BMC Genomics 2007, **8**:171 doi:10.1186/1471-2164-8-171

Received: 25 September 2006
Accepted: 14 June 2007

This article is available from: http://www.biomedcentral.com/1471-2164/8/171

Abstract

Background: *Staphylococcus aureus*, a leading cause of chronic or acute infections, is traditionally considered an extracellular pathogen despite repeated reports of *S. aureus* internalization by a variety of non-myeloid cells *in vitro*. This property potentially contributes to bacterial persistence, protection from antibiotics and evasion of immune defenses. Mechanisms contributing to internalization have been partly elucidated, but bacterial processes triggered intracellularly are largely unknown.

Results: We have developed an *in vitro* model using human lung epithelial cells that shows intracellular bacterial persistence for up to 2 weeks. Using an original approach we successfully collected and amplified low amounts of bacterial RNA recovered from infected eukaryotic cells. Transcriptomic analysis using an oligoarray covering the whole *S. aureus* genome was performed at two post-internalization times and compared to gene expression of non-internalized bacteria. No signs of cellular death were observed after prolonged internalization of *Staphylococcusaureus* 6850 in epithelial cells. Following internalization, extensive alterations of bacterial gene expression were observed. Whereas major metabolic pathways including cell division, nutrient transport and regulatory processes were drastically down-regulated, numerous genes involved in iron scavenging and virulence were up-regulated. This initial adaptation was followed by a transcriptional increase in several metabolic functions. However, expression of several toxin genes known to affect host cell integrity appeared strictly limited.

Conclusion: These molecular insights correlated with phenotypic observations and demonstrated that *S. aureus* modulates gene expression at early times post infection to promote survival. *Staphylococcusaureus* appears adapted to intracellular survival in non-phagocytic cells.

Background

Staphylococcus aureus is a versatile pathogenic bacterium capable of rapidly developing or acquiring multiple antibiotic resistances, and is now recognized as a worldwide health problem [1]. *S. aureus* is responsible for a wide spectrum of human and animal diseases, ranging from benign skin infections to severe diseases, such as arthritis, osteomyelitis, endocarditis or fatal sepsis [2]. Acute infections are related to the organisms' capacity to secrete a plethora of exotoxins [3,4] and catabolic enzymes [2,5], as documented previously in different experimental models of acute infections [6-8]. *S. aureus* is also responsible for chronic diseases such as osteomyelitis [9], rhinosinusitis [10], or otitis [11]. These infections are difficult to eradicate and often relapse even after prolonged and adapted antibiotic therapy [12,13], suggesting that *S. aureus* has developed specific strategies for intracellular persistence. In addition, anti-infective agents commonly used for the treatment of *S. aureus* infections could enhance selection of invasive intracellular strains [14].

In contrast to other persistent human pathogens, *S. aureus* is not traditionally considered as an intracellular pathogen [15]. Nevertheless, substantial evidence strongly supports that *S. aureus* can be internalized and survive in a variety of non-professional phagocytic cells *in vitro* [2,16-18] and *in vivo* [19,20]. The endocytic uptake of *S. aureus* by non-myeloid cells involves active cellular processes that depend upon F-actin polymerization and is similar in many respects to that observed in professional phagocytes [17]. Whereas entero-invasive pathogens utilize secretion systems to actively induce their own uptake by the host cell, internalization of *S. aureus* by non-professional phagocytes shows similar efficiency *in vitro* with live or killed bacteria [17,21]. The mechanism relies on an interaction between fibronectin binding protein and host-cell α5β1 integrins [17,22,23]. The role of other bacterial surface proteins like clumping-factor A or host cell Src kinase also appears important in the mediation of *S. aureus* uptake and intracellular persistence [18,24]. After internalization, the behavior of the bacterium varies according to cell-line or bacterial strain. For example, some authors reported active intracellular bacterial replication within vacuoles [25] or rapid bacterial escape from vacuole and induction of cellular apoptosis [26-28], while others described persistence for several days before induction of escape processes [29]. The production of α-toxin appears correlated with the induction of apoptosis [27,30,31]. Regulation of α-toxin expression is complex and involves multiple regulators that include *agr*, *sarA* homologues, or *svrA* [32-35].

Molecular details that govern *S. aureus* extended persistence are largely unknown. Metabolic alterations leading to small colony variant (SCV) microorganisms are one possibility that has been described [36-38]. Such *S. aureus* variants were recently shown to efficiently invade endothelial cells *in vitro* and display a markedly higher content in fibronectin-binding proteins than the parental strain [39]. SCVs display a major alteration in their ability to produce or export exotoxins [36] and reveal extensive changes in their global regulatory network [40]. Overall this persistent behavior, possibly related to alteration of regulatory networks, appears compatible with the property of *S. aureus* to generate relapsing infections even years after a first episode was apparently cured [36,41].

Several studies have examined details of cellular responses after *S. aureus* internalization in either phagocytic or non-phagocytic cells [42,43]. However, little is known about bacterial gene expression upon cellular internalization. Recent efforts in high throughput sequencing have contributed to the elucidation of numerous bacterial genomes. To date, eight fully annotated *S. aureus* genomes are publicly available [44-49] allowing the design of DNA microarrays to probe the bacterial transcriptome [50-54], or to catalogue and type variation among clinical isolates [53,55,56].

In this study, we describe an *in vitro* model where *S. aureus* is able to persist for up to two weeks in the absence of either cellular death or emergence of small colony variants. *S. aureus* gene expression changes between adherent bacteria and those arising 2 h and 6 h after cellular internalization were compared using a custom-designed and validated *S. aureus* oligoarray [53]. Profound shut-down of gene expression was observed shortly after internalization for bacterial metabolic functions and transport. This period was followed by resumption of transcriptional activity for metabolic and energy production functions correlating with moderate bacterial multiplication. These findings suggest that *S. aureus* extensively reprograms its transcriptome to adjust to the intracellular environment. Detailed study of these changes may help understanding the molecular mechanisms governing establishment and maintenance of persistent *S. aureus* infections and potentially helpful for the design of new antibacterial drugs.

Results

Infection model and bacterial RNA isolation

We developed a model system that displayed protective survival of *S. aureus* 6850 within human lung epithelial cells A549. Control experiments were performed to examine infection parameters and determine the saturation of internalization. Figure 1A shows bacterial internalization as a function of the multiplicity of infection (m.o.i). The profile describes a linear dose-response until saturation for m.o.i higher than 286 *S. aureus* per cell. We next examined the intracellular survival of strain 6850 as a function of time using the lysostaphin/gentamicin protection

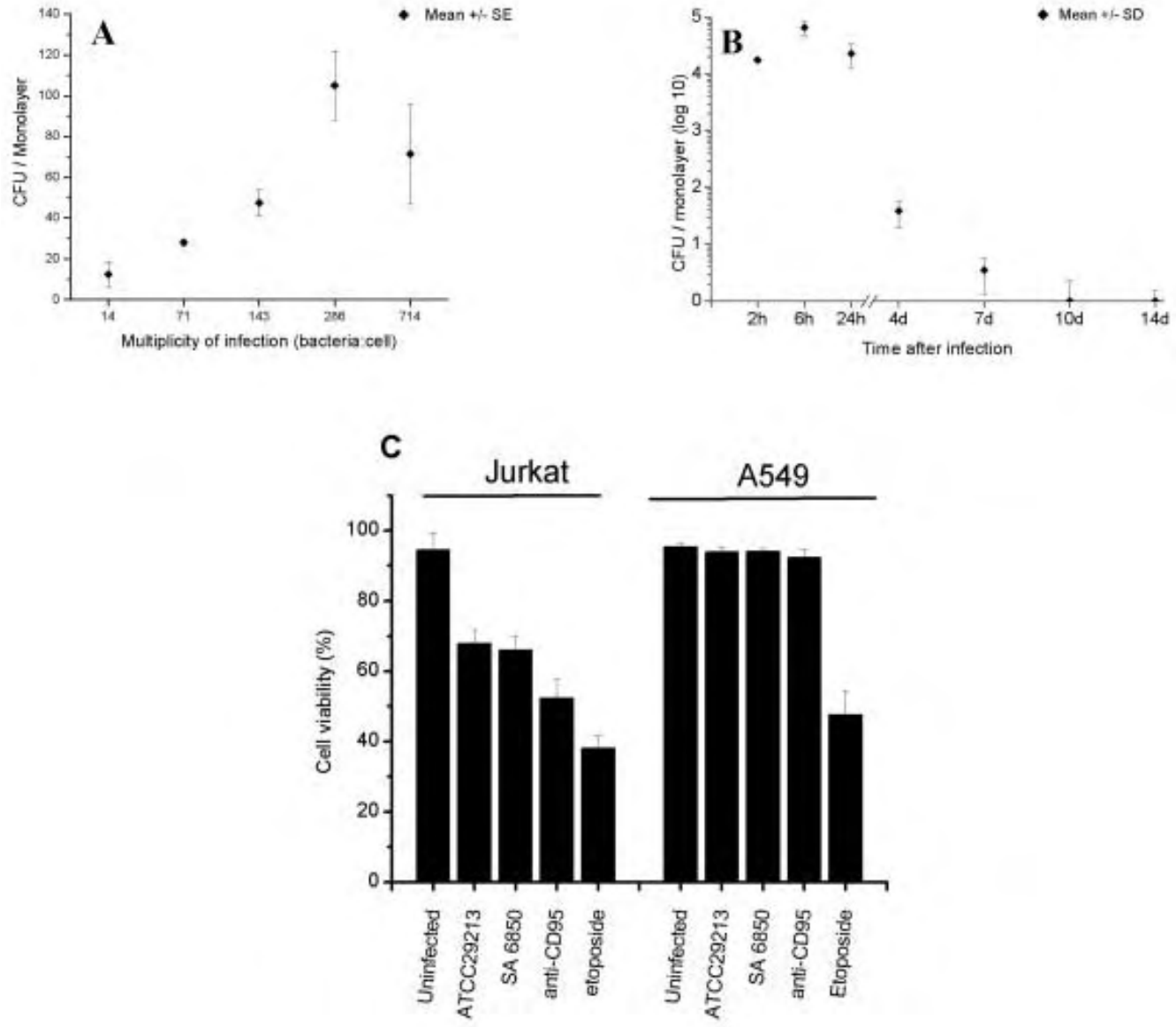


Figure 1

Internalization of *S. aureus* 6850 by A549 cells. **A**. Dose-response of internalization assay performed by adding increasing number of bacteria to constant number of cells. The internalization assay was performed as described in the experimental section. After 2 h, cells were detached with trypsin lyzed with dilute Triton X-100, and then internalized bacteria were assessed by count of colony forming units (cfu). **B**. Intracellular survival assay of *S. aureus* over a two-week interval. Note that at times after 6 hours, viable bacterial counts decreased, though viable bacteria were recovered from monolayers even 2 weeks after internalization. All values are mean ± standard deviation (SD) of at least three independent experiments. **C**. Viability tests were performed after 24 hours of internalization using bacterial m.o.i. of 100:1, or the presence of anti-CD95 (1 μg/ml) or etoposide (100 μM final concentration) as pro-apoptotic controls.

assay, as shown Figure 1B. We observed substantial intracellular survival during the first 24 hour post-internalization. Notably, we observed a 3-fold increase in viable bacteria recovered at 6 hours post internalization compared with 2 hours indicating most probably that bacterial replication, or completion of cell division, had occurred. Viable bacterial counts diminished slightly in the 6–24 h interval. A 3 log decrease in recovered bacteria was observed in the interval from 1–4 days post-infection indicating substantial infection clearance. Nevertheless, viable bacteria could be recovered in this model system up to 2 weeks post infection. Strain 6850 produces strong α-

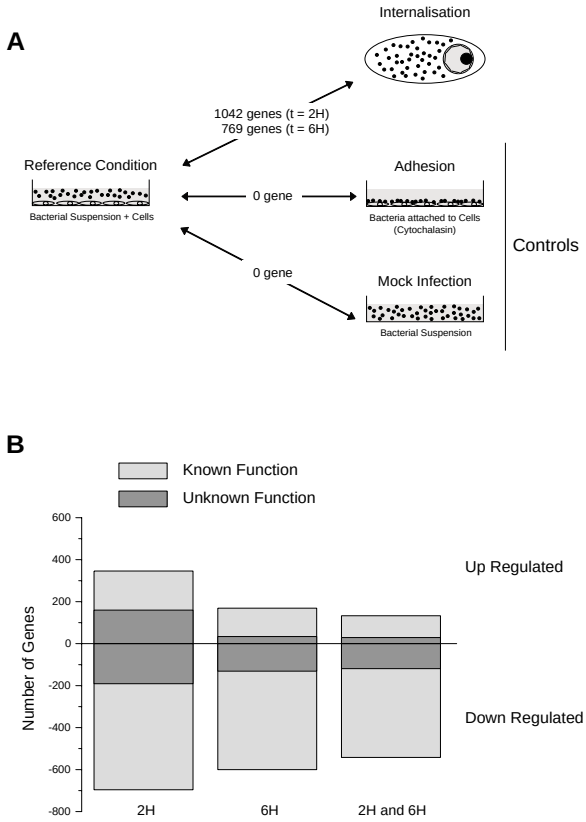


Figure 2

Schematic representation of conditions subjected to microarray analysis and number of differentially expressed genes. **A**. Similar expression profiles were observed for the comparisons between the reference condition and the control conditions consisting in i) mock infection (bacteria in cell culture medium) or ii) adherent bacteria (in the presence of cytochalasin D 1 μg/ml). On the opposite, an important number of genes were found differentially regulated between reference condition and internalized bacteria after 2 or 6 h of intracellular survival. See material and methods section for details. **B**. Numbers of genes found differentially regulated are indicated for each condition. The proportion of non-characterized genes annotated as genes with "unknown function" appears dark grey.

toxin amounts as revealed on blood agar plates as well as in strains recovered from the intracellular survival assay (data not shown). We observed no emergence of SCV at any time point in the assay. We conclude that the model reflects both a window of bacterial survival and replication as well as persistent infection and we focused in early time-points (2, 6 hours) for subsequent detailed study. Microscopic observations with trypan blue revealed no significant cytotoxicity as well as the absence of any cellular morphologic alteration during the first 14 days of co-incubation (Figure 1B). Subsequent viability assays were performed at a sub-saturating m.o.i. of 100:1 using strains 6850 or ATCC29213, previously described as cytotoxic in other cell lines [27,30,31]. Control Jurkat T cells revealed altered cellular viability in the presence of any of the two bacterial strains or by contact with pro-apoptotic compounds (Figure 1C). In contrast, a total absence of cellular toxicity was observed in A549 cells in the presence of either 6850 or ATCC strains. Similarly to previously published observation [57], anti-CD95 failed also to affect A549 viability whereas etoposide showed a significant impact on A549 viability (Figure 1C).

Purification of bacterial RNA from intracellular bacteria proved difficult due to significant excess of eukaryotic material. After protocol optimizations, the extraction and amplification procedures showed appreciable reproducibility (additional file 1) with Pearson correlation value of 0.9–0.93, an important parameter in microarray experiments [58,59]. Nevertheless, bacterial RNA preparations contained detectable contaminating RNA traces from host-cell. Absence of detectable cross-hybridization on the microarray with cellular RNA demonstrates the high specificity of the microarray for bacterial messengers resulting from the design strategy (data not shown).

Analysis of gene expression changes in extracellular and adherent bacteria

We first performed several microarray experiments to evaluate the impact of cell culture medium, presence of cells, presence of cytochalasin D, and bacterial attachment to cells, on the bacterial transcriptome expression. The experimental design and pairwise comparisons performed are depicted in Figure 2A. The following conditions were tested: i) bacteria propagated in invasive medium alone, ii) non-adherent bacteria cultivated in the presence of cells, iii) non-adherent bacteria recovered from infections in the presence of cytochalasin D, and iv) cell-adherent bacteria in the presence of cytochalasin D. No significant changes in gene regulation were recorded from all possible pair-wise comparisons. We conclude from this analysis that exposure of bacteria to our experimental parameters prior to cellular internalization does not result in detectable alteration of bacterial transcriptome. We then compared expression values obtained at time 0 in the culture medium against those obtained after 2 or 6 hours of internalization.

Analysis of intracellular bacterial gene expression changes

Previous studies performed in our laboratory indicated that bacterial uptake by both A549 and MRC5 lung epithelial cells is a rapid process (< 30 minutes) [17]. From this result, as well as those presented in Figure 1B, we reasoned that 2 and 6 hour time points would reveal new information about bacterial transcriptomic program

required for early essential adaptation to the intracellular environment. We therefore evaluated the transcriptome of *S. aureus* after 2 and 6 h post-internalization.

Comparison of gene expression between control (time 0) and 2 h or 6 h post-infection conditions revealed extensive modifications in bacterial gene regulation. The complete raw data set as imported from the array scanner as well as the 2 h, 6 h, and functional group cluster is available online [60].

Statistically significant changes in gene expression were found for 1042 and 766 genes after 2 h or 6 h of infection, respectively (including 338 and 165 hypothetical proteins). After 2 h of infection, 346 genes were up-regulated whereas 696 were down-regulated, while 169 genes were found up-regulated and 600 genes down-regulated 6 h post-infection (Figure 2B). The numbers of genes showing differential expression levels only for one of the two timepoints were 89 and 362 for 2 or 6 h, respectively. Since all tested control conditions did not result in significant changes in gene expression, we conclude that these massive alterations in transcription profiles are triggered upon bacterial internalization.

Clustering and functional group assignment

Differentially expressed genes were clustered by functional group categories according to the COG classification [61] (additional file 2). An additional category consisting in documented virulence factors was added (Table 1 and Figure 3), according to established criteria [62]. Overall, the general trend revealed a substantial down-regulation of genes involved in metabolic functions (additional file 3). For example, among the 55 ORFs encoding for ribosomal proteins, > 40 were found down-regulated at 2 h, or 6 h. This reduction concerned also the totality of the *atp* operon (*atpA* to *atpH*) responsible for ATP synthesis and genes involved in nucleotide and amino acid metabolism. Note however that the majority of these genes were up-regulated between 2 and 6 hours. Similar responses were also observed for numerous genes involved in sugar metabolism and energy production. A more balanced picture was observed among signal transduction and transcription regulators (additional file 3). Changes in gene expression of hypothetical reading frames were observed to be equally distributed between down- and up-regulated genes. Among genes that were up-regulated, we noted a significant number of targets involved in iron metabolism and other inorganic ions (additional file 3) contributing to oxidative stress protection. Among the most significantly up-regulated genes, we observed defense genes, such as *sodA*, *ahp* and also transcription regulators of virulence factors (Table 1), RNAIII, *sarS* (also named *sarH1*) and *svrA* (SA0323 or SAV0334 according to the annotation of *S.aureus* N315 and Mu50, respectively) which are known to modulate the expression of the *agr* locus [35,63]. Another category showing important changes in gene expression was transporters (addi-

Table 1: Genes differentially expressed involved in virulence

	Common	Organism	2 h fold change	6 h Fold change	Protein name
SA2423	*clfB*	N315	0.17	0.19	Clumping factor
SA2291	*fnbA*	N315	3.23		fibronectin binding proteins
SA2290	*fnbB*	N315		3.83	fibronectin binding proteins
MW1940	*hlb*	MW2	3.08		β-hemolysin
SAS065	*hld*	N315	3.05		δ-hemolysin
SA2209	*hlgB*	N315		3.12	γ-hemolysin
SA2461	*icaB*	N315	3.75		Intercellular adhesion proteins
SA1638	*lukE*	N315	3.06		leukotoxin LukE
SA1812	*SA1812*	N315	0.14	0.22	synergohymenotropic toxin precursor *S. intermedius*
SA0521	*sdrE*	N315	0.20	0.29	Ser-Asp rich proteins
SA1817	*sec3*	N315	3.41		enterotoxin type C3
MW0759	*sec4*	MW2	3.60	4.76	enterotoxin C precursor protein
SA1642	*seg*	N315	4.90	3.70	extracellular enterotoxin type G precursor
SA1816	*sel*	N315	3.46	3.09	extracellular enterotoxin L
SA1648	*seo*	N315	3.94	6.33	enterotoxin SeO
SA1009	*set1*	N315	3.73		exotoxin 1
SA0389	*set13*	N315	3.52	3.45	exotoxin 13
SA0390	*set14*	N315	3.77	3.73	exotoxin 14
MW0394	*set26*	MW2		4.95	exotoxin homolog [Genomic island nu Sa alpha2]
SA1011	*set3*	N315	5.00	4.69	exotoxin 3
SA0107	*spa*	N315	0.15	0.10	Protein A
SA0901	*sspA*	N315	0.14	0.11	V8 protease
SA0899	*sspC*	N315	3.22		Staphopain
SA1819	*tst*	N315	3.80	4.10	Toxic shock syndrome toxin-1
SA1645	*yent1*	N315		3.18	enterotoxin Yent1

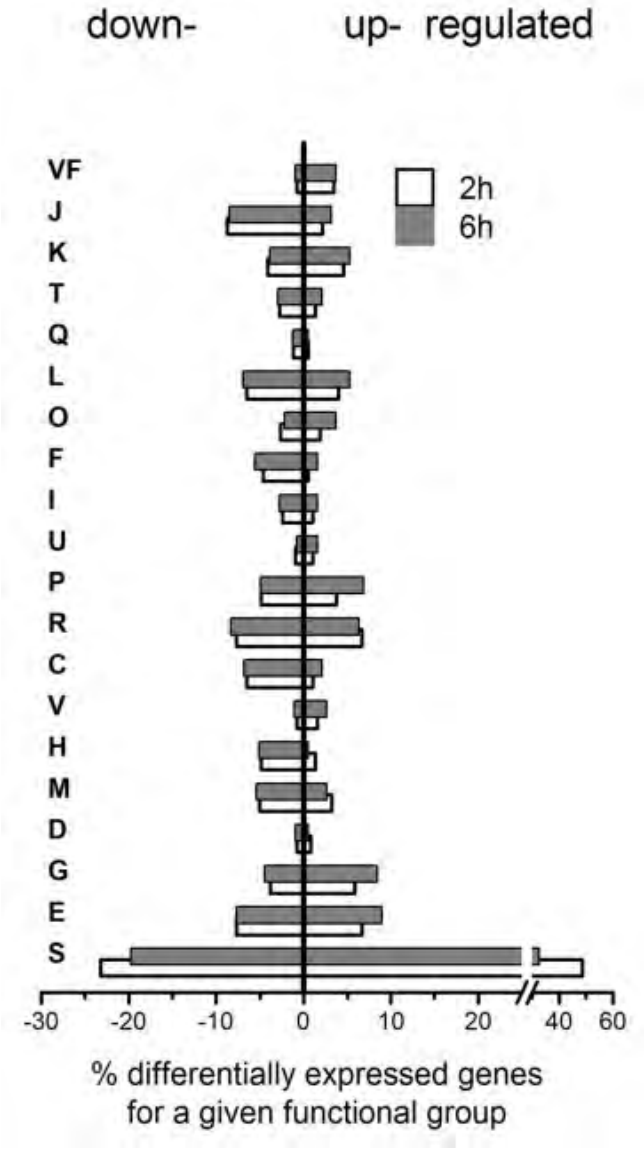


Figure 3
Changes in gene expression 2 h and 6 h after internalization shown by functional categories. Percent changes at 2 h (white) and 6 h (grey) was calculated by dividing the number of significantly changed genes by total genes at the given time point. Genes were assigned functional groups using annotated public database and metabolic pathways databases (COG). Categories are: E: Amino acid transport and metabolism, G: Carbohydrate transport and metabolism, D: Cell cycle control, cell division, chromosome partitioning, M: Cell wall/membrane/envelope biogenesis, H: Coenzyme transport and metabolism, V: Defense mechanisms, C: Energy production and conversion, R: General function prediction only, P: Inorganic ion transport and metabolism, U: Intracellular trafficking, secretion, and vesicular transport I: Lipid transport and metabolism, F: Nucleotide transport and metabolism, O: Posttranslational modification, protein turnover, chaperones, L: Replication, recombination and repair, Q: Secondary metabolites biosynthesis, transport and catabolism, T: Signal transduction mechanisms, K: Transcription, J: Translation, ribosomal structure and biogenesis, and VF for virulence factor.

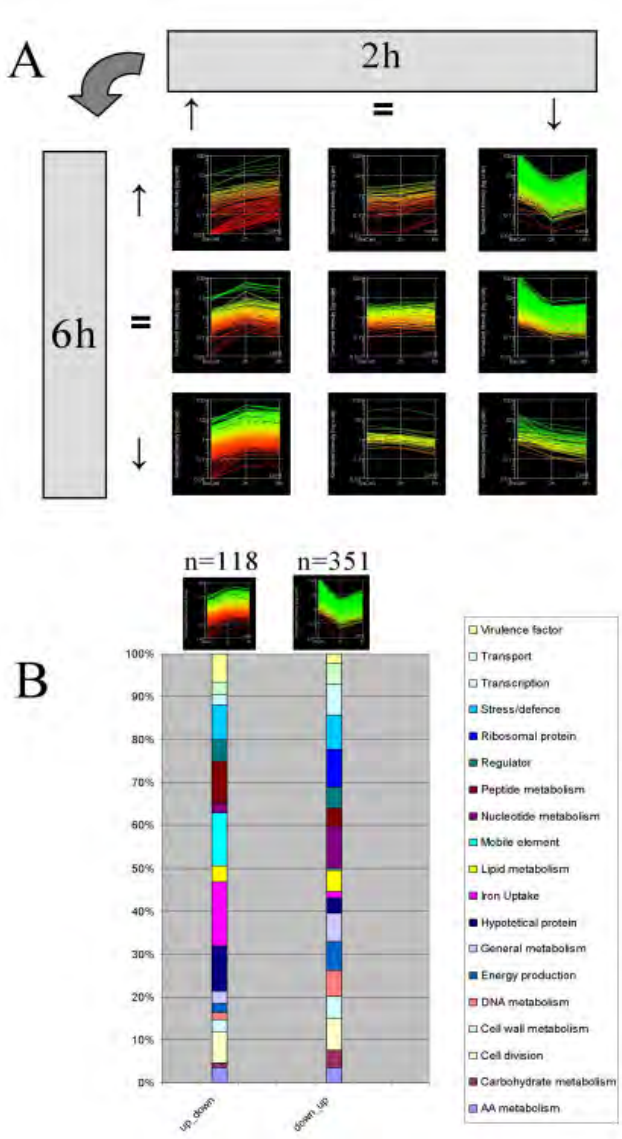


Figure 4
Global pattern of expression. Patterns of gene expression changes are shown for each of nine possible categories. The most abundant categories (B) were analyzed in depth after grouping by gene function. Numbers of genes in these two categories are also shown.

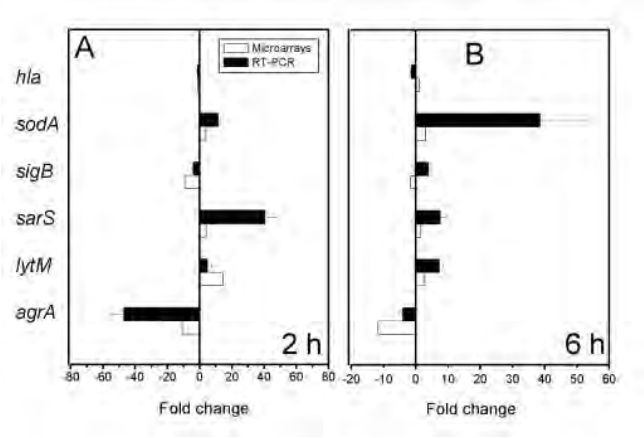


Figure 5
Quantitative PCR validation of microarray data. Dynamic of gene expression changes is shown as fold change by real-time PCR (black bars), or by microarray determination (white bars), at 2 h (A) or 6 h (B) post-infection. Data are mean ± SD of triplicate measurements from 2 independent experiments.

tional file 3). More than 40 genes were found significantly differentially regulated including transporters of iron, sugars, nucleotides and amino-acids. Importantly, expression of toxic virulence factors and α-hemolysin were found either down-regulated, or unchanged, compared to the control conditions (Table 1). Finally, the number of genes involved in biosynthesis or degradation of bacterial envelopes showed contrasted levels of expression after internalization as compared to extracellular control bacteria (see Materials and Methods). Whereas cell-wall metabolism enzymes (the autolysin *lytM*, and several amidases *lytA*, *sgtA and* SA1759) showed drastic up-regulation, transpeptidase, acid techoic synthesis genes, the *mur* operon and several *cap* genes showed massive down-regulation after internalization (additional file 3).

Based on the time-points studied, two major regulation profiles were found composed of genes showing: i) large down-regulation after 2 h, then a slight increase at 6 h, and ii) up-regulation at 2 h then only marginal change at 6 h. These two categories are presented in Figure 4A in the upper-right and lower-left corners, respectively. Grouping by functional category revealed that the first expression profile is mainly composed of mobile elements, transporters, regulators and iron metabolism (Fig. 4B). The other categories were mainly composed of genes involved in nutrient transport, general and nucleotide metabolism, and ribosomal proteins, but also in DNA replication and carbohydrate metabolism (Fig. 4B).

Analysis of gene expression changes in the interval 2 h-6 h post-infection. To examine gene expression changes occurring during its intracellular passage, we extracted a list of genes that were differentially expressed (determined by ANOVA) between times 2 h and 6 h after internalization. A striking up-regulation of gene expression was observed at 6 h post-infection in most functional categories, consistent with bacteria readapting their energy production, intermediary metabolism, protein synthesis, and nucleotide biosynthesis. This finding is consistent with metabolic restart necessary for the observed bacterial doublings in the 2–6 h time interval (Figure 1B).

qPCR validation of array data for selected genes

As an independent measure of differential gene expression, we examined the relative levels of 6 genes selected from different functional categories by quantitative real-time PCR showing up- or down-regulation or unchanged level of expression. As previously established in other studies, the absolute magnitudes of normalized signals reflects the broader dynamic range of the qPCR method compared to microarray measurements [53,64]. The results are shown in Figure 5 and are correlated in terms of up- or down-regulation with microarray determinations. Notably, the data confirms the observed down-regulation of *agr* expression correlated with the absence of *hla* transcription. Congruent changes were observed for *lytM*, *sarS* and *sodA* that showed up-regulation of their expression as previously recorded using microarrays (Figure 5).

Discussion

Recent efforts in the field of high-throughput sequencing contributed to the public release of numerous bacterial genomes of medical importance. This information allowed the development of microarrays able to monitor gene expression changes at the genome scale. We report here the utilization of a microarray covering the whole genome of *Staphylococcus aureus* to evaluate the behavior of the bacterium upon cellular internalization, a phenomenon potentially contributing to bacterial protection against host defenses, and important in the context of chronic infections. Our overall findings are in agreement with previous studies utilizing experimental cellular models of bacterial internalization, using either intracellular or facultative intracellular pathogens. If the ability of *S. aureus* to persist in different *in vitro* models is widely accepted, the capacity of *S. aureus* to escape from host-defenses and persist intracellularly *in vivo* is still debated [15]. In this respect, the call to document "the presence of staphylococci within non-professional phagocytes at sites of infection in clinicopathological tissues" [15] has been recently demonstrated by Clément and colleagues, who identified *S. aureus* residing in human nasal epithelial cells *in vivo*, a probable cause of recurrent rhinosinusitis

infection [19]. This *in vivo* survival capacity was further documented in the same population and shown to constitute a significant risk factor for recurrent episodes of rhinosinusitis refractory to antimicrobial and surgical therapy [65]. Similar behavior was also observed in an experimental model of mastitis [20].

In our model, the presence of cells or the attachment of bacteria to A549 cells (in the presence of cytochalasin) showed limited impact on bacterial transcriptome. This observation is in accordance with previous reports from our group showing that no active bacterial process is require for internalization [17]. With respect to bacterial pathogenesis, this observation suggests that *S. aureus* internalization is probably not rare as it requires only a minimal number of bacterial factors [22]. In contrast to other microorganisms known to invade eukaryotic cells and rapidly divert cellular compounds and energy to actively multiply [66], our study indicates that upon cellular internalization, *S. aureus* transiently shuts down most of its metabolic functions and adopts an apparent "dormant state" without extensively initiating classic starvation or stress process. These profound changes in the bacterial transcriptome affect approximately 40% of all putative ORFs identified in the genomic sequence of the bacterium and appear related to an extensive adaptation to the new environment constituted by the intracellular compartment. The majority of genes involved in energy production (citrate and ATP cycles, pyruvate metabolism, and oxidative phosphorylation) were drastically reduced after 2 h internalization, Consequently, > 75% of *S. aureus* ribosomal proteins showed a significant down-regulation, which constitutes a common trait with streptococcal exposition to endothelial cell [67], another Gram positive pathogen able to invade non-phagocytic cells before inducing apoptosis [68]. For these 2 pathogens, metabolic and cellular division machineries appear strongly down-regulated rapidly after internalization [67]. This situation is markedly different from that observed in invasive intracellular pathogens such as *Salmonella* [69], *Shigella* [66,70], or *Pasteurella* [71], which show up-regulation of most cellular processes upon internalization and are able to multiply actively before disseminating to other cells [69,70]. In our model, a significant increase in colony forming unit counts was observed between 2 and 6 h of intracellular persistence. This cellular multiplication parallels up-regulation of *lytM*, as well as that of numerous amidases involved in cell-wall processing [72]. Concurrently, a slight up-regulation of cell-wall and techoic acid synthesis genes was observed between these time points, suggesting evidence of intracellular replication or completion of cell division, as observed in the context of *S. aureus* phagocytosis by neutrophils [43]. This metabolic evolution was paralleled by a massive up-regulation of numerous transporters. In our study, more than 40 transporters, involved in inorganic ions, nucleotides or peptides, or urea were found significantly up-regulated. Transporters involved in metabolic pathways contribute to the direct rerouting of cellular resources by the internalized bacterium. Illustrative examples are the up-regulation of *uhpT* (observed at 2 and 6 h internalization), allowing glucose-6-phosphate acquisition, as observed during internalization of other bacterial species [66,70] or during *in vitro* [73] or intracellular exposition of *S. aureus* to acidic environments [43]. *S. aureus* genome contains 7 putative iron transporters [74] that are important mediators of bacterial virulence. Iron transporters are typically up-regulated in various internalized invasive pathogens [67,74,75]. In our study, *isdF*, SA0567 and particularly *bitC*, were drastically up-regulated in *S. aureus*, thus providing probably sufficient amounts of iron from the intracellular medium, whereas all genes composing the *fhu* operon, encoding ferric hydroxamate assimilation proteins were down-regulated, in accordance with their function of iron transporters under specific environmental conditions [76].

Intracellular survival can potentially expose the bacterium to cellular defense mechanisms such as oxygen radicals. In our study, the number of up-regulated genes involved in stress protection or bacterial defense revealed surprisingly low. Except for *sodA* and *cysM*, whose products protect bacterial cell against oxidative stress [77], a majority of genes involved in oxidative stress response appeared down-regulated. This finding is markedly different than the behavior of *Shigella* [66], but consistent with previous reports about *S. pneumoniae* [67]. Note however that experiments performed in A549 cells failed to detect production of reactive oxygen species upon bacterial uptake (not shown). Internalized *S. aureus* appeared to solicit protective functions but in a different context than that requiring preservation of the cellular machinery integrity as described by Voyish in the context of resistance to neutrophil phagocytosis [43]. In non-phagocytic cells, the expression of capsule operon, Clp proteases family and chaperones, as well as their known regulators *ctsR* and *sigB* [78], potentially conferring survival against toxic cellular metabolites was drastically down-regulated.

The expression of numerous bacterial virulence factors is modulated through complex networks of regulatory molecules including two-component systems and global regulators such as *agr* or *sar*. Multiple regulatory pathways contributed not only to the rapid adaptation of the bacteria to environmental changes but also to the regulation of toxins or cell-surface components expression during bacterial growth-phases [79,80]. In our system, upon contact with cell monolayers, strain 6850 was in early exponential phase of growth where *agr* activity is limited. Following internalization and metabolic restart, general metabolic functions appeared up-regulated in the absence of *agr*

induction. In addition, internalized bacteria showed an increase in expression of *sarS* and *svrA* simultaneously to a marked down-regulation of *saeRS* genes, which are important compounds known to modulate *agr* expression [35,63]. In parallel to a moderately increased expression of RNAIII, a compound known to induce *agr* expression *in vitro*, this regulation pattern results in the reduced expression of *hla* that potentially mediates apoptosis induction in some systems [25,27,30]. This uncoupling between *agr* and RNAIII expression has been previously observed *in vivo* in an experimental model of endocarditis [81]. The restricted level of *hla* expression suggests a different regulation behavior under these experimental conditions or the action of another unknown regulator. The latter hypothesis is further supported by the results of Goerke [32], who showed that mutation in *agr* and *sar* had no consequence for *hla* expression *in vivo*. This unknown target is probably not *sigB*, a regulator found drastically down-regulated upon internalization as well as the totality of related *rsb* partners in our model. *SigB* has been recently suggested as a potential partner involved in the persistence of SCVs recovered from cystic fibrosis [82]. In our study, we failed to detect such SCVs as *S. aureus* recovered after 2 weeks of internalization yielded significant hemolysis when plated on blood agar. However, this observation does not rule out the emergence of SCVs in our model as this phenotype appears able to revert spontaneously [83]. These observations are directly related to the absence of cytotoxicity on A549 cells even after two weeks of infection, an observation common between our study and previous work using SCVs [84].

In contrast to the behavior of internalized SCVs, Bantel and colleagues reported 6850-evoked induction of apoptosis using Jurkat T cells [30], which is consistent with our control assays confirming that strain 6850 was able of provoking cytotoxic effects. In addition, Qazi and colleagues reported the induction of *agr* and bacterial escape from endosomal vesicles rapidly after internalization using *S. aureus* RN6390 and MAC-T [85]. On the contrary, Kahl reported that i) rifampicin-treated bacteria failed to induce apoptosis and that ii) Cowan I -a strain that failed to multiply in epithelial cells- does not show appreciable cytotoxic effects [25]. These observations are important concerning *S. aureus* pathogenicity and suggest that certain combinations of host cells and bacteria can result in different biological outcomes. Altogether, results obtained with different models strongly suggest that the internalization of *S. aureus* is a key factor in chronic infections, as intracellular survival occurs in cells showing a long lifespan such as fibroblasts or epithelial cells. *S. aureus* intracellular survival appears related to its capability to adopt a discrete behavior instead of actively duplicating. *S. aureus* then benefits from natural or programmed cell death to re-emerge and trigger another episode of infection, leading to chronicity.

Conclusion

Internalization of *S. aureus* in non-phagocytic cells explaining its resistance properties against cellular or humoral host defenses is an attractive hypothesis relying on its particular ability to generate chronic diseases. In our model of cellular persistence, *S. aureus* strain 6850 is able to reside intracellularly for prolonged period of time. Microarray experiments provide a global picture of genes possibly involved in *S. aureus* pathogenesis, and in particular, the capacity of the bacterium to diminish most of its metabolic functions, despite soliciting numerous transporters allowing maintenance of vital functions but limiting its multiplication. Limited bacterial growth and expression of regulators of virulence factors leads to a finely controlled expression of cell-active toxins insuring intracellular bacterial survival. Global bacterial responses upon internalization appears complex but in accordance with the capacity of the bacterium to reside intracellularly. Finally, the study has uncovered dozens of potential targets for anti-infective therapeutics.

Methods

Bacterial strains and cell culture

A549 cells (ATCC CCL-185, human lung adenocarcinoma) were maintained under 5% CO_2 in minimal essential medium (MEM, Gibco) supplemented with 10% heat-inactivated fetal calf serum (Gibco) and 100 U penicillin/ml. Two days prior to infection, 6×10^6 cells were seeded in 10 cm wells yielding weakly confluent monolayers (2×10^7 cells/well) at the time of infection. *S. aureus* 6850 is derived from multifocal osteomyelitis, is fully pigmented, and maintains a stable virulence phenotype [86]. Overnight cultures in Mueller-Hinton broth (MHB, BD-Difco) were diluted with fresh media and grown to early log phase for 3 hours. Cells were washed twice in PBS, filtered (5 μm, Sartorius) to reduce aggregation and re-suspended in MEM at 1×10^9 cells/ml. Jurkat T cells (kind gift of D. Trono, Univ. of Geneva) were cultivated in RPMI (Gibco) supplemented with 10% heat-inactivated fetal calf serum (Gibco) and 100 U penicillin/ml.

Cell infection and control conditions

Three hours before infection, A549 cells were washed once with MEM, then placed in 5.5 ml invasion medium (IM) consisting of MEM enriched with 10% human albumin (ZLB Biopharma). Infections were initiated with a multiplicity of infection (m.o.i. 100:1 bacteria/cells) using bacteria in early logarithmic growth phase (OD_{600} = 0.2). Cultures were incubated 30 min at 37°C in 5% CO_2 with 2×10^9 washed bacteria to allow adhesion and internalization. Non-adherent bacteria were collected (control bacteria) and fixed with ice cold acetone/EtOH (1:1).

Infected monolayer cells were washed twice with MEM to remove residual non-adherent bacteria then incubated 20 min at 37°C in MEM supplemented with 4 μg/ml lysostaphin. Cells were then washed twice with MEM and placed in 20 ml MEM containing 100 μg/ml gentamicin (defines time = 0). At 2 and 6 hours post-infection, cells were washed twice with PBS, fixed in acetone/EtOH (1:1) and processed for RNA extraction.

Cell viability test

Cells (1×10^6) were seeded in triplicate into culture plates 24 h prior to treatment. Bacterial infections were performed as described for the standard lysostaphin/gentamycin protection assay. Control experiments using anti-CD95 ligation (1 μg/ml) or etoposide (100 μM final concentration) were included for cytotoxicity measurements with trypan blue. At 24 hours post-internalization, cells were washed and the viability of monolayers determined using Trypan blue dye exclusion. Viability was evaluated and expressed as a percentage of the control condition plate. In some cases, we examined combined culture supernatants and monolayers to detect the presence of non-viable detached cells. No difference was noted from viability measurements of monolayers alone.

Mock infection and cytochalasin D experiments

Fixation and RNA extractions were identical for every condition as described above. Cells were pretreated 30 minutes prior infection with 1 μg/ml cytochalasin D (Sigma) and also 60 minutes post-infection. Culture supernatants were collected and non-adherent bacterial cells were harvested, washed twice with PBS and fixed. Cell surface-attached bacteria were detached from A549 monolayer by rapid trypsin incubation, harvested, washed, and fixed. For mock infection, bacterial suspensions were prepared in IM in the absence of A549 cells.

RNA extraction and purification

For all experiments, material (host cells or free bacteria) were lysed in the presence of ice cold acetone:ethanol (1:1) for 4 minutes. Cell monolayers were maintained on ice, then scraped and centrifuged briefly at 4°C. Cellular debris and bacterial pellets were suspended in ice cold RLT buffer (Qiagen) and centrifuged, then washed four times

Table 2: List and characteristics of oligonucleotides used in the qPCR control experiments

Primer or Probe name	Sequence (5'→3')	Length	[A]5'Dye	ORF number in N315
HU				**SA1305**
F_*HU*_34	CAAAGAGAAAACATGGTTACCATTATTAA	29		
R_*HU*_135	CTCAAGCACCTCATAAGGATTATCAG	26		
P_*HU*_83[A]	AAAAGCCTATGGAAATTGCCCTCGCA	26	FAM	
agrA				**SA1844**
F_*agrA*_34	CAAAGAGAAAACATGGTTACCATTATTAA	29		
R_*agrA*_135	CTCAAGCACCTCATAAGGATTATCAG	26		
P_*agrA*_83[A]	AAAAGCCTATGGAAATTGCCCTCGCA	26	FAM	
sarS				**SA0108**
F_*sarS_553*	CACCATAAATACCCTCAAACTGTTAGAG	28		
R_*sarS_638*	TCATCTTCAGTTGAGCGTTCTTTT	24		
P_*sarS_596*[B]	AAAAGCAAGGCTATCTAA	18	FAM	
lytM				**SA0265**
F_*lytM_32*	CGATGGGCTTCGCTACATTT	20		
R_*lytM_105*	ATGTGCTTGTTGGGTGTTTGTC	22		
P_*lytM_56*[A]	TGGCGCATCAAGCAGATGCAGC	22	FAM	
sodA				**SA1382**
F_*sodA_168*	TGTTGCTAATTTAGACAGTGTACCAGCTA	29		
R_*sodA_275*	TCCCAGAATAATGAATGGTTTAAATG	26		
P_*sodA_198*[A]	CATCCAAACTGCTGTACGTAATAATGGCGG	30	FAM	
sirA				**SA0111**
F_*sirA_151*	GGGAAACCAAAGCGTGTTGT	20		
R_*sirA_246*	TGTCCATGATTCTACAGCACCTACA	25		
P_*sirA_198*[A]	ATATCAAGGTGCCACTGACGTCGCTGT	27	FAM	

[A] Taqman probes with a 3'-TAMRA quencher (Eurogentec, Seraing, Belgium).
[B] Minor groove binder probe with non-fluorescent quencher bound to the 3'end (Applied Biosystems). All experiments used concentrations of 200 nM and 100 nM for primers and probes, respectively. Primers and probe for *hla* quantification have been published previously [87,88].

in ice cold TE (10 mM Tris/HCl, 1 mM EDTA pH 8.0). Fixed and washed bacteria were lysed in TE containing 200 μg/ml lysostaphin for 15 minutes at 37°C as previously described [87,88]. RNA was purified using RNeasy columns (Qiagen) following the manufacturer's instructions. For each infection experiment, RNA from 6 wells was pooled. Purified total RNA extracts were treated with MicrobENRICH (Ambion) following the manufacturer's instructions. RNA quality and yields were monitored before and after enrichment using a 2100 BioAnalyzer (Agilent) on PicoLab chip kit (Agilent).

cDNA synthesis, amplification, and fluorescent labeling

Total RNA (300 ng per condition) was converted into cDNA by Superscript II (Invitrogen) and amplified following a previously described method [89] yielding limited bias and robust amplification [90]. Briefly, 20 ng of ds cDNA was amplified using the Genomiphi kit (Amersham Biosciences) according to the manufacturer's recommendations, then labeled with Cy3-dCTP (Perkin Elmer) using BioPrime DNA Labeling System (Invitrogen). Cy5-dCTP (Perkin Elmer) labeled genomic DNA (0.5 mg) pooled from *S. aureus* strains N315, Mu50, COL, and MW2 was used as reference for the normalization process, according to previous work [53,91].

S. aureus *microarray and hybridization conditions*

The *S. aureus* 60-mer oligonucleotide microarray contains 8455 oligonucleotides probes covering > 99% of genes of 4 sequenced *S. aureus* strains: N315, Mu50, MW2 and COL [53]. Labeled cDNA (signal channel, Cy3) and genomic DNA (control channel, Cy5) were mixed in 250 μL Agilent hybridization buffer and hybridized at 60°C for 17 hours. Slides were washed, dried under nitrogen, and scanned at 100% photomultiplier tube power for both wavelengths (Agilent). Microarrays for all experimental time-points were performed in triplicate.

Data analysis

Local background subtracted signals were calculated from three independent experiments using Feature Extraction software (Agilent version A6.1.1). To ensure spot quality, features and their respective background which were not uniform in pixel fluorescence intensity distribution in both channels were flagged (non-uniformity outlier flagging algorithm). Spots showing a genomic DNA reference signal not significantly different from corresponding background were also flagged (two-tailed Student's *t*-test feature versus background with $p < 0.05$). Raw data were imported and analyzed using GeneSpring 7.2 (Agilent). Data were normalized using both per spot (signal channel divided by the corresponding control channel and generation of $\log_{10}$ ratio) and per chip (to the 50^{th} percentile). Flagged spots of low quality were excluded from further analysis. Replicates from the same test condition were grouped and the mean calculated. Significance of normalized data was determined using Welch's approximate *t*-test adjusting individual *P* values with the Benjamini-Hochberg correction for false discovery rate multiple test. Genes showing significant change in expression between two conditions and fold change ≥ 3 were defined as differentially expressed. Fold changes for the complete dataset is available online as additional file 2.

The analysis of differential gene expression patterns in nine possible combinations when comparing 0, 2, and 6 h was performed using the Welch conditional parametric test and false discovery rate correction (variances assumed unequal, $p < 0.05$). Genes showing fold changes > 1.5 are shown.

Quantitative PCR

To confirm microarray results, the expression levels of *hla*, *sigB*, *agrA*, *sodA*, *sarS and lytM* genes were determined by real-time PCR analysis. Genomiphi-amplified material (1 ng) from independent experiments was assembled with specific primers (Invitrogen) and fluorescent probes (Table 2) designed using the PrimerExpress software (Applied Biosystems) in specific enzymatic kit mixtures (Eurogentec). PCR reactions were performed in a final volume of 20 μl in a SDS7700 (Applied Biosystems). Relative expression levels were determined by comparing cycle threshold (Ct) of each gene to the Ct value of the *hu* gene for the same cDNA preparation [92].

Authors' contributions

CG contributed to the design of the study, performed the totality of microarray experiments including RNA extraction, amplification and data analysis, PF and JS conceived the study, contributed to experimental design and analysis and coordinated final writing, CT contributed to the real-time PCR validation experiments, AH and YC contributed to the design of the microarray and developed the required bioinformatics, SL was involved in microarray interpretation, SC was involved in the design and interpretation of results obtained in the infection model, AR was involved in microarray interpretation and contributed to MS elaboration, DL and PV contributed to manuscript elaboration, WK was involved in the design, experiments and analysis related to the infection model including apoptosis experiments and in the elaboration of the manuscript.

Additional material

Additional file 1

Reproducibility of RNA amplification. *cDNA was generated in 2 separate reactions from the same RNA pool, amplified independently and hybridized in 2 different arrays. Labeled cDNA from the starting material was used as control. Dye normalized intensities of amplified material are plotted in both axis.*

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-8-171-S1.tiff]

Additional file 2

Fold changes at 2 h and 6 h of all genes compared to control bacteria. *Values represent mean of fold change from 3 or 4 independent biological replicates. * Some genes are classified in two different COG categories and appear on separate lines in the table.*

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-8-171-S2.pdf]

Additional file 3

Genes differentially expressed grouped by function. *Values represent mean of fold change from 3 or 4 independent biological replicates. * Some genes are classified in two different COG categories and appear on separate lines in the table.*

Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-8-171-S3.pdf]

Acknowledgements

This work was supported by grants from the Swiss National Science Foundation PP00B-103002/1 (JS), 3200B0-108401 (DL), 3100A0-100425 (WK) and 3100A0-116075 (PF). We are grateful to Christine Barras, Manuela Bento and Elzbieta Huggler for excellent technical assistance. CG performed the study as part of the post-graduate course in experimental medicine supported by the University of Zurich.

References

1. Fey PD, Said-Salim B, Rupp ME, Hinrichs SH, Boxrud DJ, Davis CC, Kreiswirth BN, Schlievert PM: **Comparative molecular analysis of community- or hospital-acquired methicillin-resistant Staphylococcus aureus.** *Antimicrob Agents Chemother* 2003, **47:**196-203.
2. Lowy FD: **Staphylococcus aureus infections.** *N Engl J Med* 1998, **339:**520-532.
3. Novick RP: **Autoinduction and signal transduction in the regulation of staphylococcal virulence.** *Mol Microbiol* 2003, **48:**1429-1449.
4. Archer GL: **Staphylococcus aureus: a well-armed pathogen.** *Clin Infect Dis* 1998, **26:**1179-1181.
5. Vojtov N, Ross HF, Novick RP: **Global repression of exotoxin synthesis by staphylococcal superantigens.** *Proc Natl Acad Sci U S A* 2002, **99:**10102-10107.
6. De Boer ML, Kum WW, Pang LT, Chow AW: **Co-production of staphylococcal enterotoxin A with toxic shock syndrome toxin-1 (TSST-1) enhances TSST-1 mediated mortality in a D-galactosamine sensitized mouse model of lethal shock.** *Microb Pathog* 1999, **27:**61-70.
7. De Boer ML, Kum WW, Chow AW: **Staphylococcus aureus isogenic mutant, deficient in toxic shock syndrome toxin-1 but not staphylococcal enterotoxin A production, exhibits attenuated virulence in a tampon-associated vaginal infection model of toxic shock syndrome.** *Can J Microbiol* 1999, **45:**250-256.
8. Kernodle DS, Voladri RK, Menzies BE, Hager CC, Edwards KM: **Expression of an antisense hla fragment in Staphylococcus aureus reduces alpha-toxin production in vitro and attenuates lethal activity in a murine model.** *Infect Immun* 1997, **65:**179-184.
9. Lew DP, Waldvogel FA: **Osteomyelitis.** *Lancet* 2004, **364:**369-379.
10. Gittelman PD, Jacobs JB, Lebowitz AS, Tierno PM Jr.: **Staphylococcus aureus nasal carriage in patients with rhinosinusitis.** *Laryngoscope* 1991, **101:**733-737.
11. Brook I, Finegold SM: **Bacteriology of chronic otitis media.** *JAMA* 1979, **241:**487-488.
12. Kauffman CA, Terpenning MS, He X, Zarins LT, Ramsey MA, Jorgensen KA, Sottile WS, Bradley SF: **Attempts to eradicate methicillin-resistant Staphylococcus aureus from a long-term-care facility with the use of mupirocin ointment.** *Am J Med* 1993, **94:**371-378.
13. Powers KA, Terpenning MS, Voice RA, Kauffman CA: **Prosthetic joint infections in the elderly.** *Am J Med* 1990, **88:**9N-13N.
14. Krut O, Sommer H, Kronke M: **Antibiotic-induced persistence of cytotoxic Staphylococcus aureus in non-phagocytic cells.** *J Antimicrob Chemother* 2004, **53:**167-173.
15. Lowy FD: **Is Staphylococcus aureus an intracellular pathogen?** *Trends Microbiol* 2000, **8:**341-343.
16. Alexander EH, Hudson MC: **Factors influencing the internalization of Staphylococcus aureus and impacts on the course of infections in humans.** *Appl Microbiol Biotechnol* 2001, **56:**361-366.
17. Sinha B, Francois PP, Nusse O, Foti M, Hartford OM, Vaudaux P, Foster TJ, Lew DP, Herrmann M, Krause KH: **Fibronectin-binding protein acts as Staphylococcus aureus invasin via fibronectin bridging to integrin a5b1.** *Cell Microbiol* 1999, **1:**101-117.
18. Que YA, Haefliger JA, Piroth L, Francois P, Widmer E, Entenza JM, Sinha B, Herrmann M, Francioli P, Vaudaux P, Moreillon P: **Fibrinogen and fibronectin binding cooperate for valve infection and invasion in Staphylococcus aureus experimental endocarditis.** *J Exp Med* 2005, **201:**1627-1635.
19. Clément S, Vaudaux P, Francois P, Schrenzel J, Huggler E, Kampf S, Chaponnier C, Lew D, Lacroix JS: **Evidence of an Intracellular Reservoir in the Nasal Mucosa of Patients with Recurrent Staphylococcus aureus Rhinosinusitis.** *J Infect Dis* 2005, **192:**1023-1028.
20. Brouillette E, Grondin G, Lefebvre C, Talbot BG, Malouin F: **Mouse mastitis model of infection for antimicrobial compound efficacy studies against intracellular and extracellular forms of Staphylococcus aureus.** *Vet Microbiol* 2004, **101:**253-262.
21. Cossart P: **Bacterial invasion: a new strategy to dominate cytoskeleton plasticity.** *Dev Cell* 2004, **6:**314-315.
22. Sinha B, Francois P, Que YA, Hussain M, Heilmann C, Moreillon P, Lew D, Krause KH, Peters G, Herrmann M: **Heterologously expressed Staphylococcus aureus fibronectin-binding proteins are sufficient for invasion of host cells.** *Infect Immun* 2000, **68:**6871-6878.
23. Dziewanowska K, Carson AR, Patti JM, Deobald CF, Bayles KW, Bohach GA: **Staphylococcal fibronectin binding protein interacts with heat shock protein 60 and integrins: role in internalization by epithelial cells.** *Infect Immun* 2000, **68:**6321-6328.
24. Fowler T, Johansson S, Wary KK, Hook M: **Src kinase has a central role in in vitro cellular internalization of Staphylococcus aureus.** *Cell Microbiol* 2003, **5:**417-426.
25. Kahl BC, Goulian M, Van Wamel W, Herrmann M, Simon SM, Kaplan G, Peters G, Cheung AL: **Staphylococcus aureus RN6390 replicates and induces apoptosis in a pulmonary epithelial cell line.** *Infect Immun* 2000, **68:**5385-5392.
26. Bayles KW, Wesson CA, Liou LE, Fox LK, Bohach GA, Trumble WR: **Intracellular Staphylococcus aureus escapes the endosome and induces apoptosis in epithelial cells.** *Infect Immun* 1998, **66:**336-342.
27. Haslinger-Loffler B, Kahl BC, Grundmeier M, Strangfeld K, Wagner B, Fischer U, Cheung AL, Peters G, Schulze-Osthoff K, Sinha B: **Multiple virulence factors are required for Staphylococcus aureus-induced apoptosis in endothelial cells.** *Cell Microbiol* 2005, **7:**1087-1097.
28. Wesson CA, Liou LE, Todd KM, Bohach GA, Trumble WR, Bayles KW: **Staphylococcus aureusagr and sar global regulators influence internalization and induction of apoptosis.** *Infect Immun* 1998, **66:**5238-5243.
29. Menzies BE, Kourteva I: **Internalization of Staphylococcus aureus by endothelial cells induces apoptosis.** *Infect Immun* 1998, **66:**5994-5998.
30. Bantel H, Sinha B, Domschke W, Peters G, Schulze-Osthoff K, Janicke RU: **a-Toxin is a mediator of Staphylococcus aureus-induced cell death and activates caspases via the intrinsic death pathway independently of death receptor signaling.** *J Cell Biol* 2001, **155:**637-648.
31. Mempel M, Schnopp C, Hojka M, Fesq H, Weidinger S, Schaller M, Korting HC, Ring J, Abeck D: **Invasion of human keratinocytes by**

Staphylococcus aureus and intracellular bacterial persistence represent haemolysin-independent virulence mechanisms that are followed by features of necrotic and apoptotic keratinocyte cell death. *Br J Dermatol* 2002, **146:**943-951.
32. Goerke C, Fluckiger U, Steinhuber A, Zimmerli W, Wolz C: **Impact of the regulatory loci agr, sarA and sae of Staphylococcus aureus on the induction of alpha-toxin during device-related infection resolved by direct quantitative transcript analysis.** *Mol Microbiol* 2001, **40:**1439-1447.
33. Cheung AL, Ying P: **Regulation of alpha- and beta-hemolysins by the sar locus of Staphylococcus aureus.** *J Bacteriol* 1994, **176:**580-585.
34. Fournier B, Klier A, Rapoport G: **The two-component system ArlS-ArlR is a regulator of virulence gene expression in Staphylococcus aureus.** *Mol Microbiol* 2001, **41:**247-261.
35. Garvis S, Mei JM, Ruiz-Albert J, Holden DW: **Staphylococcus aureus svrA: a gene required for virulence and expression of the agr locus.** *Microbiology* 2002, **148:**3235-3243.
36. Proctor RA, van Langevelde P, Kristjansson M, Maslow JN, Arbeit RD: **Persistent and relapsing infections associated with small-colony variants of Staphylococcus aureus.** *Clin Infect Dis* 1995, **20:**95-102.
37. von Eiff C, Bettin D, Proctor RA, Rolauffs B, Lindner N, Winkelmann W, Peters G: **Recovery of small colony variants of Staphylococcus aureus following gentamicin bead placement for osteomyelitis.** *Clin Infect Dis* 1997, **25:**1250-1251.
38. Proctor RA, Kahl B, von Eiff C, Vaudaux PE, Lew DP, Peters G: **Staphylococcal small colony variants have novel mechanisms for antibiotic resistance.** *Clin Infect Dis* 1998, **27 Suppl 1:**S68-S74.
39. Vaudaux P, Francois P, Bisognano C, Kelley WL, Lew DP, Schrenzel J, Proctor RA, McNamara PJ, Peters G, von Eiff C: **Increased expression of clumping factor and fibronectin-binding proteins by hemB mutants of Staphylococcus aureus expressing small colony variant phenotypes.** *Infect Immun* 2002, **70:**5428-5437.
40. Kahl BC, Belling G, Becker P, Chatterjee I, Wardecki K, Hilgert K, Cheung AL, Peters G, Herrmann M: **Thymidine-dependent Staphylococcus aureus small-colony variants are associated with extensive alterations in regulator and virulence gene expression profiles.** *Infect Immun* 2005, **73:**4119-4126.
41. Donati L, Quadri P, Reiner M: **Reactivation of osteomyelitis caused by Staphylococcus aureus after 50 years.** *J Am Geriatr Soc* 1999, **47:**1035-1037.
42. Matussek A, Strindhall J, Stark L, Rohde M, Geffers R, Buer J, Kihlstrom E, Lindgren PE, Lofgren S: **Infection of human endothelial cells with Staphylococcus aureus induces transcription of genes encoding an innate immunity response.** *Scand J Immunol* 2005, **61:**536-544.
43. Voyich JM, Braughton KR, Sturdevant DE, Whitney AR, Said-Salim B, Porcella SF, Long RD, Dorward DW, Gardner DJ, Kreiswirth BN, Musser JM, DeLeo FR: **Insights into mechanisms used by Staphylococcus aureus to avoid destruction by human neutrophils.** *J Immunol* 2005, **175:**3907-3919.
44. Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, Cui L, Oguchi A, Aoki K, Nagai Y, Lian J, Ito T, Kanamori M, Matsumaru H, Maruyama A, Murakami H, Hosoyama A, Mizutani-Ui Y, Takahashi NK, Sawano T, Inoue R, Kaito C, Sekimizu K, Hirakawa H, Kuhara S, Goto S, Yabuzaki J, Kanehisa M, Yamashita A, Oshima K, Furuya K, Yoshino C, Shiba T, Hattori M, Ogasawara N, Hayashi H, Hiramatsu K: **Whole genome sequencing of meticillin-resistant Staphylococcus aureus.** *Lancet* 2001, **357:**1225-1240.
45. Baba T, Takeuchi F, Kuroda M, Yuzawa H, Aoki K, Oguchi A, Nagai Y, Iwama N, Asano K, Naimi T, Kuroda H, Cui L, Yamamoto K, Hiramatsu K: **Genome and virulence determinants of high virulence community-acquired MRSA.** *Lancet* 2002, **359:**1819-1827.
46. Holden MT, Feil EJ, Lindsay JA, Peacock SJ, Day NP, Enright MC, Foster TJ, Moore CE, Hurst L, Atkin R, Barron A, Bason N, Bentley SD, Chillingworth C, Chillingworth T, Churcher C, Clark L, Corton C, Cronin A, Doggett J, Dowd L, Feltwell T, Hance Z, Harris B, Hauser H, Holroyd S, Jagels K, James KD, Lennard N, Line A, Mayes R, Moule S, Mungall K, Ormond D, Quail MA, Rabbinowitsch E, Rutherford K, Sanders M, Sharp S, Simmonds M, Stevens K, Whitehead S, Barrell BG, Spratt BG, Parkhill J: **Complete genomes of two clinical Staphylococcus aureus strains: Evidence for the rapid evolution of virulence and drug resistance.** *Proc Natl Acad Sci U S A* 2004, **101:**9786-9791.
47. Gill SR, Fouts DE, Archer GL, Mongodin EF, Deboy RT, Ravel J, Paulsen IT, Kolonay JF, Brinkac L, Beanan M, Dodson RJ, Daugherty SC, Madupu R, Angiuoli SV, Durkin AS, Haft DH, Vamathevan J, Khouri H, Utterback T, Lee C, Dimitrov G, Jiang L, Qin H, Weidman J, Tran K, Kang K, Hance IR, Nelson KE, Fraser CM: **Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant Staphylococcus aureus strain and a biofilm-producing methicillin-resistant Staphylococcus epidermidis strain.** *J Bacteriol* 2005, **187:**2426-2438.
48. Diep BA, Gill SR, Chang RF, Phan TH, Chen JH, Davidson MG, Lin F, Lin J, Carleton HA, Mongodin EF, Sensabaugh GF, Perdreau-Remington F: **Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant Staphylococcus aureus.** *Lancet* 2006, **367:**731-739.
49. Herron LL, Chakravarty R, Dwan C, Fitzgerald JR, Musser JM, Retzel E, Kapur V: **Genome sequence survey identifies unique sequences and key virulence genes with unusual rates of amino Acid substitution in bovine Staphylococcus aureus.** *Infect Immun* 2002, **70:**3978-3981.
50. Said-Salim B, Mathema B, Kreiswirth BN: **Community-acquired methicillin-resistant Staphylococcus aureus: an emerging pathogen.** *Infect Control Hosp Epidemiol* 2003, **24:**451-455.
51. Beenken KE, Dunman PM, McAleese F, Macapagal D, Murphy E, Projan SJ, Blevins JS, Smeltzer MS: **Global gene expression in Staphylococcus aureus biofilms.** *J Bacteriol* 2004, **186:**4665-4684.
52. Bischoff M, Dunman P, Kormanec J, Macapagal D, Murphy E, Mounts W, Berger-Bachi B, Projan S: **Microarray-based analysis of the Staphylococcus aureussB regulon.** *J Bacteriol* 2004, **186:**4085-4099.
53. Charbonnier Y, Gettler BM, Francois P, Bento M, Renzoni A, Vaudaux P, Schlegel W, Schrenzel J: **A generic approach for the design of whole-genome oligoarrays, validated for genomotyping, deletion mapping and gene expression analysis on Staphylococcus aureus.** *BMC Genomics* 2005, **6:**95.
54. Dunman PM, Murphy E, Hanney S, Palacio D, Tucker-Kellogg G, Wu S, Brown EL, Zagurski RJ, Shlaes D, Projan SJ: **Transcription Profiling-Based Identification of Staphylococcus aureus Genes Regulated by the agr and/or sarA Loci.** *J Bacteriol* 2001, **183:**7341-7353.
55. Fitzgerald JR, Sturdevant DE, Mackie SM, Gill SR, Musser JM: **Evolutionary genomics of Staphylococcus aureus: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic.** *Proc Natl Acad Sci U S A* 2001, **98:**8821-8826.
56. Koessler T, Francois P, Charbonnier Y, Huyghe A, Bento M, Dharan S, Renzi G, Lew D, Harbarth S, Pittet D, Schrenzel J: **Use of Oligoarrays for Characterization of Community-Onset Methicillin-Resistant Staphylococcus aureus.** *J Clin Microbiol* 2006, **44:**1040-1048.
57. Nambu Y, Beer DG: **Altered surface markers in lung cancer. Lack of cell-surface Fas/APO-1 expression in pulmonary adenocarcinoma may allow escape from immune surveillance.** *Methods Mol Med* 2003, **74:**259-266.
58. Puskas LG, Zvara A, Hackler L Jr., Van Hummelen P: **RNA amplification results in reproducible microarray data with slight ratio bias.** *Biotechniques* 2002, **32:**1330-4, 1336, 1338, 1340.
59. Wilson CL, Pepper SD, Hey Y, Miller CJ: **Amplification protocols introduce systematic but reproducible errors into gene expression studies.** *Biotechniques* 2004, **36:**498-506.
60. **Genomic Research Laboratory** 2007 [http://www.genomic.ch/sup5.php].
61. NCBI: **Phylogenetic classification of proteins encoded in complete genomes.** *NCBI* 2006 [http://www.ncbi.nlm.nih.gov/COG/index.html].
62. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q: **VFDB: a reference database for bacterial virulence factors.** *Nucleic Acids Res* 2005, **33:**D325-D328.
63. Tegmark K, Karlsson A, Arvidson S: **Identification and characterization of SarH1, a new global regulator of virulence gene expression in Staphylococcus aureus.** *Mol Microbiol* 2000, **37:**398-409.
64. Ramakrishnan R, Dorris D, Lublinsky A, Nguyen A, Domanus M, Prokhorova A, Gieser L, Touma E, Lockner R, Tata M, Zhu X, Patterson M, Shippy R, Sendera TJ, Mazumder A: **An assessment of Motorola CodeLink microarray performance for gene expression profiling applications.** *Nucleic Acids Res* 2002, **30:**e30.
65. Plouin-Gaudon I, Clément S, Huggler E, Chaponnier C, Francois P, Vaudaux P, Lew D, Schrenzel J, Lacroix JS: **Intracellular sanctuary for Staphylococcus aureus in chronic rhinosinusitis.** *Rhinology* 2006, **44:**249-254.
66. Runyen-Janecky LJ, Payne SM: **Identification of chromosomal Shigella flexneri genes induced by the eukaryotic intracellular environment.** *Infect Immun* 2002, **70:**4379-4388.
67. Orihuela CJ, Radin JN, Sublett JE, Gao G, Kaushal D, Tuomanen EI: **Microarray analysis of pneumococcal gene expression during invasive disease.** *Infect Immun* 2004, **72:**5582-5596.
68. Pracht D, Elm C, Gerber J, Bergmann S, Rohde M, Seiler M, Kim KS, Jenkinson HF, Nau R, Hammerschmidt S: **PavA of Streptococcus pneumoniae modulates adherence, invasion, and meningeal inflammation.** *Infect Immun* 2005, **73:**2680-2689.

69. Eriksson S, Lucchini S, Thompson A, Rhen M, Hinton JC: **Unravelling the biology of macrophage infection by gene expression profiling of intracellular Salmonella enterica.** *Mol Microbiol* 2003, **47:**103-118.
70. Lucchini S, Liu H, Jin Q, Hinton JC, Yu J: **Transcriptional adaptation of Shigella flexneri during infection of macrophages and epithelial cells: insights into the strategies of a cytosolic bacterial pathogen.** *Infect Immun* 2005, **73:**88-102.
71. Boyce JD, Wilkie I, Harper M, Paustian ML, Kapur V, Adler B: **Genomic scale analysis of Pasteurella multocida gene expression during growth within the natural chicken host.** *Infect Immun* 2002, **70:**6871-6879.
72. Ramadurai L, Lockwood KJ, Nadakavukaren MJ, Jayaswal RK: **Characterization of a chromosomally encoded glycylglycine endopeptidase of Staphylococcus aureus.** *Microbiology* 1999, **145 (Pt 4):**801-808.
73. Weinrick B, Dunman PM, McAleese F, Murphy E, Projan SJ, Fang Y, Novick RP: **Effect of mild acid on gene expression in Staphylococcus aureus.** *J Bacteriol* 2004, **186:**8407-8423.
74. Skaar EP, Humayun M, Bae T, DeBord KL, Schneewind O: **Iron-source preference of Staphylococcus aureus infections.** *Science* 2004, **305:**1626-1628.
75. Frisk A, Schurr JR, Wang G, Bertucci DC, Marrero L, Hwang SH, Hassett DJ, Schurr MJ: **Transcriptome analysis of Pseudomonas aeruginosa after interaction with human airway epithelial cells.** *Infect Immun* 2004, **72:**5433-5438.
76. Sebulsky MT, Speziali CD, Shilton BH, Edgell DR, Heinrichs DE: **FhuD1, a ferric hydroxamate-binding lipoprotein in Staphylococcus aureus: a case of gene duplication and lateral transfer.** *J Biol Chem* 2004, **279:**53152-53159.
77. Lithgow JK, Hayhurst EJ, Cohen G, Aharonowitz Y, Foster SJ: **Role of a cysteine synthase in Staphylococcus aureus.** *J Bacteriol* 2004, **186:**1579-1590.
78. Frees D, Chastanet A, Qazi S, Sorensen K, Hill P, Msadek T, Ingmer H: **Clp ATPases are required for stress tolerance, intracellular replication and biofilm formation in Staphylococcus aureus.** *Mol Microbiol* 2004, **54:**1445-1462.
79. Cheung AL, Bayer AS, Zhang G, Gresham H, Xiong YQ: **Regulation of virulence determinants in vitro and in vivo in Staphylococcus aureus.** *FEMS Immunol Med Microbiol* 2004, **40:**1-9.
80. Bronner S, Monteil H, Prevost G: **Regulation of virulence determinants in Staphylococcus aureus: complexity and applications.** *FEMS Microbiol Rev* 2004, **28:**183-200.
81. Xiong YQ, Van Wamel W, Nast CC, Yeaman MR, Cheung AL, Bayer AS: **Activation and transcriptional interaction between agr RNAII and RNAIII in Staphylococcus aureus in vitro and in an experimental endocarditis model.** *J Infect Dis* 2002, **186:**668-677.
82. Moisan H, Brouillette E, Jacob CL, Langlois-Begin P, Michaud S, Malouin F: **Transcription of Virulence Factors in Staphylococcus aureus Small-Colony Variants Isolated from Cystic Fibrosis Patients Is Influenced by SigB.** *J Bacteriol* 2006, **188:**64-76.
83. von Eiff C, Proctor RA, Peters G: **Staphylococcus aureus small colony variants: formation and clinical impact.** *Int J Clin Pract Suppl* 2000:44-49.
84. von Eiff C, Heilmann C, Proctor RA, Woltz C, Peters G, Gotz F: **A site-directed Staphylococcus aureus hemB mutant is a small-colony variant which persists intracellularly.** *J Bacteriol* 1997, **179:**4706-4712.
85. Qazi SN, Counil E, Morrissey J, Rees CE, Cockayne A, Winzer K, Chan WC, Williams P, Hill PJ: **agr expression precedes escape of internalized Staphylococcus aureus from the host endosome.** *Infect Immun* 2001, **69:**7074-7082.
86. Balwit JM, van Langevelde P, Vann JM, Proctor RA: **Gentamicin-resistant menadione and hemin auxotrophic Staphylococcus aureus persist within cultured endothelial cells.** *J Infect Dis* 1994, **170:**1033-1037.
87. Li D, Renzoni A, Estoppey T, Bisognano C, Francois P, Kelley WL, Lew DP, Schrenzel J, Vaudaux P: **Induction of fibronectin adhesins in quinolone-resistant Staphylococcus aureus by subinhibitory levels of ciprofloxacin or by sigma B transcription factor activity is mediated by two separate pathways.** *Antimicrob Agents Chemother* 2005, **49:**916-924.
88. Renzoni A, Francois P, Li D, Kelley WL, Lew D, Vaudaux P, Schrenzel J: **Modulation of Fibronectin Adhesins and Other Virulence Factors in a Teicoplanin-Resistant Derivative of Methicillin-Resistant Staphylococcus aureus.** *Antimicrob Agents Chemother* 2004, **48:**2958-2965.
89. Francois P, Garzoni C, Bento M, Schrenzel J: **Comparison of Amplification Methods for Transcriptomic Analysis of Low Abundance Prokaryotic RNA sources.** *J Microbiol Methods* 2006, **In Press:**.
90. Pinard R, de Winter A, Sarkis GJ, Gerstein MB, Tartaro KR, Plant RN, Egholm M, Rothberg JM, Leamon JH: **Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing.** *BMC Genomics* 2006, **7:**216.
91. Talaat AM, Howard ST, Hale W, Lyons R, Garner H, Johnston SA: **Genomic DNA standards for gene expression profiling in Mycobacterium tuberculosis.** *Nucleic Acids Res* 2002, **30:**e104.
92. Chien Y, Manna AC, Projan SJ, Cheung AL: **SarA, a global regulator of virulence determinants in Staphylococcus aureus, binds to a conserved motif essential for sar-dependent gene regulation.** *J Biol Chem* 1999, **274:**37169-37176.

BMC Genomics

BioMed Central

Research article

Open Access

Exploring glycopeptide-resistance in *Staphylococcus aureus*: a combined proteomics and transcriptomics approach for the identification of resistance-related markers

Alexander Scherl[†1], Patrice François*[†2], Yvan Charbonnier[2], Jacques M Deshusses[1], Thibaud Koessler[2], Antoine Huyghe[2], Manuela Bento[2], Jianru Stahl-Zeng[3], Adrien Fischer[2], Alexandre Masselot[4], Alireza Vaezzadeh[1], Francesca Gallé[2], Adriana Renzoni[2], Pierre Vaudaux[2], Daniel Lew[2], Catherine G Zimmermann-Ivol[1,5], Pierre-Alain Binz[4], Jean-Charles Sanchez[1], Denis F Hochstrasser[1,5,6] and Jacques Schrenzel[2]

Address: [1]Biomedical Proteomics Research Group, Geneva University, Geneva, Switzerland, [2]Service of infectious diseases, University Hospitals of Geneva, Geneva, Switzerland, [3]Applied Biosystems, Darmstadt, Germany, [4]GeneBio SA, Geneva, Switzerland, [5]Central Clinical Chemistry Laboratory, University Hospitals of Geneva, Geneva, Switzerland and [6]Pharmacy section, Faculty of Sciences, Geneva University, Switzerland

Email: Alexander Scherl - ascherl@u.washington.edu; Patrice François* - patrice.francois@genomic.ch; Yvan Charbonnier - yvan.charbonnier@genomic.ch; Jacques M Deshusses - jacques.deshusses@bluewin.ch; Thibaud Koessler - thibaud.koessler@genomic.ch; Antoine Huyghe - antoine.huyghe@genomic.ch; Manuela Bento - manuela.bento@genomic.ch; Jianru Stahl-Zeng - jianru.stahl-zeng@eur.appliedbiosystems.com; Adrien Fischer - adrien.fischer@genomic.ch; Alexandre Masselot - olav@genebio.com; Alireza Vaezzadeh - alireza.vaezzadeh@medecine.unige.ch; Francesca Gallé - francesca.galle@genomic.ch; Adriana Renzoni - adriana.renzoni@genomic.ch; Pierre Vaudaux - pierre.vaudaux@hcuge.ch; Daniel Lew - daniel.lew@hcuge.ch; Catherine G Zimmermann-Ivol - catherine.zimmerman@medcli.unige.ch; Pierre-Alain Binz - pierre-alain.binz@isb-sib.ch; Jean-Charles Sanchez - jean-charles.sanchez@medcli.unige.ch; Denis F Hochstrasser - denis.hochstrasser@sim.unige.ch; Jacques Schrenzel - jacques.schrenzel@genomic.ch

* Corresponding author †Equal contributors

Published: 22 November 2006

BMC Genomics 2006, **7**:296 doi:10.1186/1471-2164-7-296

Received: 24 April 2006
Accepted: 22 November 2006

This article is available from: http://www.biomedcentral.com/1471-2164/7/296

Abstract

Background: To unravel molecular targets involved in glycopeptide resistance, three isogenic strains of *Staphylococcus aureus* with different susceptibility levels to vancomycin or teicoplanin were subjected to whole-genome microarray-based transcription and quantitative proteomic profiling. Quantitative proteomics performed on membrane extracts showed exquisite inter-experimental reproducibility permitting the identification and relative quantification of >30% of the predicted *S. aureus* proteome.

Results: In the absence of antibiotic selection pressure, comparison of stable resistant and susceptible strains revealed 94 differentially expressed genes and 178 proteins. As expected, only partial correlation was obtained between transcriptomic and proteomic results during stationary-phase. Application of massively parallel methods identified one third of the complete proteome, a majority of which was only predicted based on genome sequencing, but never identified to date. Several over-expressed genes represent previously reported targets, while series of genes and proteins possibly involved in the glycopeptide resistance mechanism were discovered here, including regulators, global regulator attenuator, hyper-mutability factor or hypothetical proteins. Gene expression of these markers was confirmed in a collection of genetically unrelated strains showing altered susceptibility to glycopeptides.

Conclusion: Our proteome and transcriptome analyses have been performed during stationary-phase of growth on isogenic strains showing susceptibility or intermediate level of resistance against glycopeptides. Altered susceptibility had emerged spontaneously after infection with a sensitive parental strain, thus not selected *in vitro*. This combined analysis allows the identification of hundreds of proteins considered, so far as hypothetical protein. In addition, this study provides not only a global picture of transcription and expression adaptations during a complex antibiotic resistance mechanism but also unravels potential drug targets or markers that are constitutively expressed by resistant strains regardless of their genetic background, amenable to be used as diagnostic targets.

Background

The Gram-positive bacterium *Staphylococcus aureus* is an important human pathogen that has become increasingly resistant to a wide range of antibiotics over the last two decades. The emergence of multidrug-resistant isolates of methicillin-resistant *S. aureus* (MRSA) exhibiting also decreased susceptibilities to glycopeptides (glycopeptide-intermediate *S. aureus*, GISA) represents a crucial challenge for antimicrobial therapy, antimicrobial susceptibility testing, and hospital infection control. After initial description in Japan of MRSA strains with decreased susceptibility to glycopeptides [1], clinical isolates showing similar phenotypes were repeatedly reported in various countries [2-6]. These strains are distinct from high-level glycopeptide resistant isolates (VRSA) that result from the acquisition of the *vanA* gene from *Enterococcus faecalis* [7]. Their potential spreading appears of particular concern since glycopeptides represent the last barrier drugs effective against MRSA. In addition, intensive use of glycopeptides will probably contribute to the selection of other resistant strains, as already observed for numerous antimicrobial agents [8].

Vancomycin is a natural product, isolated from the bacteria *Amycolatopsis orientalis* in the early fifties [9]. Binding of this molecule to the N-acyl D-ala-D-ala residue of bacterial peptidoglycan through five strong hydrogen bonds inhibits cross-linking of the cell-wall [10]. Vancomycin acts therefore immediately upstream of the transpeptidase, the target of β-lactam antibiotics. Structure and mechanism of action of teicoplanin are similar to that of vancomycin [10]. Little is known about the underlying mechanisms which produce GISA strains. A major common marker of GISA strains is the increased cell-wall thickness [11]: the prototype GISA strain shows 30–40 cross-linked peptidoglycan layers, whereas fully susceptible strains contain only 20 layers [12].

Approximately 20% of free D-ala-D-ala residues from the peptidoglycan structure remain unprocessed by penicillin-binding proteins (PBPs) in vancomycin-susceptible strains. It was therefore suggested that important quantities of glycopeptides are trapped by these free residues [12]. In addition, it was shown that the cross-linking rate is slower in laboratory-derived GISA compared to susceptible strains [13,14], which may increase the number of trapped vancomycin molecules and contribute to the destruction of the mesh-structure of the cell-wall [15]. This cooperative clogging phenomenon has been recently shown to prevent vancomycin from reaching its target in the cytoplasmic membrane of strain Mu50, the first clinical isolate reported as GISA [16]. However, analysis of the digested cell-wall compounds with high-performance liquid chromatography (HPLC) showed different peptidoglycan structures among various clinical isolates of GISA strains showing various cross-linking frequency [17]. In other backgrounds, originating from *in vitro* selection of laboratory strains, structural changes in cell-wall composition were limited [18] suggesting that there is no single genetic or biochemical change responsible for the decreased glycopeptide sensitivity. Several genes were indeed described to play a role in glycopeptide resistance. Penicillin-binding protein 4 (PBP4) is hypothesized to cleave the terminal D-alanine residue from un-cross-linked peptidoglycan chains [19]. A decrease in PBP4 activity, as observed in clinical isolates of GISA strains, should therefore result in a higher number of D-ala-D-ala targets for vancomycin [19]. PBP2 over-expression seems also correlated with decreased glycopeptide susceptibility [20,21]. The loss of function of the accessory gene regulator (*agr*) also increases resistance to glycopeptide [22]. A recent report demonstrated the contribution of *tcaA* in the resistance mechanism. This gene encodes for a transmembrane protein showing an intracellular metal-binding motif and a large extracellular domain of unknown function. While inactivation of this gene induces increased tolerance to glycopeptides, its presence induces overproduction of the same protein, suggesting that this protein acts as a sensor and/or as a signal transducer [23]. However, the impact of such an observation on clinical isolates is still debated [24]. The two-component system *vraRS* [25] has been shown to be involved in glycopeptide resistance. The expression of *vraRS* probably mediates positive regulation of cell-wall synthesis pathway in *S. aureus*, through induction of *pbp2* [21], thus increasing the sensitivity of GISA strains to cell-wall synthesis inhibitors [25]. In another study performed with clinical isolates and Mu50 or Mu50 derivative strains showing increased levels of resistance after passages onto vancomycin-containing medium, most of the genes involved in purine biosynthesis and transport were found up-regulated in the highly vancomycin-resistant strain as compared to the parental strain [26]. The deduced hypothesis relies on an increase in AMP pool able to generate ATP and compensate the difference in energy requirement observed between susceptible and resistant isolates.

Altogether, glycopeptide-resistance in *Staphylococcus aureus* appears due to multiple factors, including cell-wall synthesis and processing [27], autolysis [28,29], or regulatory events [23,24,30]. A major limitation of this type of comparative study is the difficulty to obtain isogenic strains showing various susceptibility phenotypes within the same genetic background and relevance in the context of human infections. Our group reported the isolation of GISA subpopulations emerging from a glycopeptide-susceptible parental strain in an experimental model of subcutaneous infection [31]. In the same model of infection, we then isolated a spontaneous revertant strain showing restored susceptibility to glycopeptides [32]. To improve

our understanding of mechanisms contributing to this spontaneous evolution, these three isogenic strains presenting different glycopeptide susceptibility levels were compared at the transcriptome and proteome levels during the stationary phase of growth. Transcriptional profiles were performed using a customized and extensively validated oligoarray [33]. In addition, mass spectrometry (MS)-based protein quantification was performed on membrane-enriched extracts. Enriched membrane extracts were chosen for proteomic experiments, since membrane proteins are often involved in bacterial antibiotic resistance and contribute also to numerous important metabolic pathways and transports, not only of nutrients but also of chemicals [34-38]. The recently introduced isobaric tagging technology was used for simultaneous quantification in all strains [39,40]. Isoelectric focusing (IEF) on immobilized pH gradients served as the first dimensional peptide separation [41], followed by LC-MS/MS analysis with a MALDI tandem MS instrument. These combined approaches performed on isogenic clinical isolates showing stable resistance, without antibiotic pressure revealed genes and proteins identified as potentially involved in the acquisition of glycopeptide resistance through the involvement of a complex and multi-factorial biological network. In addition, most of genes or proteins previously suspected to contribute to glycopeptide resistance were identified in our study, confirming the robustness of our approach. Finally, six potential targets found up-regulated at the transcript and protein levels are also shown to be over-expressed in a collection of *S. aureus* clinical isolates from unrelated genetic background showing reduced susceptibility to glycopeptides.

Results

Clonality of the studied strains

MLST performed with partial sequences of 7 genes revealed that strains MRGR3, 14-4 and 14-4Rev displayed the same allele profile 2.3.1.1.4.4.3 for *arcC*, *aroE*, *glpf*, *gmk*, *pta*, *tpi*, *yqiL*, respectively, corresponding to MLST 239.

To further confirm the clonality of these strains, comparative genome hybridization was performed on an oligoarray covering the whole genome of *Staphylococcus aureus* COL, N315, Mu50 and MW2. Additional file 1 illustrates that susceptible strain MRGR3 and the GISA strain 14-4 displayed strictly identical fluorescence patterns on the 5427 unique oligonucleotide probes [33].

Proteomic experiments

Analysis by 2-DE revealed striking similarities in protein expression of the two glycopeptide-susceptible strains (MRGR3 and 14-4Rev). On the opposite, the protein pattern between the glycopeptide intermediate strain (14-4) and both susceptible strains appeared totally different (Figure 1). To document this high variability, quantitative proteomic experiments were performed on membrane-enriched fractions with MS-based approaches, using isobaric tags. Isoelectric focusing of peptides was used as first dimension separation, prior to LC-MS/MS analysis. Enrichment was evaluated by assaying the lactate dehydrogenase activity, a marker of cytosolic contents. Enzymatic activity reached 1330 U/g in total protein extract and 60 U/g in membrane protein.

All quantitative proteomics experiments were performed in duplicates. In the first experiment (PR1), 3'724 unique peptides corresponding to 632 proteins were identified from the bacterial membrane fraction. In the second experiment (PR2), 3'719 peptides corresponding to 754 proteins were identified. A total of 551 proteins were commonly identified in both experiments. Together, these experiments yielded a total of 835 unique proteins, covering approximately 32% of the whole deduced proteome of *S. aureus* strain N315 (see Additional files 2, 3, 4).

Among the 2'575 predicted ORFs of strain N315, 637 protein products (24%) are predicted to be integral membrane proteins (e.g. containing at least one transmembrane domain). Among the proteins identified during these two experiments, approximately 20% are predicted to be integral membrane proteins.

Intense signal from the reporter fragment ions (derived from the reporter fragment ions abundance at m/z 114.1, 115.1, 116.1 and 117.1 Da) was obtained for almost all peptides (>95%). Relative quantification could therefore be performed on all 835 proteins differentially expressed between strain MRGR3 and 14-4, on 826 proteins

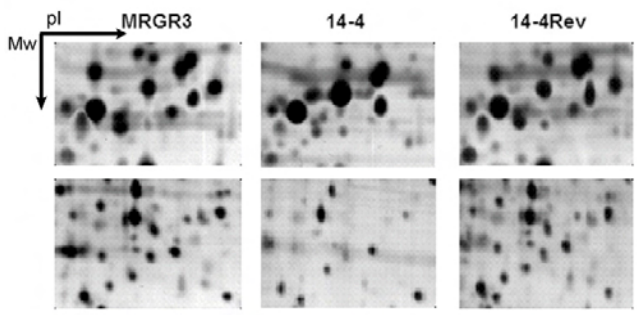


Figure 1
Two windows for each strain are represented to illustrate 2-D gel electrophoresis performed on membrane-enriched protein fractions of GISA (14-4) and glycopeptide-susceptible strains (MRGR3 and 14-4Rev). While similar patterns are observed between susceptible strains, drastic differences are visible when comparing the GISA with either of the susceptible strains.

expressed between 14-4 and 14-4Rev, and on 826 proteins between MRGR3 and 14-4Rev. A list of these proteins is provided separately (see Additional files 5, 6). Tables contain the corresponding ORF number, their relative abundance ratio as well as the coefficient of variation (CV) of ratios from individual peptides belonging to the same protein. In average, 4.9 peptides per protein were used for quantification. As shown in additional tables, the CV of individual peptide ratios belonging to the same protein is relatively high. When comparing glycopeptide-susceptible strains, mean CV value is 23.5%, whereas it increased to 32.6% between 14-4 and 14-4Rev, and 36.5% between 14-4 and MRGR3. Most identified peptides matched those of the nucleotide sequence-deduced [42] peptides of strain N315. However, when peptides were not found in N315 genome, ORF numbers from other sequenced *S. aureus* strains are provided in the tables (see Additional files 5, 6).

Scatter plots of the proteomics ratio revealed very similar profiles in both glycopeptide-susceptible strains (MRGR3 and 14-4Rev). The number of proteins out of the range -0.5 to 0.5 ($\log_{10}$ of protein expression ratio) was 0, and only 10 proteins appeared out of the range of -0.3 to 0.3 (Figure 2A). On the opposite, important modifications of the distribution were observed between the GISA strain and any of the 2 glycopeptide-susceptible ones (Figures 2B and 2C). The number of proteins found more abundant in one strain over the other was approximately 60 in the range of -0.5 to 0.5 and 160 in the range of -0.3 to 0.3, resulting in a scatter distribution instead of a 0-centered representation. The lists of identified proteins between the GISA strain and either of the susceptible strains revealed almost the same content as all but only 9 proteins were found in the two analyses, indicating excellent reproducibility of the membrane protein preparations. The experimental variability observed between quantifications of the two replicates was relatively important. Figures 2B and 2C depict several proteins with divergent expression ratios between replicates. Due to experimental variation, inherent to quantitative mass spectrometry techniques [43] and illustrated by the relative high CV values of individual peptides from the same protein, a conservative approach was selected that considered only proteins significantly differentially expressed in both experiments. To this end, we considered the distribution between both glycopeptide-susceptible strains (MRGR3 and 14-4Rev) as a Gaussian normal distribution. Threshold for significant differential expression was set at the 5th percentile of the most over- or under-expressed proteins (2.5 percentile most over-expressed and 2.5 percentile most under-expressed) between both glycopeptide-susceptible strains (MRGR3 and 14-4Rev).

A total of 155 proteins are differentially expressed between strains 14-4 and MRGR3, and 110 proteins between strains 14-4 and 14-4Rev, which corresponds to approximately 4% of the deduced proteome. Together, 178 unique proteins are differentially expressed in the GISA strain. A list of these proteins is given in supplementary material (see Additional files 5 and 6). Comparison reveals that more than 65% of these proteins are common between both lists obtained from comparison of the resistant strain with either of the two susceptible isolates. Among these 2 lists, the most important categories involved proteins playing a role in: energy metabolism, amino-acids transport, cell envelope biosynthesis, protein turnover and inorganic ion transport (corresponding to

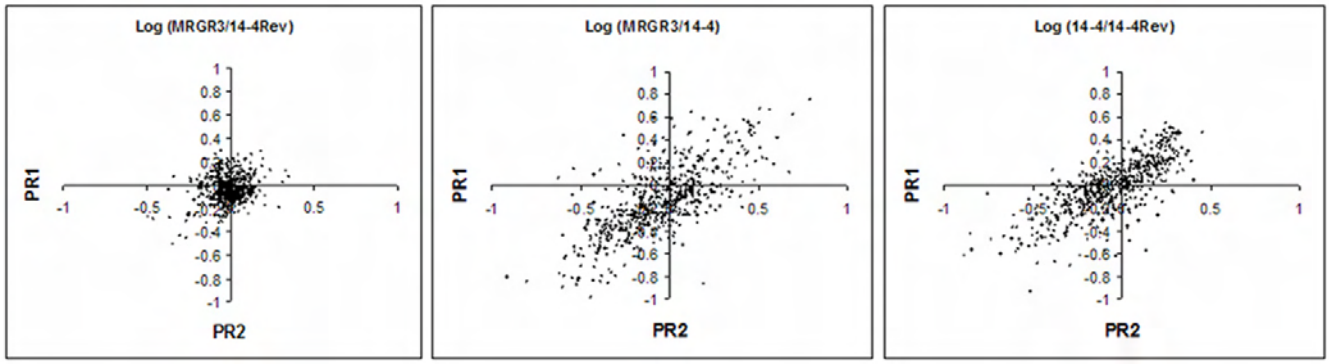


Figure 2
Dispersion of expression ratios obtained for two independent quantitative proteomic experiments illustrates the reproducibility of the whole procedure and the homogeneity observed when comparing the two susceptible strains (A). A good correlation but a large difference in protein expression is observed when comparing GISA and either MRGR3 or 14-4Rev (panels B and C with correlation coefficients of 0.92 and 0.93; respectively).

COG categories C, E, M, O, P, respectively). Among the 19 categories represented, these 5 categories constituted 40% of the identified proteins.

Transcriptomic analysis

Genes with statistically significant changes in the level of expression and differences superior to two-fold changes were identified and listed for pairwise comparisons (see additional files 7, 8). According to these criteria, only 10 genes were differentially regulated between strain MRGR3 and 14-4Rev, consisting in 0.4% of the transcriptome. However, strain 14-4 compared to MRGR3 showed 67 differentially regulated genes, whereas 14-4 compared to 14-4Rev showed 64. In total, 94 differentially expressed genes were identified between GISA and both glycopeptide-susceptible strains. Finally, these tables reveal that more than 65% of these genes were common. Among the up- or down-regulated genes, the most represented transcripts were found in COG E, K and P corresponding to genes involved in amino acids transport, transcription and inorganic ions transport, respectively. Thus, 2 out of these 3 categories were also found predominant in the quantitative proteome study.

Correlation between transcriptomic and proteomic data

The transcriptomic expression trends compared to the proteomic quantification for all identified proteins are shown in Tables 1 and 2 for the comparison between the two sensitive strains and the GISA. This analysis reveals that an equivalent proportion of genes are regulated similarly in both comparisons.

The comparison between the proteomic expression direction (Tables 1 and 2) and the corresponding transcriptomic expression direction according to the functional classes (according to the COG database) shows important similarities for most categories. However, two functional classes (COG O and P, corresponding respectively to post-translational modification, protein turnover, chaperones and inorganic ion transport) show opposed quantification trends (Figure 3A and 3B).

Table 1: Transcriptomic expression direction for the differentially expressed proteins between strains 14-4 and MRGR3.

		Proteomic		
		Up	**=**	**Down**
Transcriptomic	**Up**	57		12
	=	-	684	
	Down	36		26

Table 2: Transcriptomic expression direction for the differentially expressed proteins between strains 14-4 and 14-4Rev.

		Proteomic		
		Up	**=**	**Down**
Transcriptomic	**Up**	30		12
	=		718	
	Down	15		31

Finally, strict comparison of significantly differentially expressed proteins and genes allowed us to identify 7 common ORFs (Tables 3 and 4) showing the same transcript and proteins expression trends. Among these genes, 5 are over-expressed whereas only two are under-expressed. In addition, SA0532 that revealed strongly up-regulated in microarray experiments was also evaluated in quantitative analysis.

Genetic relationships between strains showing glycopeptide intermediate level of resistance and identified markers expression

The expression ratio of the five over-expressed ORFs at the gene and protein level in glycopeptide-intermediate strain 14-4 as well as the over-expressed transcript SA0532 was measured on a collection of genetically unrelated strains (Table 5). Rapid genotype clustering (Figure 4A) shows that our collection of strains displays at least five distinct genomic contents confirming that these strains are not genetically related. Cluster A composed of Mu50 and Mu3 shows some relatedness with NRS17, a strain isolated in the USA. This relatedness between the two Japanese isolates has been documented previously [44]. Cluster B contains the closely-related French isolates. Patterns C and F contain only a single isolate (NRS3 and NRS35, respectively). Cluster D confirms that our set of isogenic strains is unique, and appeared clonal which is in accordance with the MLST and CGH results (see Additional file 1), and partially related to cluster E. Finally, the 3 strains extensively documented in the present study appear clonal but clearly different from at least 4 other genotypes identified. As previously documented using pulse-field gel electrophoresis, our three strains appeared isogenic [45].

Expression levels of the 6 mRNA markers potentially involved in the mechanism of glycopeptide resistance, assessed by quantitative PCR using specific oligonucleotides (Table 6) is shown on Figure 4B. Important differences were observed between strains and their genetic backgrounds. All reference strains (e.g. N315, COL) showed marginal differences in terms of gene expression

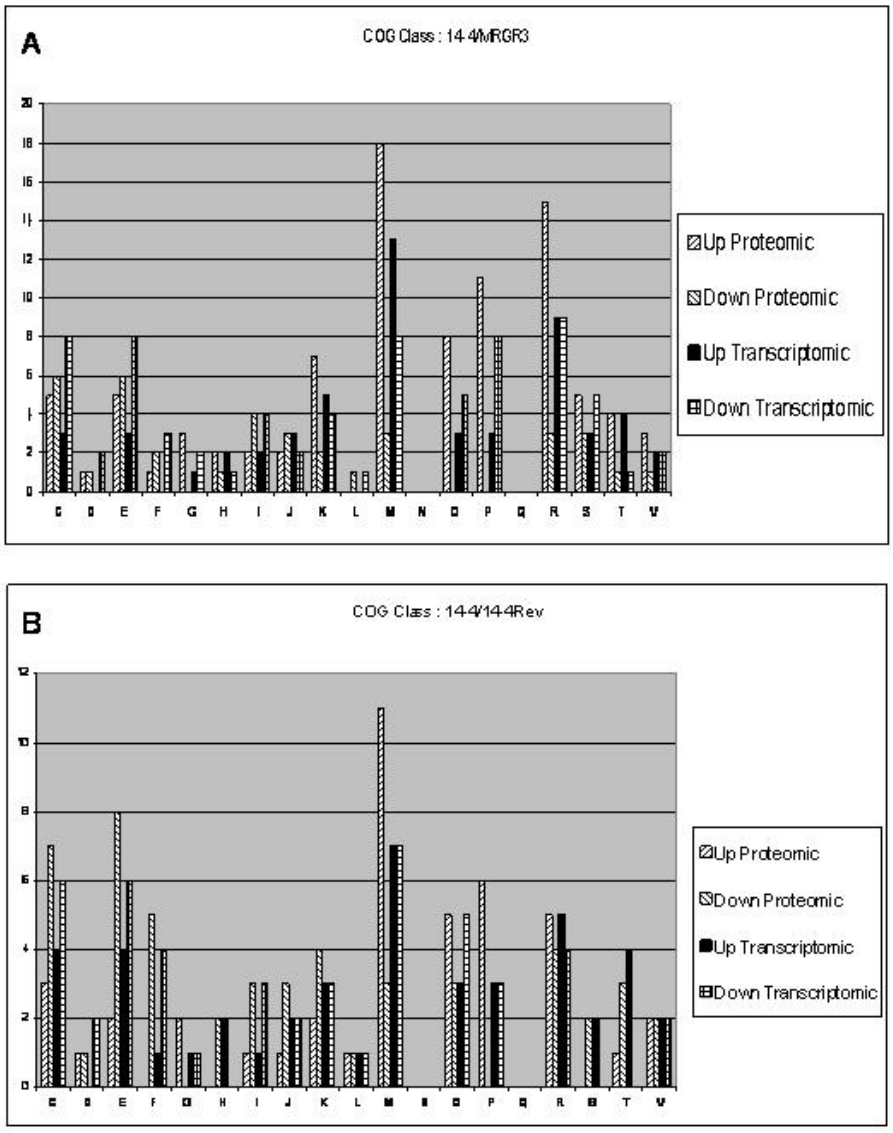


Figure 3
Differentially expressed proteins with the corresponding transcripts expression trend regrouped by functions for the comparison between GISA strain 14-4, and MRGR3 or 14-4Rev (A and B). Similar general profiles were obtained for the two comparisons between the GISA and either of the susceptible strains while totally different profile was observed for the two susceptible strains. Functional classes are: C = energy production and conversion; D = cell division and chromosome partitioning; E = amino acid transport and metabolism; F = nucleotide transport and metabolism; G = carbohydrate transport and metabolism; H = coenzyme metabolism; I = lipid metabolism; J = translation, ribosomal structure and biogenesis; K = transcription; L = DNA replication, recombination and repair; M = cell envelope biogenesis, outer membrane; N = cell motility; O = post-translational modification, protein turnover, chaperones; P = inorganic ion transport and metabolism; Q = secondary metabolites biosynthesis, transport and catabolism; R = general function prediction; S = function unknown; T = signal transduction mechanisms; U = secretion; V = defense mechanism. Number of proteins for each category is indicated on top of the corresponding columns.

Table 3: List of significantly differentially expressed genes corresponding to over- or under-expressed proteins between 14-4 and MRGR3.

ORF number	Function	Transcript	Protein
SA0536.1	vraX	up	up
SA0591	hypothetical protein	up	up
SA1195	peptide methionine sulfoxide reductase regulator MsrR	up	up
SA2113	hypothetical protein	up	up

compared to other strains reported as glycopeptides-intermediate. The most variable markers were SA0536.1, SA1691 and SA0591, which revealed importantly up-regulated in the vast majority of unrelated GISA strains. A similar expression pattern of these markers was found in our resistant strain 14-4 as in the ancestral GISA prototype, Mu50 and its clonal derivative Mu3 showing important up-regulation of all identified markers. Taken together, these results illustrate the benefits of performing combined transcriptomic and proteomic analysis. Indeed, both were necessary to obtain a global and holistic view of the state of the bacterial cells. On the opposite, analysis restricted to gene-expression or protein abundance may lead to important bias.

Discussion

Technical issues and challenges

The use of isobaric-tagging during isoelectric focusing prior to LC-MS/MS provided the possibility to quantitatively examine a large number of proteins for differential expression [39,40]. Using membrane-enriched protein extracts from *Staphylococcus aureus* during stationary phase, we consistently identified and quantified relative protein abundance for approximately 32% of the proteome when comparing pairs of strains. Previously, we showed that only 23% of the proteome was accessible using different combinations of protein or peptide fractionation and numerous separation techniques [46]. The capacity to obtain extended proteome recovery, particularly from membrane fractions, represents one of the major bottlenecks in proteomic experiments. As expected from multiple previous reports, numerous soluble contaminants are identified from the membrane-enriched fractions. However, these compounds greatly enhance the number of identified proteins and contribute to the global picture of the proteome from a single cellular fraction. In addition, the removal of important fractions of abundant cytosolic proteins increases the number of low-abundance proteins linked to the bacterial membrane [46]. In this report, we concentrated on membrane-enriched fraction. Our membrane purification process allowed reducing drastically the content of cell-wall anchored and cytosolic proteins, which constitute the most abundant fraction of cellular proteins by a 200-fold factor. Based on the content of *Staphylococcus aureus* proteome, this corresponds to approximately 4% contamination of membrane fraction by cytosolic proteins Western blot performed with antibodies raised against *S. aureus* protein A showed that the contamination by cell-wall anchored proteins was marginal. The new approach with an IEF separation of peptides leads to the identification of an appreciable part of the bacterium proteome, maximizing the chances to discover new targets involved in *S. aureus* glycopeptide resistance. In this respect, our analysis of *S. aureus* proteome allowed satisfactory representation of low-abundance or membrane proteins. In this work, 19% of the identified proteins constituted integral membrane proteins, a value close to the predicted transmembrane proteome (19.7% of proteome showed at least 2 putative transmembrane domains) as deduced from the nucleotide sequence [42,46]. Numerous examples of bacterial resistance against different antibiotic families involve membrane

Table 4: List of significantly differentially expressed genes corresponding to over- or under-expressed proteins between 14-4 and 14-4Rev.

ORF number	Function	Transcript	Protein
SA0022	hypothetical protein, similar to 5'-nucleotidase	down	down
SA0591	hypothetical protein	up	up
SA0977	cell surface protein	down	down
SA1195	peptide methionine sulfoxide reductase regulator MsrR	up	up
SA1691	similar to penicillin-binding protein 1A/1B	up	up
SA2113	hypothetical protein	up	up

SA0591. Shows the **VLTHEFGHVL** motif, a characteristic of neutral Zn metallo-protease. Membrane associated protein.
SA2113. Shows partial homology with DNA repair protein from *Psychrobacter arcticus* and is particularly conserved among the different *Staphylococcus aureus* strains (often strictly identical nucleotide sequence).
SA0536.1. 1–19 N-terminal amino-acids are signal peptide (probably exported).

Table 5: Characteristics of strains subjected to real-time PCR quantification

Strain Name	CMI		Characteristics and origin
	Vanco	Teico	
NRS 3	4	8	Peritoneal fluid, USA
NRS 11	2	2	Ophthalmologic infection, France
NRS 12	8	8	Ophthalmologic infection, France
NRS 13	4	4	Ophthalmologic infection, France
NRS 14	8	8	Ophthalmologic infection, France
NRS 17	8	8	Bloodstream, USA
NRS 35	2	2	Bloodstream, France
NRS 56	8	8	Wound/skin/soft tissue, Brazil
NRS 63	8	8	Bloodstream, Oman
NRS 64	2	2	Bloodstream, Oman
Mu50 (NRS1)	8	8	Wound/skin/soft tissue, Japan [42]
Mu3 (NRS2)	0.5	2	Purulent sputum, Japan
MRGR3	1	1	Bloodstream, Switzerland
14-4	8	8	Sub-cutaneous infection with MRGR3, Switzerland [31]
14-4Rev	1	1	Sub-cutaneous infection with 14-4, Switzerland [32]
MW2	1	1	Community-acquired *Staphylococcus aureus* [68]
COL	0.5	1	Ancestor strain of MRSA, England [80]
N315	0.5	1	Pharyngeal smear, Japan [42]

proteins [34-37]. Membrane proteins are notoriously difficult to solubilize during tryptic digestion and, depending on the number of transmembrane domains, most of the fractionated peptides remain highly hydrophobic which is an important limitation during isoelectric focusing [47]. Due to important differences between the three strains analyzed in this study, this aspect is of utmost importance. The mean CV of individual peptide ratios belonging to the same protein varies between 23.5% (ratios between MRGR3 and 14-4Rev) and 36.5% (ratios between 14-4 and susceptible strains) which is relatively low, based on the complexity of the whole experimental procedure. The sample variability certainly contributed to these variations, as we observed higher CVs for the comparison between susceptible and GISA strains. The high peptide recovery due to combination of IEF, LC-MS/MS and labeling strategy is therefore essential for reliable protein quantification. Similar quantitative profiles were obtained in each individual experiment, as shown on Figure 2. However, totally different profiles were observed between either of the two susceptible strains and the GISA (Figures 2A and 2B). This observation is concordant with 2D-gel electrophoresis results which clearly illustrates divergent patterns between GISA and the susceptible strains.

Whereas high-level glycopeptide resistance in *S. aureus* has been shown to rely on the horizontal transfer of *vanA* gene from *Enterococcus faecalis* [7,48], the mechanisms involved in glycopeptide-intermediate resistance remain poorly understood and numerous potential mechanisms are currently explored [11,27]. They probably involve a complex and global gene expression and protein translation changes as well as protein processing alterations. The most documented modifications contributing to this phenotype involve profound modifications of cell-wall biosynthesis including increased cell-wall thickness, cross-linking, or accumulation of cell-wall components [11,27]. In this study, several targets involved in cell-wall synthesis were found differentially expressed either at the protein or transcript level, or at both levels. Other potential hypothesis rely on the increased abundance of D-Ala-D-Ala residues, forming false-target trapping glycopeptides or to an altered regulation of cell-wall synthesis [12,14]. Other potential mechanisms are related to the requirement of increased energy synthesis related to larger cell-wall contents, involving phosphoglycerate kinase [49], or purine metabolism [26].

In a cell, protein abundance is not strictly correlated to that of its cognate mRNA levels. Thus, a combined transcriptomics and proteomics approach is warranted for genome-wide identification of molecular targets involved in the resistance mechanism. Previous experiments performed in our laboratory showed that resistance to glycopeptides expressed by strain 14-4 remains stable even after numerous sub-cultures in antibiotic-free medium [31]. This resistant phenotype relies on constitutive changes in gene expression and is stable, thus not requiring drug induction. At stationary phase, a decoupling between protein and transcripts abundance is generally observed except for most of constitutive processes [50], supporting our exploration of the transcriptome and proteome of *S. aureus* during this stage.

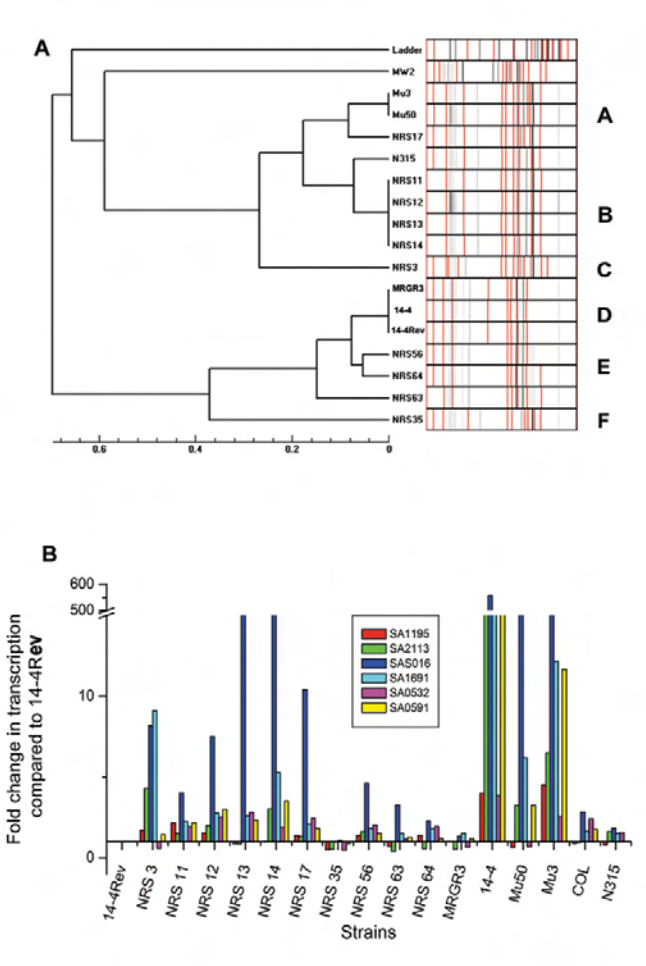


Figure 4

A. Genotyping of strains evaluated by VNTR. Based on the limit of clonality of the assay (0.1 unit of divergence) [77], our collection contains at least 4 unrelated genetic backgrounds (noted A-D). Control strains whose genome has been sequenced have been simultaneously analyzed. **B.** Gene expression of identified markers on a collection of strains showing altered susceptibility to glycopeptides. Relative expression of 6 gene markers found up-regulated in the resistant strain 14-4 compared to the two sensitive strains. The same markers were also quantified in a collection of strains from the NARSA, selected for their resistance to glycopeptide. Values are mean + standard deviation of triplicate measurements from 2 independent experiments. Increase expression of selected markers, linked to glycopeptide resistance is independent on the background. The complete raw microarray dataset has been posted on the Gene Expression Omnibus database [78], accession number GSE5188. All additional files are freely accessible [79].

Differentially expressed targets in GISA compared to susceptible strains

Proteomics as well as transcriptomics studies revealed moderate numbers of differentially regulated genes or proteins. Comparison of GISA strain with either of the 2 susceptible isolates showed a majority of common differentially expressed targets. In addition, a perfect correlation was observed for those common genes in the transcriptomic evaluation and the proteomic studies. Genes or gene products known to be involved in different antibiotic resistance mechanisms, such as PBP2 (SA1283/Q7A5K8), Ant (SA2385/P04827), MurE (SA0876/P65480), and the methicillin resistance-related autolysis protein FmtA (SA0909/Q7A6A2) were over-expressed in the GISA strain. In addition to these targets involved in cell-wall biosynthesis, many other proteins involved in cell-wall metabolism (formation or post-processing and hydrolysis) were found over-expressed, such as SgtB (SA1691/Q7A4S6) and SsaA (SA2093/Q7A423). Part of these compounds involved in cell-wall processing was reported to alter glycopeptide susceptibility [27]. These results were shown in gene expression study [27] and are now confirmed on protein expression levels.

Another important protein target found differentially expressed at the mRNA and protein levels is the product of *vraX* gene (SA0536.1/Q99W32, accession number AB050664). This gene belongs to the *vra* operon (SA0533–SA0535) encoding for genes initially described to be involved in imipenem resistance. Most of these genes were up-regulated at the mRNA levels together with SA0532 and SA0536.1, located upstream and downstream of the *vra* region, respectively and suggesting contribution in glycopeptide resistance. Motive search revealed that both genes, SA0532 and SA0536 harbored a putative phosphorylation site, suggesting a role in a regulatory process. Regulatory systems contributed to the GISA phenotype, such as the two-component sensor histidine kinase, coded by the *vraSR* genes (SA1701-SA1700), previously shown as important in the development of resistance to either imipenem or glycopeptides [25]. Our study reveals also that protein regulators are over-expressed, such as the signal transduction protein TRAP (Q7A4W3/SA1653) [51], acting as a signal-transducer protein during quorum sensing [51,52], the divIVA protein (SA1279/Q7A5L1) known to regulate cell division, or the putative transcription factor (SA2296/Q7A3J2) and *sarH1*, whose up-regulation is in accordance with a decreased expression of *spa*. Finally, the over-expression of proteins involved in stress protection such as proteinases CtpA (SA1253/Q7A5M9), MsrA and MsrB (SA1257/P99065 and SA1256/P65446) and the regulator MsrR (SA1195/Q99Q02), contributing to the reparation of proteins inactivated by oxidation. These attenuators of transcription, mainly expressed during exponential phase are over-

Table 6: List and characteristics of primers and probes used for validation of identified markers.

Gene number[A]	Sequence (5'→3')	Length	5'-Dye	Concentration (nM)
SA0591				
F_SA0591-195F	TGGCCGTGCAAAAGATTTAGT	21		200
R_SA0591-275R	TGTGTAATTGCGAATCCTTGTTG	23		200
P_SA0591-219T	TGTAACAAGTCCACGAGAGCGACAACAGA	29	FAM	100
SA2113				
F_SA2113-40F	GCCGGTTTAAATGACCGTACTACT	24		200
R_SA2113-117R	TTGCAACGCATCACAAAATTG	21		200
P_SA2113-75T	CACTTTTGCATAACTTACTT[B]	20	VIC	100
SA1195				
F_SA1195-799F	CCTGATTCAGGGATTTTCCAAA	22		200
R_SA1195-877R	TTGGCACAGTCAATGACTTAACATC	25		200
P_SA1195-822T	TCACCTCGGATGCCAAAACTCAAACCT	27	FAM	100
SA1691				
F_SA1691-319F	CATCATGGATTCGATTTGAAAGG	23		200
R_SA1691-389R	CCTTGCACATCTCTGTCGCTAA	22		200
P_SA1691-344T	CAACTAGAGCTTTATTTTCA[B]	20	FAM	100
SA0536.1				
F_SA0536.1-31F	GGCGCACCAGTTTATGAAATTAT	23		200
R_SA0536.1-111R	TTCAGTATCACTAAATGAATCGTCACAT	28		200
P_SA0536.1-60T	AACGTTTCAGCATGTTT[B]	17	VIC	100
SA0532				
F_SA0532-15F	GATTGCGATAACTTCTTTTCTCTAT	25		200
R_SA0532-115R	AATCAAACTTAAGCCCTAGATAAA	24		200
P_SA0532-51T	CAGTATCAAATTTATCTAGGGCTTA[B]	25	VIC	100
rRNA16S				
F_rRNA16s-551F	GGCAAGCGTTATCCGGAATT	20		200
R_rRNA16S-651R	GTTTCCAATGACCCTCCACG	20		200
P_rRNA16s-573T	CCTACGCGCGCTTTACGCCCA	21	FAM	100

[A] Based on the annotation of N315 genome.
[B] Minor groove binder (MGB) probes with non-fluorescent quencher bound to the 3'-end (Applied Biosystems).

expressed by the GISA strain at the stationary phase and are also abundant at the protein level. This observation suggests that the GISA strain is constitutively in a stress state, despite absence of antibiotic exposure.

Some proteins involved in purine biosynthesis were found slightly down-regulated in 14-4 compared to sensitive strains at the protein level, whereas none of the genes involved in this operon was found differentially expressed by microarray. This observation is in contradiction with a recent report showing a massive up regulation of the *pur* operon [26] in GISA strains isolated *in vitro*. Our study, performed on *ex vivo* derived clinical isolates involves probably different mechanisms yielding to the resistant phenotype. These differences illustrate also the importance of growth conditions and antibiotic resistance selection. Indeed, these different conditions allow pointing out complementary mechanisms leading to reduced susceptibility to glycopeptides. In our comparison, metabolic targets showing differential expression between GISA and susceptible strains concerned mainly the metabolism of arginine. Arginine repressor is up-regulated and consequently argininosuccinate lyase expression is reduced as observed previously in response to cell-wall active antibiotics [53]. The potential effect is the decreased succinate and fumarate levels and the accumulation of aspartate. Aspartate is a central compound involved in numerous metabolic functions such as amino acids synthesis, urea cycle or energetic transports and cell-wall synthesis. Another category of targets found differentially regulated in our study is ABC transporters. Some of these compounds have been proposed as targets for immunotherapy [38]. In our study, 4 ABC transporters were found down-regulated in GISA compared to susceptible strain, suggesting that the GISA phenotype does not involve this type of target.

Correlation of proteomic and transcriptomic profiles

In order to compare the transcriptomic profiles with the proteomic expression ratios, the differentially expressed proteins were compared to trend of gene-expression ratio (Tables 1 and 2). A significant part of the over-expressed proteins corresponds to down-regulated transcripts and suggests altered protein degradation. Comparison of GISA with either of the sensitive strains showed similarities between up- or down-regulated functions between the transcriptome and the proteome for all functional classes. However, these tendencies are inverted for the COG func-

tions O and P, corresponding to post-translational modification, protein turnover, chaperones; and inorganic ion transport and metabolism, respectively. Indeed, most proteins in these functional groups are up-regulated whereas most transcripts are down-regulated. Similarities are observed by comparing the glycopeptide intermediate strain with the two susceptible strains. The analysis revealed also that most of the expressions have the same trend in proteomic and in transcriptomic (mainly related to a reduced metabolism in the GISA) but some regions showed different directions. These regions are probably composed of genes showing altered turn-over rates [50,54], as previously observed [46]. Finally, genes and proteins found differentially expressed showing the same trend represent potential targets for rapid diagnostic of evolution to GISA phenotype. These lists contain SA0536.1 (*vraX*), SA0591 and SA2113, two hypothetical proteins and *msrR*. Additionally, SA0532 showing an elevated difference in expression between 14-4 and MRGR3 appeared also as a potential marker of interest. The expression at the mRNA level was evaluated in a collection of unrelated strains showing altered susceptibility to glycopeptides. Among our collection, 5 unrelated genetic backgrounds were identified and all of them displayed intermediate level of resistance to glycopeptides. The level of expression of 6 markers found up-regulated in our *in vivo* isolated strain was very similar to that observed in the ancestral GISA strains. One of these markers, SA0536.1, has been very recently identified as up-regulated in GISA strains [30]. In accordance with this report, SA0536.1 was found up-regulated in all the genetically unrelated strains of our study. However this study observed that the up-regulation of SA0536.1 is not related to vancomycin induction [30] but is constitutive in all our strains. McAleese suggested as "provocative findings" that the expression of several markers is constitutive as revealed by our work.

Compared to the different studies published in the field and elaborating on potential mechanisms involved in glycopeptide resistance, this study was performed using GISA strains recovered in the absence of antibiotic pressure. However, most genes previously discovered in different GISA strains, either selected *in vitro* or *in vivo*, were also documented in our study. To the involvement of cell-wall biosynthesis as contributor of the GISA phenotype, our analysis now adds the discovery of new targets present in higher abundance in the GISA strains. Our determination revealed also the contribution of the *msr* locus, as well as that of several transporters that could contribute to a more direct resistance mechanism. In that sense, the presence at the protein levels of these targets appeared as a strong evidence. Thus, our results suggest that glycopeptide resistance is a stable phenomenon acquired by *Staphylococcus aureus* and expressed in the absence of any selection pressure. Finally, coupled to a rapid genotyping method, we identified markers potentially involved in the resistance mechanisms, potentially amenable to be used as diagnostic markers.

Conclusion

Combined proteomic and transcriptomic analyses allowed obtaining a global view of complex processes involving differentially regulated factors contributing to antibiotic resistance. This combined information is essential for the global integration of the data. Several potential genes and proteins leading to glycopeptide resistance were identified. This study provides the identity of targets potentially related to the bacterial evolution toward a GISA phenotype, thus a contribution to the elucidation of complex mechanisms leading to glycopeptide resistance. In view of the results found in the literature and additional information obtained in this study, we showed that our genetic background appeared particularly relevant and that multiple mechanisms are mobilized by *Staphylococcus aureus* to evolve to the GISA phenotype. In addition, several markers reflecting the potential evolution of *Staphylococcus aureus* to the GISA phenotype have been identified.

Methods

Reagents and chemicals

All chemicals purchased were of the highest purity grade, unless otherwise stated. MilliQ water (Millipore, Bedford, MA) was used for the preparation of all buffers and solvents. Methanol, hydrochloric acid, magnesium chloride, potassium chloride, saccharose were purchased from Merck (Darmstadt, Germany). Acetonitrile (AcN) was purchased from Biosolve (Valkenswaard, The Netherland). High boiling-point petroleum ether, SDS, orthophosphoric acid and 2,2,2-trifluoroethanol (99.0%) were purchased from Fluka (Buchs, Switzerland). Trifluoroacetic acid (TFA), α-cyano-4-hydroxycinnamic acid, 1,4-dithioerythritol (DTE), ammonium bicarbonate, potassium chloride, potassium dihydrogenophosphate, iodoacetamide, glycerol, glycine, phosphate buffered saline, porcine trypsin and Tris were from Sigma-Aldrich (St. Louis, MO). IPG strips (3–10, nonlinear, 18 cm) were purchased from GE Healthcare (Piscataway, NJ). Agarose, ampholines (3–10) and molecular mass markers were purchased from BioRad (Hercules, CA). Mueller Hinton broth was obtained from Difco (Detroit, MI), and saccharose from Merck (Darmstadt, Germany). The murolytic enzyme lysostaphin (Ambicin) was purchased from Applied Microbiology Inc (Tarrytown, NY). dCTP coupled to cyanine dyes were obtained from NEN (Perkin Elmer, Boston, MA, USA).

Strains, growth conditions and time point

MRSA strain MRGR3, a glycopeptide-susceptible strain was isolated from a patient with catheter-related sepsis

[31]; it is highly pathogenic in a rat model of chronic infections [55]. Strain 14-4, a stable isogenic GISA strain was recovered from an experimental infection with MRGR3 [31]. Strain 14-4Rev that spontaneously emerged during infection with strain 14-4 is a revertant glycopeptide-susceptible isolate [32]. For protein extracts, strains MRGR3, 14-4 and 14-4Rev were grown with agitation at 37°C in Mueller Hinton Broth (MHB; 200 mL in 1000-mL flask), as previously described [56]. At stationary phase ($OD_{540\ nm}$ = 6 corresponding to 2–3 × 10^9 cells/ml), cells were chilled on ice and harvested by centrifugation at 8'000 × *g* for 5 min at 4°C. For preparation of crude membrane extracts, 20 mL culture aliquots were washed in 1.1 M saccharose-containing buffer [46], then suspended in 2 mL aliquots of the same buffer containing 50 μg/mL of the hydrolytic enzyme lysostaphin for 10 min at 37°C. Protoplasts were recovered after centrifugation (30 min at 8,000 × *g*) and hypo-osmotic shock was applied in the presence of 10 μg/mL DNase I (Fluka, Buchs, Switzerland) to decrease the viscosity of the medium. Crude membrane pellets were obtained after ultracentrifugation at 50,000 × *g* for 50 min in a Beckman Optima TLX (Beckman Coulter Intl SA, Nyon, Switzerland).

Evaluation of membrane extracts purity-Evaluation of membrane enrichment was performed by assaying the lactate dehydrogenase following a previously described method [57,58] in a LX20 Beckman-Coulter (Beckman Coulter Intl SA, Nyon, Switzerland). In addition, total and membrane protein extracts (20 μg) were separated on SDS-PAGE, transferred to PVDF membrane [59] then incubated with anti-protein A (clone spa-27, Sigma, at a 1:750 dilution). Anti-mouse IgG coupled to phosphatase (1:7500 dilution) was used to reveal specific binding. A large 50 kD smeary band was obtained in total protein extracts whereas in membrane-enriched fraction, the staining was close to the limit of detection of the Western blot procedure (see Additional file 9).

A strain collection showing altered susceptibility levels to glycopeptides (Table 5) was obtained from the Network on Antimicrobial Resistance in *Staphylococcus aureus* (NARSA, Virginia, USA) [60].

Proteomics experiments

Two-dimensional gel electrophoresis was performed using previously described conditions [47]. Quantitative-MS based proteomic experiments were performed in duplicates, on two individual cultures. For each experiment, an estimated amount of 300 μg of crude membrane protein extract (BCA method, Pierce, Rockford, IL) from each strain was separately dissolved, reduced, alkylated, digested and labeled with iTRAQ compounds according to the manufacturer's procedure (Applied Biosystems, Framingham MD). For the first proteomic experiment (PR1), strain MRGR3 was labeled with iTRAQ 114, strain 14-4 with iTRAQ 116 and 14-4Rev with iTRAQ 117. For the second proteomic replicate experiment (PR2), strain MRGR3 was labeled with iTRAQ 117, strain 14-4 with iTRAQ 114 and 14-4Rev with iTRAQ 115. After mixing, peptides from all strains were concentrated and desalted using an Oasis HLB 1 cc 10 mg solid-phase extraction cartridge (Waters, Milford, MA). Desalted peptides were re-suspended in 300 μl isoelectric focusing buffer containing 4 M urea and 0.5% ampholines in 50% TFE. The IPG strips were rehydrated overnight with the peptide solution. Isoelectric focusing was performed with the following conditions: Linear gradient from 0 to 3'500 V in 3 hours, and 3'500 V during 20 hours. After isoelectric focusing, the IPG strip was washed 3 times 10 seconds in 3 distinct baths containing high boiling point petroleum ether in order to remove the paraffin oil. The strip was then manually cut in 58 (PR1) and 64 (PR2) fractions with a scalpel, and gel pieces were placed in polypropylene tubes containing 70 μl 0.1% TFA. After 30 minutes, TFA solution was transferred into another tube and replaced with 0.1% TFA in 50% Acetonitrile (AcN). After 30 minutes, the solution was also transferred to the second tube and replaced with 0.1% TFA in AcN, also transferred after 30 minutes. The combined solution was then dry-evaporated, re-suspended in 25 μl HPLC buffer A (0.1% formic acid in 5% AcN) and stored at -20°C. A volume of 5 μl of peptide solution of each fraction was loaded on a 10 cm long home-made column with an ID of 100 μm, packed with C_{18} reverse phase (YMS-ODS-AQ200, Michrom BioResource, Auburn, CA), and eluted directly on a MALDI target using a home-made spotting robot. The elution gradient ranged from 4% to 38% solvent B (0.1% formic acid in 80% AcN) in 40 minutes. Peptides were analyzed in MS and MS/MS mode using a 4700 MALDI-TOF/TOF tandem mass spectrometer (Applied Biosystems, Framingham, MA).

After MS/MS analysis, peak lists from each fraction were created with embedded software (4700 explorer 2.0 peak-to-mascot) and merged together before database searching using the Phenyx platform from GeneBio (Geneva, Switzerland). Searching was performed against a home-made database containing all predicted ORFs from genome-sequenced strain N315 [42] and proteins from other *S. aureus* and *S. epidermidis* strains with less than 90% identity (5515 entries extracted from UniProt (release 46, 25 Jan. 2005) and TrEMBL (release 29, 25 Jan. 2005)). Multiple-peptide hits were accepted with individual peptide z-score higher than 6, single-peptide hits with z-scores higher than 8, corresponding to a predicted peptide false-positive ratio of respectively less than 6% and 3.5%. These values were obtained after plotting false positive rates against true positive rates on a ROC-like curve. False positive hits where obtained from searches in the

entire SwissProt TrEMBL database without species restriction and true positive were selected from validated, high-score peptides matching to *S. aureus* entries.

Areas of the iTRAQ reporter ions for each peptide were extracted with Phenyx software. Peptides with highest iTRAQ reporter fragment ion areas below 2000 (arbitrary unit from "peak to mascot" software) were excluded. Mean peptide expression ratio was mathematically centered on value 1. Before prediction of transmembrane (TM) segments, the signal peptide of the predicted protein sequences was removed using the SignalP V2.0 tool [61], accessible through the internet [62]. The number of TM segments was then predicted using the TMHMM2.0 tool [63], also accessible through the internet [64]. Functional classes of all predicted ORFs were retrieved from the COG database [65,66], publicly available from the internet [67].

Microarray manufacturing

The microarray was manufactured by *in situ* synthesis of 8'454 long oligonucleotide probes (Agilent, Palo Alto, CA, USA), selected as previously described [33]. It covers >99% of all ORFs annotated in strains N315, Mu50 [42], MW2 [68] and COL [69]. Briefly, the microarray contains 8'192 *S. aureus* specific oligonucleotides. Based on available sequence information 89% of the probes are common to the 4 strains used for the design, whereas 11% are strain-specific capture elements. Extensive experimental validation of this array has been described previously, using CGH, mapping of deletion and quantitative RT-PCR [33].

Preparation of labeled nucleic acids

S. aureus strains were grown overnight in MHB, as described for proteomics analysis. Total RNA was extracted from 2 mL of cells at 2–3 × 10^9 cells/ml, using the RNeasy kit (Qiagen, Basel, Switzerland), as previously described [32,33]. After additional DNase treatment, the absence of remaining DNA traces was evaluated by quantitative PCR (SDS 7700; Applied Biosystems, Framingham, MA) with assays specific for 16s rRNA and the *HU* genes [32,70], encoding for a DNA-binding protein. Batches of 10 μg total *S. aureus* RNA were labeled by Cy-3 dCTP using the SuperScript II (Invitrogen, Basel, Switzerland) following manufacturer's instructions. Labeled was then purified onto QiaQuick columns (Qiagen). Purified genomic DNA from the 4 sequenced strains was extracted (DNeasy, Qiagen), labeled with Cy-5 dCTP using the Klenow fragment of DNA polymerase I (BioPrime, Invitrogen, Carlsbad, CA) [33,71].

Hybridization and scanning parameters

Cy5-labeled DNA (0.125 μg per stain) and Cy3-labeled cDNA (10 μg) mixture was diluted in 250 μl Agilent hybridization buffer, and hybridized at a temperature of 60°C for 17 hours in a dedicated hybridization oven (Robbins Scientific, Sunnyvale, CA, USA). Slides were washed, dried under nitrogen flow, and scanned (Agilent, Palo Alto, CA, USA) using 100% PMT power for both wavelengths. Data were extracted and processed using Feature Extraction™ software (version 6.1.1, Agilent). For expression analysis, local background-subtracted signals were corrected for unequal dye incorporation or unequal load of labeled product. The algorithm consisted of a rank consistency filter and a curve fit using the default LOWESS (locally weighted linear regression) method. Data consisting of three independent biological experiments were expressed as Log_{10} ratios and analyzed using GeneSpring 7.0 (SiliconGenetics, Redwood City, CA, USA). Statistical significance of differentially expressed genes was identified by variance analysis (ANOVA) [72], performed using GeneSpring, including the Benjamini and Hochberg false discovery rate correction (5%).

Real-time PCR validation of discovered markers

Gene-specific probes were designed using Primer Express 2.0 (Applied Biosystems). Oligonucleotide primers and probes (Table 6) obtained from Eurogentec (Seraing, Belgium) or Applied Biosystems (minor groove binder coupled to dark quencher) were solubilized in water and reactions were assembled in a one-step RT-PCR enzymatic mixture (Invitrogen) in a final volume of 15 μl. Reaction was performed in a SDS 7500 (Applied Biosystems). All measurements were performed in triplicate from two independent samples of purified RNA (0.5 ng/reaction) obtained from overnight and log-phase cultures isolated as previously described [33,73]. Cycle thresholds were assessed using default parameters. Briefly, the standard deviation of fluorescent values recorded from cycles 3–15 was multiplied by 10 to define the cycle threshold line. Cycle thresholds (Ct) were then derived from the intercept between this line and the signal obtained during the RT-PCR reaction. Results were normalized using intensity levels recorded for the rRNA 16s gene as previously described [32]. Figure shows relative gene expression for all strains as compared to 14-4Rev.

Genotyping of the related strains – Multi-locus Sequence Typing (MLST) was performed using previously described procedure and primers [74]. Allele numbers were assigned according to MLST Web site [75]. Comparative genome hybridization was performed according to the previously described procedure [76].

Genotyping of Staphylococcus aureus *collection by a variable number of tandem repeat (VNTR) approach*

A previously published method was used to evaluate the genomic content of strains composing the analyzed collection [76,77]. Briefly, this assay is based on a multiplex

PCR using ten primer pairs targeting genes showing variable number of tandem repeats. This method shows at least similar discriminatory power as that of pulse field-gel electrophoresis [77]. In addition to the NARSA collection, fully sequenced control strains were subjected to the assay.

Authors' contributions

AS performed all proteomic experiments including data analysis and interpretation. MB performed transcriptomic experiments. PF contributed to the design, analysis and interpretation of transcriptomic and expression experiments. YC, TK and AH were involved in transcriptomic and proteomic data comparison. AF and FG performed genotyping and Taqman experiments including data analysis. JMD participated to the design and optimization of the isoelectric focusing procedure. PV and DL were involved in revising the article critically. ARV contributed to estimate membrane enrichment procedure, AR contributed to the isogenic characterization of strains. JSZ helped with the use of iTRAQ compounds. AM and PAB optimized the protein identification and quantification software for this study as well as the corresponding statistical analysis. CGZI and JCS participated in the design and coordination of the study. DH and JS designed the study and coordinated it. All authors read and approved the final manuscript.

Additional material

Additional file 1

Comparative genome hybridization using microarray. CGH performed with labeled gDNA from strains MRGR3 and 14-4. showing that strains MRGR3 and 14-4 generated strictly identical patterns.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-7-296-S1.pdf]

Additional file 2

Raw quantitative proteomic data obtained for the comparison between MRGR3 vs 14-4Rev. Table showing quantitative proteomics measurements of protein expression between strains MRGR3 (parental strain) and 14-4Rev (susceptible revertant of GISA 14-4)
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-7-296-S2.pdf]

Additional file 3

Raw quantitative proteomic data obtained for the comparison between MRGR3 vs 14-4. Table showing quantitative proteomics measurements of protein expression between strains MRGR3 (parental strain) and 14-4 (GISA)
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-7-296-S3.pdf]

Additional file 4

Raw quantitative proteomic data obtained for the comparison between 14-4 vs 14-4Rev. Table showing quantitative proteomics measurements of protein expression between strains 14-4 (GISA) and 14-4Rev (susceptible revertant of GISA 14-4)
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-7-296-S4.pdf]

Additional file 5

Relative quantification of all identified proteins and trend of mRNA expression obtained by microarray for the comparison between strains MRGR3 and 14-4. Table showing the comparison between protein expression and trend of mRNA expression measured by microarray for strain MRGR3 (parental) strain and 14-4 (GISA)
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-7-296-S5.pdf]

Additional file 6

Relative quantification of all identified proteins and trend of mRNA expression obtained by microarray for the comparison between strains 14-4 and 14-4Rev. Table showing the comparison between protein expression and trend of mRNA expression measured by microarray for strain 14-4 (GISA) strain and 14-4Rev (susceptible revertant of GISA 14-4)
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-7-296-S6.pdf]

Additional file 7

Differentially expressed transcripts between 14-4 and MRGR3. Table showing all differentially expressed genes measured by using microarray between strains 14-4 (GISA) and MRGR3 (parental strain)
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-7-296-S7.pdf]

Additional file 8

Differentially expressed transcripts between 14-4 and 14-4Rev. Table showing all differentially expressed genes measured by using microarray between strains 14-4 (GISA) and 14-4Rev (susceptible revertant of GISA 14-4)
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-7-296-S8.pdf]

Additional file 9

Evaluation of membrane purity by Western blot. Coomassie blue stained SDS-PAGE and Western blot performed with anti-protein A antibody allowing to evaluate enrichment in membrane proteins of the protein fractions.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-7-296-S9.pdf]

Acknowledgements

This work was supported by grants 4049-63250 from National Research Program NRP49 ("Antibiotic Resistance"), 3200B0-108401 (to D.L.), 3200B0-103951 to (P.V.), PP00B-103002/1 (to J.S.) and 404940-106296/1 (to P.F.).

References

1. Hiramatsu K, Aritaka N, Hanaki H, Kawasaki S, Hosoda Y, Hori S, Fukuchi Y, Kobayashi I: **Dissemination in Japanese hospitals of strains of *Staphylococcus aureus* heterogeneously resistant to vancomycin.** *Lancet* 1997, **350:**1670-1673.
2. Kim MN, Pai CH, Woo JH, Ryu JS, Hiramatsu K: **Vancomycin-intermediate Staphylococcus aureus in Korea.** *J Clin Microbiol* 2000, **38:**3879-3881.
3. Tenover FC, Lancaster MV, Hill BC, Steward CD, Stocker SA, Hancock GA, O'Hara CM, McAllister SK, Clark NC, Hiramatsu K: **Characterization of staphylococci with reduced susceptibilities to vancomycin and other glycopeptides.** *J Clin Microbiol* 1998, **36:**1020-1027.
4. Trakulsomboon S, Danchaivijitr S, Rongrungruang Y, Dhiraputra C, Susaemgrat W, Ito T, Hiramatsu K: **First Report of Methicillin-Resistant Staphylococcus aureus with Reduced Susceptibility to Vancomycin in Thailand.** *J Clin Microbiol* 2001, **39:**591-595.
5. Woods CW, Cheng AC, Fowler VG Jr, Moorefield M, Frederick J, Sakoulas G, Meka VG, Tenover FC, Zwadyk P, Wilson KH: **Endocarditis caused by *Staphylococcus aureus* with reduced susceptibility to vancomycin.** *Clin Infect Dis* 2004, **38:**1188-1191.
6. Borg MA, Zerafa R, Morrison D, Cuschieri P: **Incidence of glycopeptide hetero-intermediate *Staphylococcus aureus* strains in Maltese hospitals.** *Clin Microbiol Infect* 2005, **11:**405-407.
7. Weigel LM, Clewell DB, Gill SR, Clark NC, McDougal LK, Flannagan SE, Kolonay JF, Shetty J, Killgore GE, Tenover FC: **Genetic analysis of a high-level vancomycin-resistant isolate of *Staphylococcus aureus*.** *Science* 2003, **302:**1569-1571.
8. Chambers HF: **The Changing Epidemiology of *Staphylococcus aureus*?** *Emerg Infect Dis* 2001, **7:**178-182.
9. McCormick MH, McGuire JM, Pittenger GE, Pittenger RC, STARK WM: **Vancomycin, a new antibiotic. I. Chemical and biologic properties.** *Antibiot Annu* 1955, **3:**606-611.
10. Barna JC, Williams DH: **The structure and mode of action of glycopeptide antibiotics of the vancomycin group.** *Annu Rev Microbiol* 1984, **38:**339-357.
11. Cui L, Ma X, Sato K, Okuma K, Tenover FC, Mamizuka EM, Gemmell CG, Kim MN, Ploy MC, El Solh N, *et al.*: **Cell wall thickening is a common feature of vancomycin resistance in *Staphylococcus aureus*.** *J Clin Microbiol* 2003, **41:**5-14.
12. Hiramatsu K: **Vancomycin-resistant *Staphylococcus aureus*: a new model of antibiotic resistance.** *Lancet Infect Dis* 2001, **1:**147-155.
13. Sieradzki K, Tomasz A: **Gradual alterations in cell wall structure and metabolism in vancomycin-resistant mutants of *Staphylococcus aureus*.** *J Bacteriol* 1999, **181:**7566-7570.
14. Sieradzki K, Tomasz A: **Inhibition of cell wall turnover and autolysis by vancomycin in a highly vancomycin-resistant mutant of *Staphylococcus aureus*.** *J Bacteriol* 1997, **179:**2557-2566.
15. Cui L, Murakami H, Kuwahara-Arai K, Hanaki H, Hiramatsu K: **Contribution of a thickened cell wall and its glutamine nonamidated component to the vancomycin resistance expressed by *Staphylococcus aureus* Mu50.** *Antimicrob Agents Chemother* 2000, **44:**2276-2285.
16. Cui L, Iwamoto A, Lian JQ, Neoh HM, Maruyama T, Horikawa Y, Hiramatsu K: **Novel Mechanism of Antibiotic Resistance Originating in Vancomycin-Intermediate *Staphylococcus aureus*.** *Antimicrob Agents Chemother* 2006, **50:**428-438.
17. Boyle-Vavra S, Labischinski H, Ebert CC, Ehlert K, Daum RS: **A spectrum of changes occurs in peptidoglycan composition of glycopeptide-intermediate clinical *Staphylococcus aureus* isolates.** *Antimicrob Agents Chemother* 2001, **45:**280-287.
18. Koehl JL, Muthaiyan A, Jayaswal RK, Ehlert K, Labischinski H, Wilkinson BJ: **Cell wall composition and decreased autolytic activity and lysostaphin susceptibility of glycopeptide-intermediate *Staphylococcus aureus*.** *Antimicrob Agents Chemother* 2004, **48:**3749-3757.
19. Finan JE, Archer GL, Pucci MJ, Climo MW: **Role of penicillin-binding protein 4 in expression of vancomycin resistance among clinical isolates of oxacillin-resistant *Staphylococcus aureus*.** *Antimicrob Agents Chemother* 2001, **45:**3070-3075.
20. Boyle-Vavra S, Yin S, Challapalli M, Daum RS: **Transcriptional induction of the penicillin-binding protein 2 gene in *Staphylococcus aureus* by cell wall-active antibiotics oxacillin and vancomycin.** *Antimicrob Agents Chemother* 2003, **47:**1028-1036.
21. Yin S, Daum RS, Boyle-Vavra S: ***VraSR* two-component regulatory system and its role in induction of *pbp2* and *vraSR* expression by cell wall antimicrobials in *Staphylococcus aureus*.** *Antimicrob Agents Chemother* 2006, **50:**336-343.
22. Sakoulas G, Eliopoulos GM, Moellering RC Jr, Wennersten C, Venkataraman L, Novick RP, Gold HS: **Accessory gene regulator (agr) locus in geographically diverse *Staphylococcus aureus* isolates with reduced susceptibility to vancomycin.** *Antimicrob Agents Chemother* 2002, **46:**1492-1502.
23. Maki H, McCallum N, Bischoff M, Wada A, Berger-Bachi B: ***tcaA* inactivation increases glycopeptide resistance in *Staphylococcus aureus*.** *Antimicrob Agents Chemother* 2004, **48:**1953-1959.
24. Wootton M, MacGowan AP, Walsh TR: **Expression of *tcaA* and *mprF* and glycopeptide resistance in clinical glycopeptide-intermediate *Staphylococcus aureus* (GISA) and heteroGISA strains.** *Biochim Biophys Acta* 2005, **1726:**326-327.
25. Kuroda M, Kuroda H, Oshima T, Takeuchi F, Mori H, Hiramatsu K: **Two-component system VraSR positively modulates the regulation of cell-wall biosynthesis pathway in *Staphylococcus aureus*.** *Mol Microbiol* 2003, **49:**807-821.
26. Mongodin E, Finan J, Climo MW, Rosato A, Gill S, Archer GL: **Microarray transcription analysis of clinical Staphylococcus aureus isolates resistant to vancomycin.** *J Bacteriol* 2003, **185:**4638-4643.
27. Cui L, Lian JQ, Neoh HM, Reyes E, Hiramatsu K: **DNA microarray-based identification of genes associated with glycopeptide resistance in *Staphylococcus aureus*.** *Antimicrob Agents Chemother* 2005, **49:**3404-3413.
28. Boyle-Vavra S, Challapalli M, Daum RS: **Resistance to autolysis in vancomycin-selected *Staphylococcus aureus* isolates precedes vancomycin-intermediate resistance.** *Antimicrob Agents Chemother* 2003, **47:**2036-2039.
29. Renzoni A, Barras C, Francois P, Charbonnier Y, Huggler E, Garzoni C, Kelley WL, Majcherczyk P, Schrenzel J, Lew D, *et al.*: **Transcriptomic and functional analysis of an autolytic-deficient teicoplanin-resistant derivative of methicillin-resistant *Staphylococcus aureus*.** *Antimicrob Agents Chemother* 2006, **50:**3048-61.
30. McAleese F, Wu SW, Sieradzki K, Dunman P, Murphy E, Projan S, Tomasz A: **Overexpression of genes of the cell wall stimulon in clinical isolates of *Staphylococcus aureus* exhibiting vancomycin-intermediate-*S. aureus*-type resistance to vancomycin.** *J Bacteriol* 2006, **188:**1120-1133.
31. Vaudaux P, Francois P, Berger-Bachi B, Lew DP: **In vivo emergence of subpopulations expressing teicoplanin or vancomycin resistance phenotypes in a glycopeptide-susceptible, methicillin-resistant strain of *Staphylococcus aureus*.** *J Antimicrob Chemother* 2001, **47:**163-170.
32. Renzoni A, Francois P, Li D, Kelley WL, Lew D, Vaudaux P, Schrenzel J: **Modulation of Fibronectin Adhesins and Other Virulence Factors in a Teicoplanin-Resistant Derivative of Methicillin-Resistant *Staphylococcus aureus*.** *Antimicrob Agents Chemother* 2004, **48:**2958-2965.
33. Charbonnier Y, Gettler BM, Francois P, Bento M, Renzoni A, Vaudaux P, Schlegel W, Schrenzel J: **A generic approach for the design of whole-genome oligoarrays, validated for genomotyping, deletion mapping and gene expression analysis on *Staphylococcus aureus*.** *BMC Genomics* 2005, **6:**95.
34. Cambau E, Gutmann L: **Mechanisms of resistance to quinolones.** *Drugs* 1993, **45(Suppl 3):**15-23.
35. Brown MH, Skurray RA: **Staphylococcal multidrug efflux protein QacA.** *J Mol Microbiol Biotechnol* 2001, **3:**163-170.
36. Putman M, van Veen HW, Konings WN: **Molecular properties of bacterial multidrug transporters.** *Microbiol Mol Biol Rev* 2000, **64:**672-693.
37. Kwon DH, Lu CD: **Polyamines induce resistance to cationic peptide, aminoglycoside, and quinolone antibiotics in *Pseudomonas aeruginosa* PAO1.** *Antimicrob Agents Chemother* 2006, **50:**1615-1622.
38. Otto M, Gotz F: **ABC transporters of staphylococci.** *Res Microbiol* 2001, **152:**351-356.
39. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Johnstone R, Neumann T, Hamon C: **Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.** *Anal Chem* 2003, **75:**1895-1904.

40. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, *et al.*: **Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents.** *Mol Cell Proteomics* 2004, **3:**1154-1169.
41. Cargile BJ, Talley DL, Stephenson JL Jr: **Immobilized pH gradients as a first dimension in shotgun proteomics and analysis of the accuracy of pI predictability of peptides.** *Electrophoresis* 2004, **25:**936-945.
42. Kuroda M, Ohta T, Uchiyama I, Baba T, Yuzawa H, Kobayashi I, Cui L, Oguchi A, Aoki K, Nagai Y, *et al.*: **Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*.** *Lancet* 2001, **357:**1225-1240.
43. Anderle M, Roy S, Lin H, Becker C, Joho K: **Quantifying reproducibility for differential proteomics: noise analysis for protein liquid chromatographymass spectrometry of human serum.** *Bioinformatics* 2004, **20:**3575-3582.
44. Ohta T, Hirakawa H, Morikawa K, Maruyama A, Inose Y, Yamashita A, Oshima K, Kuroda M, Hattori M, Hiramatsu K, *et al.*: **Nucleotide substitutions in *Staphylococcus aureus* strains, Mu50, Mu3, and N315.** *DNA Res* 2004, **11:**51-56.
45. Majcherczyk PA, McKenna T, Moreillon P, Vaudaux P: **The discriminatory power of MALDI-TOF mass spectrometry to differentiate between isogenic teicoplanin-susceptible and teicoplanin-resistant strains of methicillin-resistant Staphylococcus aureus.** *FEMS Microbiol Lett* 2006, **255:**233-239.
46. Scherl A, Francois P, Bento M, Deshusses JM, Charbonnier Y, Converset V, Huyghe A, Walter N, Hoogland C, Apel RD, *et al.*: **Correlation of proteomic and transcriptomic profiles of Staphylococcus aureus during the post-exponential phase of growth.** *Journal Microbiological Methods* 2005, **60:**247-257.
47. Deshusses JM, Burgess JA, Scherl A, Wenger Y, Walter N, Converset V, Paesano S, Corthals GL, Hochstrasser DF, Sanchez JC: **Exploitation of specific properties of trifluoroethanol for extraction and separation of membrane proteins.** *Proteomics* 2003, **3:**1418-1424.
48. Chang S, Sievert DM, Hageman JC, Boulton ML, Tenover FC, Downes FP, Shah S, Rudrik JT, Pupp GR, Brown WJ, *et al.*: **Infection with vancomycin-resistant *Staphylococcus aureus* containing the *vanA* resistance gene.** *N Engl J Med* 2003, **348:**1342-1347.
49. MacKenzie FM, Greig P, Morrison D, Edwards G, Gould IM: **Identification and characterization of teicoplanin-intermediate *Staphylococcus aureus* blood culture isolates in NE Scotland.** *J Antimicrob Chemother* 2002, **50:**689-697.
50. Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN: **Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays.** *Proc Natl Acad Sci USA* 2002, **99:**9697-9702.
51. Balaban N, Goldkorn T, Gov Y, Hirshberg M, Koyfman N, Matthews HR, Nhan RT, Singh B, Uziel O: **Regulation of *Staphylococcus aureus* pathogenesis via target of RNAIII-activating protein (TRAP).** *J Biol Chem* 2001, **276:**12476.
52. Gov Y, Borovok I, Korem M, Singh VK, Jayaswal RK, Wilkinson BJ, Rich SM, Balaban N: **Quorum sensing in Staphylococci is regulated via phosphorylation of three conserved histidine residues.** *J Biol Chem* 2004, **279:**14665-14672.
53. Utaida S, Dunman PM, Macapagal D, Murphy E, Projan SJ, Singh VK, Jayaswal RK, Wilkinson BJ: **Genome-wide transcriptional profiling of the response of *Staphylococcus aureus* to cell-wall-active antibiotics reveals a cell-wall-stress stimulon.** *Microbiology* 2003, **149:**2719-2732.
54. Gerth U, Kirstein J, Mostertz J, Waldminghaus T, Miethke M, Kock H, Hecker M: **Fine-tuning in regulation of Clp protein content in *Bacillus subtilis*.** *J Bacteriol* 2004, **186:**179-191.
55. Lucet JC, Herrmann M, Rohner P, Auckenthaler R, Waldvogel FA, Lew DP: **Treatment of experimental foreign body infection caused by methicillin-resistant *Staphylococcus aureus*.** *Antimicrob Agents Chemother* 1990, **34:**2312-2317.
56. Scherl A, Francois P, Converset V, Bento M, Burgess JA, Sanchez JC, Hochstrasser DF, Schrenzel J, Corthals GL: **Nonredundant mass spectrometry: a strategy to integrate mass spectrometry acquisition and analysis.** *Proteomics* 2004, **4:**917-927.
57. Amador E, Dorfman LE, Wacker WE: **Serum Lactic dehydrogenase activity: an analytical assessment of current assays.** *Clin Chem* 1963, **12:**391-399.
58. Ulmer DD, Vallee BL, Wacker WE: **Metalloenzymes and myocardial infarction. II. Malic and lactic dehydrogenase activities and zinc concentrations in serum.** *N Engl J Med* 1956, **255:**450-456.
59. McDevitt D, Francois P, Vaudaux P, Foster TJ: **Molecular characterization of the clumping factor (fibrinogen receptor) of *Staphylococcus aureus*.** *Mol Microbiol* 1994, **11:**237-248.
60. **Network on Antimicrobial Resistance in *Staphylococcus aureus*** [http://www.narsa.net]
61. Nielsen H, Engelbrecht J, Brunak S, Von Heijne G: **Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.** *Protein Eng* 1997, **10:**1-6.
62. **Center for biological sequence analysis** [http://www.cbs.dtu.dk/services/SignalP-2.0/]
63. Sonnhammer EL, Von Heijne G, Krogh A: **A hidden Markov model for predicting transmembrane helices in protein sequences.** *Proc Int Conf Intell Syst Mol Biol* 1998, **6:**175-182.
64. **Center for biological sequence analysis** [http://www.cbs.dtu.dk/services/TMHMM-2.0/]
65. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV: **The COG database: new developments in phylogenetic classification of proteins from complete genomes.** *Nucleic Acids Res* 2001, **29:**22-28.
66. Tatusov RL, Koonin EV, Lipman DJ: **A genomic perspective on protein families.** *Science* 1997, **278:**631-637.
67. **Phylogenetic classification of proteins encoded in complete genomes** [http://www.ncbi.nlm.nih.gov/COG/index.html]
68. Baba T, Takeuchi F, Kuroda M, Yuzawa H, Aoki K, Oguchi A, Nagai Y, Iwama N, Asano K, Naimi T, *et al.*: **Genome and virulence determinants of high virulence community-acquired MRSA.** *Lancet* 2002, **359:**1819-1827.
69. **Comprehensive Microbial Resource (CMR)** [http://cmr.tigr.org/tigr-scripts/CMR/CmrHomePage.cgi]
70. Garzoni C, Francois P, Couzinet S, Tapparel C, Charbonnier Y, Huyghe A, Renzoni A, Lucchini S, Lew DP, Vaudaux P, *et al.*: **A global view of the *Staphylococcus aureus* whole genome expression upon internalization in human epithelial cells.** 2006. submitted
71. Talaat AM, Howard ST, Hale W, Lyons R, Garner H, Johnston SA: **Genomic DNA standards for gene expression profiling in *Mycobacterium tuberculosis*.** *Nucleic Acids Res* 2002, **30:**e104.
72. Churchill GA: **Using ANOVA to analyze microarray data.** *Biotechniques* 2004, **37:**173-5, 177.
73. Vaudaux P, Francois P, Bisognano C, Kelley WL, Lew DP, Schrenzel J, Proctor RA, McNamara PJ, Peters G, von Eiff C: **Increased expression of clumping factor and fibronectin-binding proteins by *hemB* mutants of *Staphylococcus aureus* expressing small colony variant phenotypes.** *Infect Immun* 2002, **70:**5428-5437.
74. Enright MC, Day NP, Davies CE, Peacock SJ, Spratt BG: **Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*.** *J Clin Microbiol* 2000, **38:**1008-1015.
75. **Multi-locus sequence typing** [http://www.mlst.net]
76. Koessler T, Francois P, Charbonnier Y, Huyghe A, Bento M, Dharan S, Renzi G, Lew D, Harbarth S, Pittet D, *et al.*: **Use of Oligoarrays for Characterization of Community-Onset Methicillin-Resistant *Staphylococcus aureus*.** *J Clin Microbiol* 2006, **44:**1040-1048.
77. Francois P, Huyghe A, Charbonnier Y, Bento M, Herzig S, Topolski I, Fleury B, Lew D, Vaudaux P, Harbarth S, *et al.*: **Use of an automated multiple-locus, variable-number tandem repeat-based method for rapid and high-throughput genotyping of *Staphylococcus aureus* isolates.** *J Clin Microbiol* 2005, **43:**3346-3355.
78. **Gene Expression Omnibus** [http://www.ncbi.nlm.nih.gov/geo/]
79. **Genomic Research Laboratory** [http://www.genomic.ch/sup4.php]
80. Gill SR, Fouts DE, Archer GL, Mongodin EF, Deboy RT, Ravel J, Paulsen IT, Kolonay JF, Brinkac L, Beanan M, *et al.*: **Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain.** *J Bacteriol* 2005, **187:**2426-2438.

JOURNAL OF CLINICAL MICROBIOLOGY, Mar. 2006, p. 1040–1048 Vol. 44, No. 3
0095-1137/06/$08.00+0 doi:10.1128/JCM.44.3.1040–1048.2006

Use of Oligoarrays for Characterization of Community-Onset Methicillin-Resistant *Staphylococcus aureus*

Thibaud Koessler,[1] Patrice Francois,[1]* Yvan Charbonnier,[1] Antoine Huyghe,[1] Manuela Bento,[1] Sasi Dharan,[3] Gesuele Renzi,[2] Daniel Lew,[1,2] Stephan Harbarth,[3] Didier Pittet,[3] and Jacques Schrenzel[1,2]

Genomic Research Laboratory[1] and Clinical Microbiology Laboratory,[2] Service of Infectious Diseases, and Infection Control Program, Department of Internal Medicine,[3] University of Geneva Hospitals, CH-1211 Geneva 14, Switzerland

Received 28 June 2005/Returned for modification 6 September 2005/Accepted 3 October 2005

Until recently, methicillin-resistant *Staphylococcus aureus* (MRSA) was considered the prototype of a hospital-acquired bacterial pathogen. However, recent reports have shown that MRSA has now emerged in the community. Characterization of specific markers for distinguishing the origin of isolates could contribute to improved knowledge of MRSA epidemiology. The release of whole-genome sequences of hospital- and community-acquired *S. aureus* strains allowed the development of whole-genome content analysis techniques, including microarrays. We developed a microarray composed of 8,191 open reading frame-specific oligonucleotides covering >99% of the four sequenced *S. aureus* genomes (N315, Mu50, MW2, and COL) to evaluate gene contents of hospital- and community-onset *S. aureus* strains. In parallel, pulsed-field gel electrophoresis, variable number of tandem repeats, antibiogram, staphylococcal cassette chromosome-*mec* element typing, and presence of the Panton-Valentine leukocidin gene were evaluated in a collection of 15 clinical isolates. Clusters obtained with microarrays showed a high degree of similarity with those obtained by pulsed-field gel electrophoresis or variable number of tandem repeats. Clusters clearly segregated hospital-onset strains from community-onset strains. Moreover, the microarray approach allowed definition of novel marker genes and chromosomal regions specific for given groups of isolates, thus providing better discrimination and additional information compared to pulsed-field gel electrophoresis and variable number of tandem repeats. Finally, the comparative genome hybridization approach unraveled the occurrence of multiple horizontal transfer events leading to community-onset MRSA as well as the need for a specific genetic background in recipient strains for both the acquisition and the stability of the *mec* element.

Methicillin-resistant *Staphylococcus aureus* (MRSA) is the causative agent of a wide diversity of diseases ranging from benign skin infections to life-threatening diseases such as endocarditis, osteomyelitis, sepsis, and toxic shock syndrome. MRSA was previously described as a typical nosocomial pathogen (2, 4, 30, 37), but recently, several outbreaks in the community have been reported (9, 29). Although exportation of nosocomial MRSA lineages to the community explains some cases, epidemiologic and genetic elements suggest that the community holds specific strains harboring different genetic backgrounds (11, 20, 52).

Specific markers have been identified to distinguish community- from hospital-acquired strains, such as antibiotic susceptibility profiles (antibiograms), toxin contents (toxinograms), and staphylococcal cassette chromosome (SCC)-*mec* typing. Most of these markers are carried by mobile genetic elements or located in genomic islands, suggesting transmission by horizontal transfer (3, 38, 57). Phylogenic trees arising from molecular techniques such as sequencing of highly conserved genes (multilocus sequence typing [MLST]) or from enzymatic cleavage of genomic DNA (pulsed-field gel electrophoresis [PFGE]) demonstrated that community and hospital strain populations are clearly divergent (57), originating from different ancestral clones. These molecular methods contributed to the solution of central phylogenetic issues. MLST revealed a powerful macroevolutionary tool recognizing common ancestries, while PFGE displayed appropriate resolution power for distinguishing outbreak events and providing microevolutionary analysis (6, 36). However, these methods appear poorly informative for the study of specific gene contents in whole bacterial genomes. Identification of specific genes would considerably deepen our knowledge of epidemiologic markers and bacterial mechanisms implicated in *Staphylococcus aureus* dissemination. Clearly, such information is accessible only with a whole-genome approach.

Recent progress in high-throughput sequencing techniques yielded the publication of numerous *S. aureus* genomes, thus facilitating the discovery of sequences that are variably repeated in the different sequenced strains. Those variable numbers of tandem repeats (VNTR) facilitate a new type of multilocus analysis. VNTR was successfully applied to the molecular study of numerous bacteria (40, 45, 46), including MRSA (22, 27, 47).

Genome sequence data also permitted the development of DNA oligoarrays for evaluating the presence and/or expression of genes in whole genomes of several bacterial pathogens (5, 28, 49). The use of microarrays for studying MRSA genomic contents showed that 22% of the genome is dispensable, containing mainly virulence and resistance factors. Moreover, these regions contain mediators of lateral gene transfer such as transposase and integrase genes (21). Recently, Saunders and colleagues used a microarray only composed of virulence-as-

* Corresponding author. Mailing address: University of Geneva Hospitals, Service of Infectious Diseases/Genomic Research Laboratory, CH-1211 Geneva 14, Switzerland. Phone: 41 22 372 9338. Fax: 41 22 372 9830. E-mail: patrice.francois@genomic.ch.

TABLE 1. Antimicrobial susceptibility profiles

Strain No.	Sampling date (day.mo.yr)	Age (yr)	Sex[b]	PVL[c]	SCC-*mec* type	ST	Result for[a]:											
							PEN	OXA	AMI	GEN	NOR	CLI	ERY	FUS	MUP	SXT	TET	FOS
I	19.06.2003	66	M	−	I	228	R	R	R	R	R	R	R	R	S	S	S	S
II	16.05.2003	65	M	−	I	228	R	R	R	R	R	R	R	S	R	S	S	S
III	14.04.2003	74	M	−	III	239	R	R	R	R	R	S	R	S	S	R	R	R
IV	22.05.2003	59	M	−	IV	8	R	R	S	S	R	S	S	S	S	S	S	S
V	10.04.2003	26	F	−	IV	8	R	R	S	S	S	S	S	I	S	S	S	S
VI	08.05.2003	36	M	+	V	152	R	R	R	R	S	S	S	S	S	S	S	S
VII	20.05.2003	37	M	+	V	152	R	R	R	R	S	S	S	S	S	S	S	S
VIII	24.08.2003	29	M	+	V	152	R	R	R	R	S	S	S	S	S	S	S	S
IX	24.07.2003	40	M	+	IV	80	R	R	R	S	S	R	R	R	R	S	S	S
X	24.07.2003	9	F	+	IV	80	R	R	R	S	S	R	R	R	R	S	S	S
XI	26.02.2003	40	M	+	IV	80	R	R	R	S	S	S	S	R	S	S	S	S
XII	26.02.2003	40	M	+	IV	80	R	R	R	S	S	S	S	R	S	S	S	S
XIII	17.02.2003	50	M	+	IV	80	R	R	R	S	S	S	S	R	S	S	S	S
XIV	02.08.2003	39	F	+	IV	NT[d]	R	R	R	S	S	S	S	R	S	S	S	S
XV	07.08.2003	73	M	+	IV	80	R	R	R	S	S	S	S	R	R	S	S	S

[a] Abbreviations: PEN, penicillin; OXA, oxacillin; AMI, amikacin; GEN, gentamicin; NOR, norfloxacin; CLI, clindamycin; ERY, erythromycin; FUS, fusidic acid; MUP, mupirocin; SXT, cotrimoxazole; TET, tetracycline; FOS, fosfomycin; R, resistant; S, susceptible; I, intermediate.

[b] M, male; F, female.

[c] PVL, presence (+) or absence (−) of the Panton-Valentine leukocidin gene.

[d] NT, nontypeable. Allelic profile is 1.3.1.14.11.51.NT, which matches to 6/7 alleles of ST80. The sequence of the seventh gene, corresponding to allele "4," suggested that this strain is probably a new ST.

sociated factors and core genes (used for MLST) to study the evolution and pathogenic potential of *S. aureus* isolates. This array yielded a coherent phylogenetic tree (using core genes) yet provided interesting epidemiological information on the potential evaluation of strain virulence (50).

To further analyze *S. aureus* gene contents and provide a more detailed molecular map, we developed an oligoarray (10) based on the genome of four fully sequenced *S. aureus* strains: MW2 (3), N315 (38), Mu50 (38), and COL (http://www.tigr.org/tdb/mdb/mdbinprogress.html). This array is composed of 8,191 open reading frame (ORF)-specific oligonucleotides allowing >99% coverage of these four genomes. To find molecular signatures specific for hospital-acquired or community-onset strains, we examined the genomic contents of 15 MRSA isolates. Strains are classified as community-onset MRSA (CO-MRSA; 13 strains) or hospital-acquired MRSA (HA-MRSA; 2 strains) according to genetic markers previously described; we also included clinical data (age, sex, sampling date). We compared clusters obtained by hybridizing genomic DNA on microarrays (genomotyping (34), PFGE, and VNTR. The microarray approach provided better discrimination. Furthermore, it allowed definition of novel marker genes and chromosomal regions specific for given groups of isolates, thus enriching our molecular arsenal for epidemiological monitoring.

MATERIALS AND METHODS

Strain collection. CO-MRSA strains were selected from a collection recovered during a screening study performed at hospital admission between February and August 2003 (26) and from ongoing surveillance of MRSA in the Geneva community. Thirteen strains were randomly selected as potential CO-MRSA according to their antimicrobial susceptibility profiles, SCC-*mec* element type, and presence of the Panton-Valentine leukocidin (PVL) gene. Moreover, 2 MRSA isolates (strains I and II) were added as controls, representative of our predominant HA-MRSA clone (Table 1). Note that strains IX through XII and strain XV originated from three members of the same family.

Microbiologic methods. Identification of MRSA was performed on oxacillin-resistant *S. aureus* plates (Oxoid, Basingstoke, United Kingdom). Further identification of MRSA was based on Pastorex agglutination (Bio-Rad, Reinach, Switzerland), DNase reaction on agar, and growth on Mueller-Hinton oxacillin plates (6 mg of oxacillin per ml). MRSA identification was confirmed with the Vitek 2 identification and susceptibility testing cards for gram-positive bacteria (bioMérieux, Marcy l'Etoile, France).

DNA extraction and purification. Genomic DNA (gDNA) was prepared from isolated colonies grown overnight on Mueller-Hinton agar at 37°C. Briefly, 10^9 cells were lysed in 100 μL Tris-EDTA buffer (10 mM Tris–1 mM EDTA, pH 8) containing 50 μg/ml lysostaphin (Ambicin; Applied Microbiology, Tarrytown, NY) for 10 min at 37°C. DNA was then isolated and purified using a DNeasy kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions, including RNase treatment. DNA quantification and protein contaminations were assessed by using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Inc., Rockland, DE).

PVL, *mecA* detection, and SCC-*mec* typing. PCR assays were conducted to evaluate the presence of the *mecA* gene (23), the Panton-Valentine leukocidin gene, and the SCC-*mec* type element (24).

DNA preparation for pulsed-field gel electrophoresis. Epidemiological typing of MRSA isolates was performed by PFGE of chromosomal DNA digested with SmaI (Bio-Rad) using a CHEF MAPPER system (Bio-Rad, Hercules, CA), according to established protocols (8). The banding patterns of the different gels were analyzed with GelCompar software (v. 4.1; Applied Maths, Belgium). Interpretation of the fragment patterns was based on published criteria (7). Patterns differing by one to six DNA fragments were considered subtypes, and those distinguished by seven or more DNA fragments were considered distinct types (14, 53).

Multiple-locus VNTR typing. The VNTR typing assay was performed as previously described (22) but with the addition of the following two primer pairs for assessment of a total of 10 target genes: SAS-F (5′-TTG-GAA-CAT-TCG-AAT-ATA-CAG-AGT) and SAS-R (5′-TCG-ATG-TAC-TGT-CAC-TTA-ATG-ATG); plsR2-F (5′-AAT-TAC-AAC-GCC-TCA-AGC-TG) and plsR2-R (5′-GCA-CCA-TGG-ATG-ATT-ACT-TC).

SCC-*mec* sequencing. Genomic DNA of strains III, VI, VII, and VIII was extracted as previously described (24). The amplification reaction was performed in a PTC 200 Peltier thermal cycler (MJ Research, Inc., Watertown, Mass.) in a 20-μl reaction volume. Amplification conditions and sequencing primers for the cassette chromosome recombinase (*ccr*) genes were selected according to the method of a previously published study (43). DNA sequencing was performed with an ABI Prism 3100 sequencer (Applied Biosystems). Homologies were searched using BLAST (http://www.ncbi.nlm.nih.gov/BLAST/).

Multilocus sequence typing. MLST was performed on all isolates by PCR amplification of internal fragments of seven housekeeping genes by using previously described procedure and primers (17). PCR products were sequenced with an ABI Prism 3100 DNA sequencer (Applied Biosystems, Foster City, CA, USA). Allele numbers were assigned according to the program available from the MLST website (http://www.mlst.net) (Table 1).

TABLE 2. Primer sequences for amplification of 11 genes and 2 regions of interest[a]

Gene or region	Accession no.	Primer	Primer sequence (5′–3′)	Amplicon size or length (bp)	% GC
Genes					
I	MW0329	F	GCATTAACACCAAAACATTTAGCT	1,089	30.1
		R	ATAAATCAGCATGATGCAGAARGT		
II	MW1040	F	ACTACAACTACAATTGCGTCAACA	367	30.7
		R	AAACTAAGTTGACTGCCTTTTGTG		
III	MW1753	F	CCTCAGTAACTGGAATAAATGCTG	648	32.77
		R	TAAAATCTTTGATTTGAGGCGTAA		
IV	MW0305	F	CATGCGAATTATTTCACGATTATT	946	35.38
		R	TGCTAAAATTGCTTCTTYTTGTGT		
V	MW1327	F	CAAGCATTAAACCATTTATTCGTC	929	35.16
		R	ATGTTCAATGACACCWGAAACTCT		
VI	MW0760	F	ACAGTCTTATCTAACGGCGATGTA	646	28.21
		R	CGACATCTAGATGAAATTGTGTTG		
VII	MW0447	F	AACGTTTATTTGAAGAGTCGAATG	359	35.19
		R	CATATCCATTGATAGCGTTTCTCT		
VIII	MW0622	F	GCATGAACTGGATATTTTGGATAT	963	30.64
		R	CAATTTCATTTTGTAATGGGAAAA		
IX	MW2515	F	AGTTAGTGACATAGCACGTGTGAA	414	32.83
		R	GCCATTATTGCTGTATTACTTTCG		
X	MW0105	F	GTTACTTATCGGTTTAGCGGTTTT	1,189	30.1
		R	ATCCCTTTCCATCTTTTCATATTG		
XI	MW1864	F	CCCTCAAAATGATATTTCRCGATA	125	31.03
		R	TGATTTTTAACATCATTTTTGGATG		
Regions					
A	MW1204–MW1211	F	GAAAGAACATTCCCAAATAATGAA	5,987	28.77
		R	TATTTGCCATTTGGTGAAAAATAC		
B	MW2308–MW2313	F	TACATCCAAATACCGCTAAGAAAA	3,957	31.38
		R	TCAAAATGATATGGAAGTTGTTGC		

[a] Theoretical size and GC content of amplicons were determined based on the genome sequence of MW2.

Microarray hybridization and scanning. Our whole-genome *Staphylococcus aureus* DNA microarray was designed and validated as described by Charbonnier et al. (10). Test and reference gDNAs (1 μg) were labeled with cyanine-3 or cyanine-5 dCTP (NEN, Perkin Elmer) using the BioPrime DNA labeling kit (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Unincorporated fluorescent nucleotides were removed using Centrisep columns (Princeton separations, EMP Biotech, Berlin, Germany). Cy-3-labeled gDNAs from the four reference strains used to design the microarray: 0.125 μg from each strain (54) were mixed with 0.5 μg of Cy-5-labeled test gDNA in hybridization buffer (Agilent Technologies, CA), for a total volume of 250 μl. The hybridization mixture was heated to 95°C for 2 min, and then hybridization was performed for 17 h at 60°C with rotation in a dedicated hybridization oven (Robbins Scientific, Sunnyvale, CA). Stringent washings were then performed according to the manufacturer's instructions. Slides were dried under nitrogen flow and scanned (Agilent Technologies, CA) using 100% photon multiplier tube power for both wavelengths using the Agilent scanner.

Microarray analysis. Fluorescence intensities were extracted using Feature extraction software (version 6.1.1; Agilent). Local background-subtracted signals were corrected for unequal dye incorporation or unequal load of the labeled product. The algorithm consisted of a rank consistency filter and a curve fit using the default LOWESS (locally weighted linear regression) method. Additional software was developed in-house to analyze the processed data. This software filtered the data to exclude irrelevant values, as flagged by the extraction software. The background noise of each experiment was evaluated by computing the standard deviation of negative-control intensities. Features whose intensities were smaller than the standard deviation value of the negative controls were considered inefficient hybridization and discarded from further analysis. The software calculated for each spot the logarithm of the ratio between the test channel and the control channel (log ratio). Since the control signal is present in each spot, this log ratio corresponds to a per feature normalization. Computed log ratio values were further sorted into 150 bin categories and fitted with a Gaussian distribution curve, using the Levenberg-Marquardt algorithm. The software estimated the presence probability of each oligonucleotide probe (EPP), as previously described (34). As clearly documented in the work of Kim et al., we used the most stringent EPP value, as our study was focused on "strict divergent gene analysis parameter." EPP values of ≤1% (each oligonucleotide probe) were extracted and considered absent features in the test channel; we concentrated on the subset of probes predicted to reliably detect MW2 gene targets (10). The list of absent features from each experiment was then clustered by the software using Dice distance and the group average linkage algorithm to construct a hierarchical cluster tree (15). The extensive list of these genes, absent in at least one of the tested strains is shown in a supplemental table (http://www.genomic.ch/sup3.php).

Verification of divergent genes. Projection of absent features onto the genome map of MW2 (Genome Viewer Software) (33) identified regions of difference between our collection and MW2. Flanking primers were designed with Jellyfish software (LabVelocity) to control the size of the amplicons spanning these regions of difference.

We also designed primers to assess a selection of 11 cluster-specific genes. As each gene is covered by one to eight oligonucleotide probes on our microarray, selected gene targets ought to have every covering oligonucleotide probe present (or absent) to be selected. Target genes and regions within different clusters were amplified using 13 PCRs after protocol optimization using MW2 genomic DNA as a target control (primer sequences are shown in Table 2). PCR assay was performed in a PTC 200 Peltier thermal cycler in a 20-μl reaction volume. Reaction mixtures contained 0.2 mM concentrations of each deoxynucleoside triphosphate, 0.3 μM concentrations of each primer, 1 mM $MgSO_4$, 0.8 U of KOD hot-start DNA polymerase (Novagen, Madison, Wis.). Cycling conditions were as follows: denaturation for 2 min at 94°C; 35 cycles of 15 s at 94°C, 20 s at 60°C, and *x* s (20 s/kbp, depending on the size of genes and regions) at 72°C; and a postextension of 10 min at 72°C. PCR amplification results were evaluated using the Bioanalyzer 2100 with the DNA 7500 chip kit (Agilent Technologies, CA).

RESULTS

Strain selection and characterization. Among 13 strains defined as potential CO-MRSA by their antimicrobial susceptibility profiles, only 10 (77%) possessed the gene of the PVL

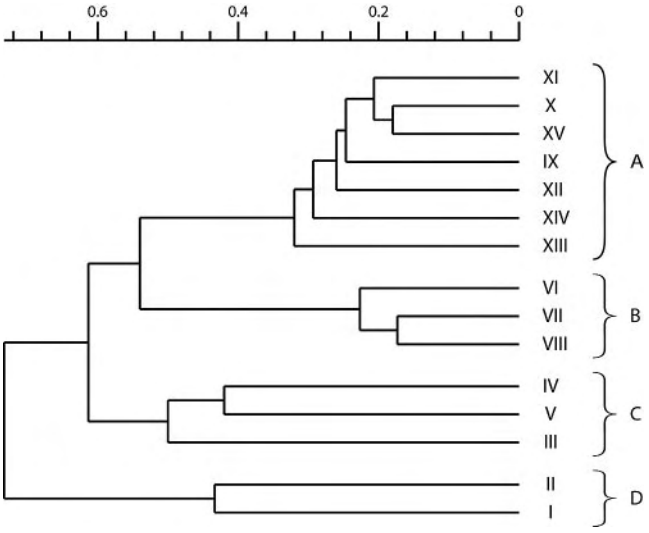


FIG. 1. Cluster analysis by microarrays, using an EPP value of ≤1%. Roman numerals represent strain numbers. The scale above the dendrogram shows the percentage of similarity among different strains.

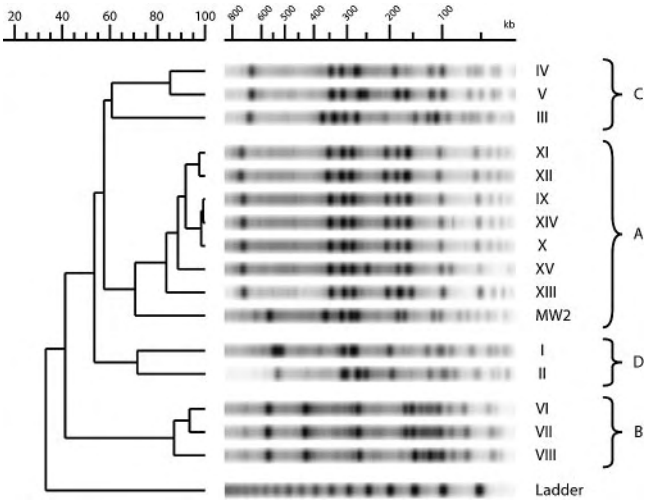


FIG. 2. Cluster analysis by PFGE. The scale above the dendrogram shows the percentage of similarity among different strains.

toxin, contrasting with previous reports suggesting that it was an efficient marker of CO-MRSA (57); however, further study confirmed our local prevalence (39, 42). The two other selected strains were negative for the PVL toxin yet displayed a highly resistant antibiotic profile (Table 1). The type of SCC-*mec* element has also been reported as an important marker for distinguishing community acquisition from hospital acquisition (22, 48). Indeed, among our PVL-positive strains, 7/10 carried a type IV cassette and 3/10 carried a type V cassette. Among the remaining 3 strains, two additional were type IV (15%) and one revealed type III (7%). Antibiotic susceptibility profiles (Table 1) showed that the two nosocomial strains displayed a broad resistance spectrum (strains I and II), as opposed to the CO-MRSA, which showed more susceptible and variable resistance patterns.

Genomotyping and hierarchical clustering by microarray. The listing of gene targets reported to be absent by genomotyping was employed for cluster analysis. We arbitrarily defined a cutoff of 52% that segregated our strain collection into four clusters (Fig. 1). Cluster A was composed of seven strains containing the PVL toxin and a SCC-*mec* IV element. These isolates were resistant to penicillin and oxacillin but susceptible to the majority of antibiotics tested (Table 1). We found in this cluster all isolates collected from patients of the same family (strains IX, XI, and XII were isolates from the father, XV was an isolate from the grandfather, and X was from the daughter). This cluster contained also two isolates (strains XIII and XIV) epidemiologically unrelated to the described family outbreak. We determined a similarity of >65% between strains that composed this cluster and of >70% between isolates of the same family (included in cluster A). Cluster B was composed of PVL-positive strains with a SCC-*mec* V element but was epidemiologically unrelated to cluster A. As shown in Table 1, these strains were resistant to gentamicin and susceptible to fusidic acid, as opposed to strains from cluster A. In this cluster, we noted a >75% similarity to and a <55% difference from cluster A. Cluster C was composed of three PVL-negative strains with variable antibiograms, two of which possessed a SCC-*mec* IV cassette and the remaining one possessed a type III cassette. Finally, cluster D contained the 2 HA-MRSA control strains, displaying a SCC-*mec* I element. This final cluster showed the least similarity (only about 30%) to the other ones.

Pulsed-field gel electrophoresis. Figure 2 shows the 16 PFGE lanes that segregated into four major clusters. These four clusters showed the same strain composition as those previously assessed by microarray analysis. Cluster A included all five strains from the family outbreak, two unrelated isolates (strains XIII and XIV), and strain MW2, considered a model of community-acquired MRSA (3). Note that strains IX, X, and XIV appeared clonally related by PFGE. Cluster C and cluster D were composed of *PVL*-negative CO-MRSA strains and control HA-MRSA, respectively. The most distantly related cluster, cluster B, included three *PVL*-positive strains with a SCC-*mec* V cassette.

Multiple-locus VNTR typing. VNTR results were analyzed with in-house software, as described elsewhere (22). Again, cluster composition was similar to that previously determined by microarray analysis or PFGE (Fig. 3). A high degree of similarity was observed within cluster A, except for strain XIII, which displayed a 10% divergence from the other strains composing cluster A. This represents the lowest discriminatory power among the three methods describing this cluster. As for PFGE, clusters C and D were closest to cluster A. Strains constituting cluster D appeared to be clonally related by VNTR. MW2, which coclustered with cluster A by PFGE, appeared very distantly related using VNTR. These data and clustering analysis were confirmed by another independent determination (not shown).

MLST analysis. All 15 MRSA isolates were further analyzed by MLST. Strains I and II belong to ST228. Strains III and IV and strain V from cluster C revealed ST239 and ST8, respectively, both being major clones belonging to CC8. Strains VI, VII, and VIII from cluster B belong to ST152, while six of

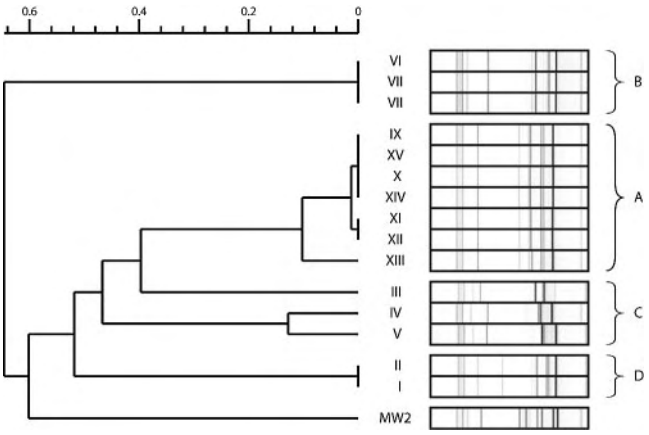


FIG. 3. Cluster analysis by VNTR. Roman numerals represent strain numbers. The scale above the dendrogram shows the percentage of similarity among different strains.

seven strains from cluster A belong to ST80. The last one, strain XIV, was nontypeable (Table 1).

Circular genome view. The previous identification of four different clusters was used to identify and map oligonucleotide probes into cluster-specific groups. A cluster-specific oligonucleotide probe is defined as failing to detect its target gene in all strains of a given cluster. Analyses were performed with an EPP of ≤1%, corresponding to the most stringent available condition. Of 6,201 MW2-specific oligonucleotide probes covering 2,632 ORFs, analysis revealed that 93 (3.5%) to 520 (19.7%) probes failed to detect their cognate targets. Fifty-seven to 335 of these probes were cluster specific, i.e., absent in all strains that compose a given cluster. Surprisingly, strains of the B cluster show the largest number of divergence from MW2 (335 absent oligonucleotide probes). Cluster-specific oligonucleotide probes were displayed on the genome of MW2 (Fig. 4), thus revealing that genome regions appeared discriminatory between different clusters. To confirm this observation, cluster-specific regions and selected genes were further analyzed by PCR amplification.

Validation of microarray results. We analyzed by PCR the presence of 11 genes showing a divergent presence between the four observed clusters. To select a gene, we required that all probes covering that gene revealed its absence in all strains of a given cluster. We also studied two chromosomal regions consisting of more than three contiguous ORFs that were absent in one or more clusters compared with the community-acquired reference strain MW2. Those targets appeared widely distributed on a full-genome scale (Fig. 4). Among our selection, genes I, II, III, and IV were present in every cluster except cluster B. They encode a hypothetical protein similar to low-temperature-requirement A protein, a fibrinogen-binding protein, a serine protease, and a hypothetical protein similar to the dihydroflavone-4-reductase, respectively (corresponding to the following ORF numbers on the genome sequence of MW2: MW0329, MW1040, MW1753, and MW0305). Gene X (MW0105), encoding a capsular polysaccharide synthesis protein, was expected to be absent in all the strains composing cluster B. However, an amplification band was observed in all strains composing this cluster (figure not shown), suggesting a too-stringent selection of the cutoff. Two genes were selected to characterize cluster C: a hypothetical protein similar to the two-component sensor histidine kinase (gene VIII, MW0622) and a hypothetical protein (gene IX, MW2515). PCR confirmed their absence in strains from cluster C only. Gene V (MW1327) was absent in cluster D only, as determined by microarray and PCR; it encodes for a threonine deaminase homolog. Gene VI (extracellular enterotoxine L, MW0760) was absent in all the strains excepted in MW2 and strain XIII. Gene XI (MW1864) encodes for a truncated transposase present in all strains and was thus considered a positive control for genomotyping. The last gene (gene VII, MW0447) was absent from strains of cluster A; it encodes a conserved hypothetical protein.

Two regions of difference were selected. As expected, regions A and B displayed 28.77 and 31.38% GC content, respectively, a lower value than the average 32.80% GC content measured in the MW2 genome (3). Primers were designed flanking the zone of interest to generate amplicons of different sizes. Jellyfish software (LabVelocity) was used for primer selection and amplicon size determination. As PCR primers were designed in conserved regions, flanking the region of divergence, we were able to calculate the theoretical size of the PCR product. Region A (MW1206 to MW1209) is composed of ABC transporters (MW1206 to MW1207) and a two-component histidine kinase sensor and regulator (MW1208 to MW1209). Based on the MW2 genome, amplification of region A (Fig. 5a) was expected to yield a 5,987-bp fragment, in agreement with our MW2 control strain that yielded a 5,585-bp amplicon. Strains constituting cluster A yielded an amplicon size of 2,761 bp (minimum, 2,600 bp; maximum, 2,944 bp) (Fig. 5A), in agreement with the predicted size of that conserved region. However, no amplification signal was recorded from strains of cluster B. In cluster C, strains IV and V displayed a shorter band than MW2, but their coclustering strain III yielded a band of 6,333 bp, with a probable insertion of approximately 300 bp.

Region B (ORFs MW2308 to MW2313) contains two hypothetical proteins and a transcriptional regulator. This second region (Fig. 5B) generated a 3,893-bp band when using strain MW2, in agreement with the predicted size of 3,957 bp. Cluster A yielded an average size of 1,929 bp (minimum, 1,821; maximum, 2,028), in agreement with the predicted conserved region. In cluster B, only strain VI yielded a 4,564-bp amplicon, thus showing a probable insertion in this region. On average, amplicons measured 4,138 bp and 4,163 bp when originating from clusters C and D, respectively. Taken together, these results suggest that our conservative genomotyping approach allowed accurate detection of two regions of difference.

DISCUSSION

Numerous recent reports about community-onset MRSA contributed to convince experts that its epidemiology represents an emerging and worldwide concern (57). The study of evolutionary relationship of MRSA in the community is controversial; while some studies report a clonal origin for CO-MRSA, others suggest a more distant relatedness between clinical isolates (16, 41). We describe here the characterization and analysis of a collection of CO-MRSA strains using whole-

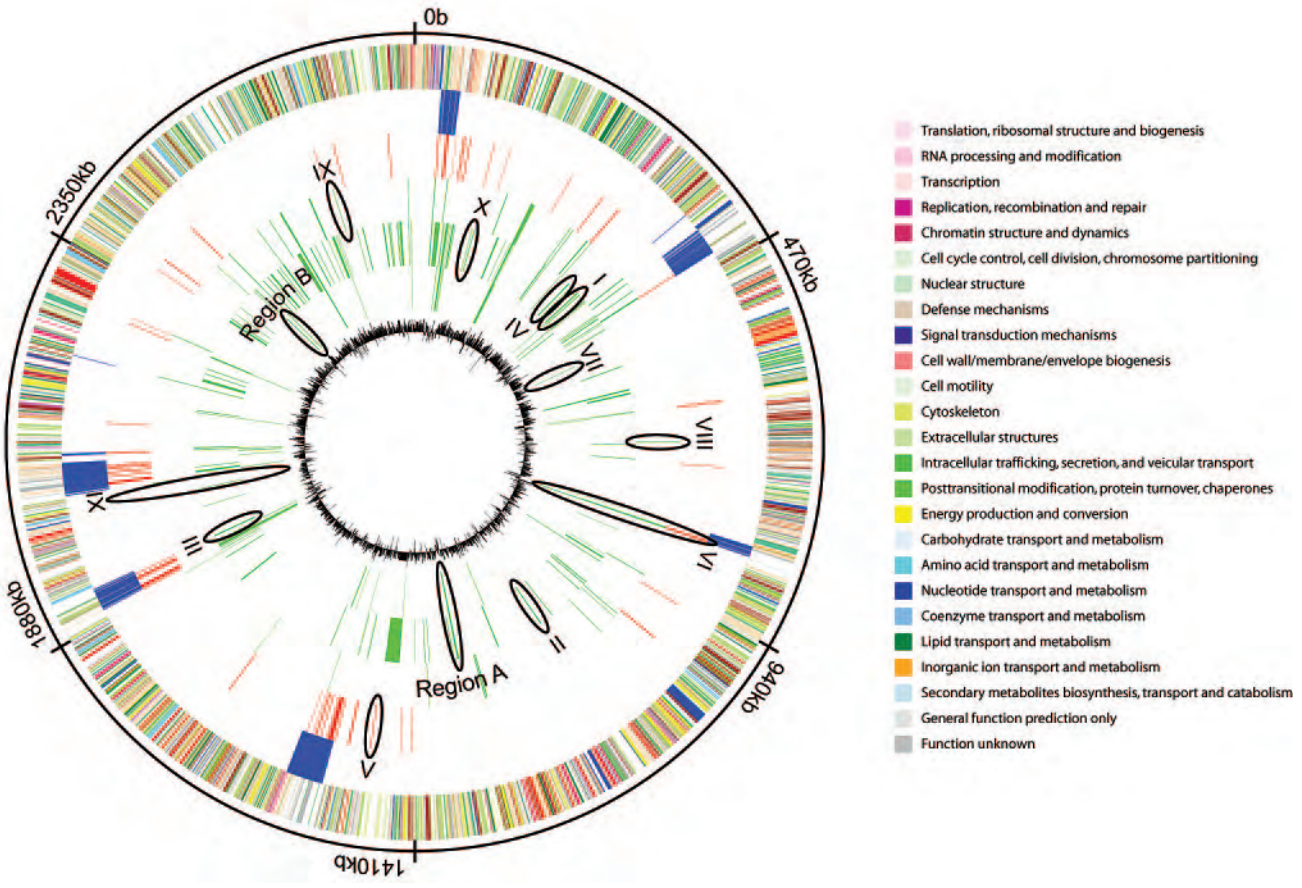


FIG. 4. Functional genomic organization of the MW2 chromosome compared with 15 clinical strains. From the outside inward, the first circle represents nucleotide position in Mbp. The second circle shows ORFs on the plus and minus (37) strands, with colors according to their COG functional categories. The third circle shows genetic mobile elements and virulence factors from MW2 (3). The fourth circle highlights gene targets that are absent from cluster D compared to MW2. The fifth, sixth, and seventh circles depict gene targets that are absent in clusters C, B, and A, respectively. The eighth circle represents G/C content of MW2. Roman numerals represent gene numbers that are further analyzed.

genome oligonucleotide microarrays together with different standardized genotyping methods.

The combined use of different genotyping methods applied on the same panel of strains yields an improvement in overall resolving power. Frequently, laboratories associate two genotyping techniques, one focusing on slow evolutionary genetic markers and another one addressing rapid evolutionary elements. MLST is considered the reference method for studying relatedness between strains (18), while PFGE is recognized as the gold standard method for outbreak analysis (32, 36). However, PFGE can display an overly discriminant power by segregating clonally related strains (6); and under special analysis criteria, PFGE and MLST have the same level of discrimination (44). Coombs et al. (13) have proved the efficiency of a combined approach using MLST and analysis of the *mec* element; moreover, the latter provided substantial complementary epidemiological information. One should also keep in mind that MLST can group unrelated strains under a common profile (55, 58), thus revealing insufficient discriminatory power. From our study, discriminatory power obtained by VNTR appears lower than that of PFGE and comparative genome hybridization techniques. However, strain relatedness, as established with cluster composition and arrangement, was identical using VNTR (Fig. 3) and PFGE approaches (Fig. 2). The main advantages of VNTR reside in its low cost, moderate time consumption, and high typing resolution (23). Cluster analysis by microarrays is clearly the most discriminative of the three methods. The cluster composition is identical to the two previously cited methods, although it shows a slightly different arrangement; this observation was also described by Koreen et al. (36). Moreover, the comparative genome hybridization technique, consisting of mapping the whole chromosome (at a frequency of 1 probe every 400 to 500 bp using our array), allows precise evaluation of variation between clusters and even between strains within each cluster. This global overview of the genomic composition highlights the biological signature of different clusters.

MLST studies revealed that the main clones were ST5, ST8, and ST239. Among them, some specific genetic backgrounds appear able to acquire SCC-*mec* elements I, II, III, and IV, while others appear capable of displaying only one of them (18, 19). This observation strongly suggests that the acquisition of the *mec* element occurred during several independent acquisition events, i.e., the horizontal transfer of genetic material

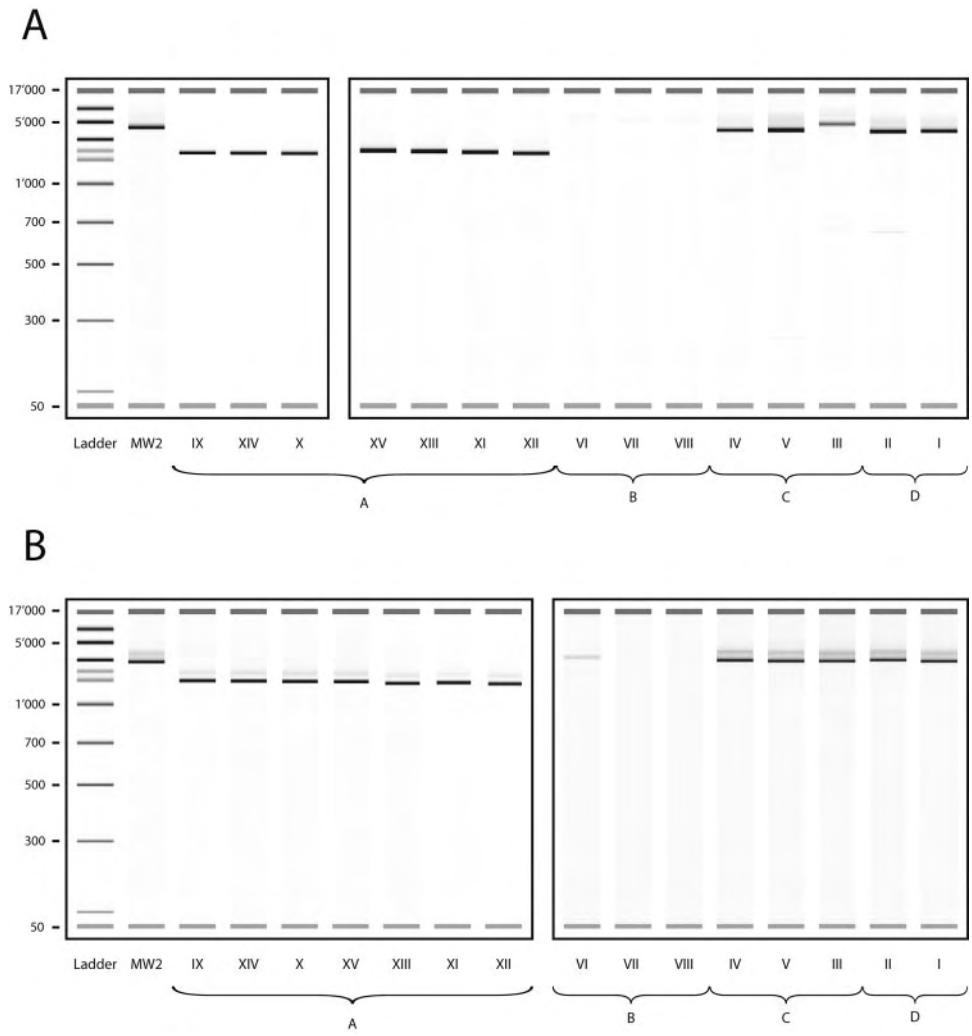


FIG. 5. Size variation of amplicons spanning regions A (A) and B (B). Roman numerals represent strain numbers. Each box represents a new assay.

between different strains (52). For instance, the SCC-*mec* type IV (associated with the presence of the *PVL* gene and considered a specific marker of community-onset strains) is found integrated in a wide diversity of genetic backgrounds and in different clones, including ST5 and ST8 (19). On the contrary, our study and a previous one performed in Switzerland (25) suggest that the new SCC-*mec* type V cassette is recovered only in a particular genetic background, namely ST152. An Australian study was recently able to link the presence of the *mec* V element to three different sequence types: ST152, ST8, and ST45 (13). By combining VNTR, PFGE, and microarrays, we confirmed that SCC-*mec* V-related strains appear extremely homogeneous and distantly related to all other strains, either CO-MRSA or HA-MRSA. This may signify that the *mec* V mobile element needs a particular genetic background to be acquired and stabilized.

In our study, CO-MRSA strains were found to belong to major epidemic clones: strain IV and V *mec* IV belong to ST8 (epidemic MRSA strains EMRSA-2 and EMRSA-6), while strain III, carrier of the *mec* III element, belongs to ST239. As we do not differentiate *mec* III from *mec* IIIA, we were unable to attribute this strain to the Brazilian clone (*mec* IIIA) (56) or to the Hungarian clone (*mec* III) (35, 51). All but one strain from cluster A belonged to ST80, a preponderant clone in Greece and also carrier of the Panton-Valentine toxin (1).

We were unable to attribute a sequence type to strain XIV, as sequencing results yielded a novel allelic profile. Interestingly, this profile differed from that of ST80 by one nucleotide only.

Taken together, these observations within that small strain collection suggest (i) the occurrence of multiple horizontal transfer events leading to CO-MRSA and (ii) the presence of specific genetic backgrounds for the acquisition and stability of the *mec* element (31).

An advantage of typing the whole bacterial genome is the possibility of detecting regions of difference between clusters of strains. In this study, we were able to identify different types of CO-MRSA strains: MRSA strains with the SCC-*mec* V element, CO-MRSA related to the reference strain MW2 (*mec* IV type), and other CO-MRSA strains (*mec* IV type or *mec* III type). Using our genomotyping analysis, our strain collection

differed from MW2 by 3.5 to 19.7%. Using microarrays, Fitzgerald et al. evaluated the dispensable part of the genome of *Staphylococcus aureus* as 22% (21). Discrepancies can be explained by the use of different analytical methods and diversity of strain collections. For instance, we used a very conservative EPP value, whereas other authors reported the absence of genes with EPP values as high as 19% (12). Thus, we expect that our approach might have missed genes that were truly absent in different strains. Whereas Fitzgerald et al. (21) compared MRSA and methicillin-susceptible *S. aureus* strains of human, ovine, and bovine origin to the lab strain COL, we focused on the CO-MRSA isolated recovered in our region and compared them with the human pathogen MW2. As expected, regions of difference are mainly localized in genomic regions exhibiting low GC contents and thus likely to derive from horizontal transfer (21).

Genomotyping enables a precise overview of the genetic contents of a strain during a single experiment. Such microarray experiments highlight nonessential regions specific to some strains and permit definition of clusters based on such biological signatures. Thus, they permit not only typing and characterization of clinical isolates with high resolution capacity but also identification of genetic markers characterizing the origin of the strains. In this context, microarrays appear to be a powerful tool for comprehensive genotyping and further understanding of genetic plasticity and strain dissemination.

ACKNOWLEDGMENTS

We thank the many staff members of the University Hospitals of Geneva in Bacteriology, Infectious Diseases, and Infection Control Program who contributed to this study.

This work was supported by grants 631-057950.99, PP00B—103002/1 (to J.S.), and 404940-106296/1 (to P.F.) from the Swiss National Science Foundation. T.K. was supported by a grant from the Department of Internal Medicine from the University Hospitals of Geneva.

REFERENCES

1. **Aires, D. S., C. Bartzavali, I. Spiliopoulou, I. S. Sanches, M. I. Crisostomo, and H. de Lencastre.** 2003. Two international methicillin-resistant *Staphylococcus aureus* clones endemic in a university hospital in Patras, Greece. J. Clin. Microbiol. **41:**2027–2032.
2. **Archer, G. L.** 1998. *Staphylococcus aureus*: a well-armed pathogen. Clin. Infect. Dis. **26:**1179–1181.
3. **Baba, T., F. Takeuchi, M. Kuroda, H. Yuzawa, K. Aoki, A. Oguchi, Y. Nagai, N. Iwama, K. Asano, T. Naimi, H. Kuroda, L. Cui, K. Yamamoto, and K. Hiramatsu.** 2002. Genome and virulence determinants of high virulence community-acquired MRSA. Lancet **359:**1819–1827.
4. **Barrett, F. F., R. F. McGehee, Jr., and M. Finland.** 1968. Methicillin-resistant *Staphylococcus aureus* at Boston City Hospital. Bacteriologic and epidemiologic observations. N. Engl. J. Med. **279:**441–448.
5. **Bischoff, M., P. Dunman, J. Kormanec, D. Macapagal, E. Murphy, W. Mounts, B. Berger-Bachi, and S. Projan.** 2004. Microarray-based analysis of the *Staphylococcus aureus sigmaB* regulon. J. Bacteriol. **186:**4085–4099.
6. **Blanc, D. S., P. Francioli, and P. M. Hauser.** 2002. Poor value of pulsed-field gel electrophoresis to investigate long-term scale epidemiology of methicillin-resistant *Staphylococcus aureus*. Infect. Genet. Evol. **2:**145–148.
7. **Blanc, D. S., D. Pittet, C. Ruef, A. F. Widmer, K. Muhlemann, C. Petignat, S. Harbarth, R. Auckenthaler, J. Bille, R. Frei, R. Zbinden, P. Moreillon, P. Sudre, and P. Francioli.** 2002. Molecular epidemiology of predominant clones and sporadic strains of methicillin resistant *Staphylococcus aureus* in Switzerland and comparison with European epidemic clones. Clin. Microbiol. Infect. **8:**419–426.
8. **Blanc, D. S., M. J. Struelens, A. Deplano, R. de Ryck, P. M. Hauser, C. Petignat, and P. Francioli.** 2001. Epidemiological validation of pulsed-field gel electrophoresis patterns for methicillin-resistant *Staphylococcus aureus*. J. Clin. Microbiol. **39:**3442–3445.
9. **Campbell, K. M., A. F. Vaughn, K. L. Russell, B. Smith, D. L. Jimenez, C. P. Barrozo, J. R. Minarcik, N. F. Crum, and M. A. Ryan.** 2004. Risk factors for community-associated methicillin-resistant *Staphylococcus aureus* infections in an outbreak of disease among military trainees in San Diego, California, in 2002. J. Clin. Microbiol. **42:**4050–4053.
10. **Charbonnier, Y., B. M. Gettler, P. Francois, M. Bento, A. Renzoni, P. Vaudaux, W. Schlegel, and J. Schrenzel.** 2005. A generic approach for the design of whole-genome oligoarrays, validated for genomotyping, deletion mapping and gene expression analysis on *Staphylococcus aureus*. BMC Genomics **6:**95.
11. **Charlebois, E. D., F. Perdreau-Remington, B. Kreiswirth, D. R. Bangsberg, D. Ciccarone, B. A. Diep, V. L. Ng, K. Chansky, and H. F. Chambers.** 2004. Origins of community strains of methicillin-resistant *Staphylococcus aureus*. Clin. Infect. Dis. **39:**47–54.
12. **Chen, T., Y. Hosogi, K. Nishikawa, K. Abbey, R. D. Fleischmann, J. Walling, and M. J. Duncan.** 2004. Comparative whole-genome analysis of virulent and avirulent strains of *Porphyromonas gingivalis*. J. Bacteriol. **186:**5473–5479.
13. **Coombs, G. W., G. R. Nimmo, J. M. Bell, F. Huygens, F. G. O'Brien, M. J. Malkowski, J. C. Pearson, A. J. Stephens, and P. M. Giffard.** 2004. Genetic diversity among community methicillin-resistant *Staphylococcus aureus* strains causing outpatient infections in Australia. J. Clin. Microbiol. **42:**4735–4743.
14. **Deplano, A., W. Witte, W. J. van Leeuwen, Y. Brun, and M. J. Struelens.** 2000. Clonal dissemination of epidemic methicillin-resistant *Staphylococcus aureus* in Belgium and neighboring countries. Clin. Microbiol. Infect. **6:**239–245.
15. **Dice, L. R.** 1945. Measures of the amount of ecologic association between species. Ecology **26:**297–302.
16. **Dufour, P., Y. Gillet, M. Bes, G. Lina, F. Vandenesch, D. Floret, J. Etienne, and H. Richet.** 2002. Community-acquired methicillin-resistant *Staphylococcus aureus* infections in France: emergence of a single clone that produces Panton-Valentine leukocidin. Clin. Infect. Dis. **35:**819–824.
17. **Enright, M. C., N. P. Day, C. E. Davies, S. J. Peacock, and B. G. Spratt.** 2000. Multilocus sequence typing for characterization of methicillin- resistant and methicillin-susceptible clones of *Staphylococcus aureus*. J. Clin. Microbiol. **38:**1008–1015.
18. **Enright, M. C., D. A. Robinson, G. Randle, E. J. Feil, H. Grundmann, and B. G. Spratt.** 2002. The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). Proc. Natl. Acad. Sci. USA **99:**7687–7692.
19. **Feil, E. J., and M. C. Enright.** 2004. Analyses of clonality and the evolution of bacterial pathogens. Curr. Opin. Microbiol. **7:**308–313.
20. **Fey, P. D., B. Said-Salim, M. E. Rupp, S. H. Hinrichs, D. J. Boxrud, C. C. Davis, B. N. Kreiswirth, and P. M. Schlievert.** 2003. Comparative molecular analysis of community- or hospital-acquired methicillin-resistant *Staphylococcus aureus*. Antimicrob. Agents Chemother. **47:**196–203.
21. **Fitzgerald, J. R., D. E. Sturdevant, S. M. Mackie, S. R. Gill, and J. M. Musser.** 2001. Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. Proc. Natl. Acad. Sci. USA **98:**8821–8826.
22. **Francois, P., A. Huyghe, Y. Charbonnier, M. Bento, S. Herzig, I. Topolski, B. Fleury, D. Lew, P. Vaudaux, S. Harbarth, W. Van Leeuwen, A. Van Belkum, D. S. Blanc, D. Pittet, and J. Schrenzel.** 2005. Use of an automated multiple-locus, variable-number tandem repeat-based method for rapid and high-throughput genotyping of *Staphylococcus aureus* isolates. J. Clin. Microbiol. **43:**3346–3355.
23. **Francois, P., D. Pittet, M. Bento, B. Pepey, P. Vaudaux, D. Lew, and J. Schrenzel.** 2003. Rapid detection of methicillin-resistant *Staphylococcus aureus* directly from sterile or nonsterile clinical samples by a new molecular assay. J. Clin. Microbiol. **41:**254–260.
24. **Francois, P., G. Renzi, D. Pittet, M. Bento, D. Lew, S. Harbarth, P. Vaudaux, and J. Schrenzel.** 2004. A novel multiplex real-time PCR assay for rapid typing of major staphylococcal cassette chromosome *mec* elements. J. Clin. Microbiol. **42:**3309–3312.
25. **Harbarth, S., P. Francois, J. Schrenzel, C. Fankhauser-Rodriguez, S. Hugonnet, T. Koessler, A. Huyghe, and D. Pittet.** 2005. Community-associated methicillin-resistant *Staphylococcus aureus*, Geneva, Switzerland. Emerg. Infect. Dis. **11:**962–965.
26. **Harbarth, S., H. Sax, C. Fankhauser-Rodriguez, J. Schrenzel, A. Agostinho, and D. Pittet.** Evaluating the probability of previously unknown methicillin-resistant *Staphylococcus aureus* carriage at hospital admission. Am. J. Med., in press.
27. **Hardy, K. J., D. W. Ussery, B. A. Oppenheim, and P. M. Hawkey.** 2004. Distribution and characterization of staphylococcal interspersed repeat units (SIRUs) and potential use for strain differentiation. Microbiology **150:**4045–4052.
28. **Hinchliffe, S. J., K. E. Isherwood, R. A. Stabler, M. B. Prentice, A. Rakin, R. A. Nichols, P. C. Oyston, J. Hinds, R. W. Titball, and B. W. Wren.** 2003. Application of DNA microarrays to study the evolutionary genomics of *Yersinia pestis* and *Yersinia pseudotuberculosis*. Genome Res. **13:**2018–2029.
29. **Jones, M. E., D. C. Mayfield, C. Thornsberry, J. A. Karlowsky, D. F. Sahm, and D. Peterson.** 2002. Prevalence of oxacillin resistance in *Staphylococcus aureus* among inpatients and outpatients in the United States during 2000. Antimicrob. Agents Chemother. **46:**3104–3105.
30. **Jorgensen, J. H.** 1986. Laboratory and epidemiologic experience with methicillin-resistant *Staphylococcus aureus* in the USA. Eur. J. Clin. Microbiol. **5:**693–696.

31. **Katayama, Y., D. A. Robinson, M. C. Enright, and H. F. Chambers.** 2005. Genetic background affects stability of *mecA* in *Staphylococcus aureus*. J. Clin. Microbiol. **43:**2380–2383.
32. **Kazakova, S. V., J. C. Hageman, M. Matava, A. Srinivasan, L. Phelan, B. Garfinkel, T. Boo, S. McAllister, J. Anderson, B. Jensen, D. Dodson, D. Lonsway, L. K. McDougal, M. Arduino, V. J. Fraser, G. Killgore, F. C. Tenover, S. Cody, and D. B. Jernigan.** 2005. A clone of methicillin-resistant *Staphylococcus aureus* among professional football players. N. Engl. J. Med. **352:**468–475.
33. **Kerkhoven, R., F. H. van Enckevort, J. Boekhorst, D. Molenaar, and R. J. Siezen.** 2004. Visualization for genomics: the Microbial Genome Viewer. Bioinformatics **20:**1812–1814.
34. **Kim, C. C., E. A. Joyce, K. Chan, and S. Falkow.** 2002. Improved analytical methods for microarray-based genome-composition analysis. Genome Biol. **3:**RESEARCH0065.1–0065.17.
35. **Ko, K. S., J. Y. Lee, J. Y. Suh, W. S. Oh, K. R. Peck, N. Y. Lee, and J. H. Song.** 2005. Distribution of major genotypes among methicillin-resistant *Staphylococcus aureus* clones in Asian countries. J. Clin. Microbiol. **43:**421–426.
36. **Koreen, L., S. V. Ramaswamy, E. A. Graviss, S. Naidich, J. M. Musser, and B. N. Kreiswirth.** 2004. *spa* typing method for discriminating among *Staphylococcus aureus* isolates: implications for use of a single marker to detect genetic micro- and macrovariation. J. Clin. Microbiol. **42:**792–799.
37. **Kreiswirth, B., J. Kornblum, R. D. Arbeit, W. Eisner, J. N. Maslow, A. McGeer, D. E. Low, and R. P. Novick.** 1993. Evidence for a clonal origin of methicillin resistance in *Staphylococcus aureus*. Science **259:**227–230.
38. **Kuroda, M., T. Ohta, I. Uchiyama, T. Baba, H. Yuzawa, I. Kobayashi, L. Cui, A. Oguchi, K. Aoki, Y. Nagai, J. Lian, T. Ito, M. Kanamori, H. Matsumaru, A. Maruyama, H. Murakami, A. Hosoyama, Y. Mizutani-Ui, N. K. Takahashi, T. Sawano, R. Inoue, C. Kaito, K. Sekimizu, H. Hirakawa, S. Kuhara, S. Goto, J. Yabuzaki, M. Kanehisa, A. Yamashita, K. Oshima, K. Furuya, C. Yoshino, T. Shiba, M. Hattori, N. Ogasawara, H. Hayashi, and K. Hiramatsu.** 2001. Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. Lancet **357:**1225–1240.
39. **Liassine, N., R. Auckenthaler, M. C. Descombes, M. Bes, F. Vandenesch, and J. Etienne.** 2004. Community-acquired methicillin-resistant *Staphylococcus aureus* isolated in Switzerland contains the Panton-Valentine leukocidin or exfoliative toxin genes. J. Clin. Microbiol. **42:**825–828.
40. **Lindstedt, B. A., T. Vardund, and G. Kapperud.** 2004. Multiple-locus variable-number tandem-repeats analysis of *Escherichia coli* O157 using PCR multiplexing and multi-colored capillary electrophoresis. J. Microbiol. Methods **58:**213–222.
41. **Naimi, T. S., K. H. LeDell, D. J. Boxrud, A. V. Groom, C. D. Steward, S. K. Johnson, J. M. Besser, C. O'Boyle, R. N. Danila, J. E. Cheek, M. T. Osterholm, K. A. Moore, and K. E. Smith.** 2001. Epidemiology and clonality of community-acquired methicillin-resistant *Staphylococcus aureus* in Minnesota, 1996–1998. Clin. Infect. Dis. **33:**990–996.
42. **Naimi, T. S., K. H. LeDell, K. Como-Sabetti, S. M. Borchardt, D. J. Boxrud, J. Etienne, S. K. Johnson, F. Vandenesch, S. Fridkin, C. O'Boyle, R. N. Danila, and R. Lynfield.** 2003. Comparison of community- and health care-associated methicillin-resistant *Staphylococcus aureus* infection. JAMA **290:**2976–2984.
43. **Okuma, K., K. Iwakawa, J. D. Turnidge, W. B. Grubb, J. M. Bell, F. G. O'Brien, G. W. Coombs, J. W. Pearman, F. C. Tenover, M. Kapi, C. Tiensasitorn, T. Ito, and K. Hiramatsu.** 2002. Dissemination of new methicillin-resistant *Staphylococcus aureus* clones in the community. J. Clin. Microbiol. **40:**4289–4294.
44. **Peacock, S. J., G. D. de Silva, A. Justice, A. Cowland, C. E. Moore, C. G. Winearls, and N. P. Day.** 2002. Comparison of multilocus sequence typing and pulsed-field gel electrophoresis as tools for typing *Staphylococcus aureus* isolates in a microepidemiological setting. J. Clin. Microbiol. **40:**3764–3770.
45. **Pourcel, C., F. Andre-Mazeaud, H. Neubauer, F. Ramisse, and G. Vergnaud.** 2004. Tandem repeats analysis for the high resolution phylogenetic analysis of *Yersinia pestis*. BMC Microbiol. **4:**22.
46. **Ramisse, V., P. Houssu, E. Hernandez, F. Denoeud, V. Hilaire, O. Lisanti, F. Ramisse, J. D. Cavallo, and G. Vergnaud.** 2004. Variable number of tandem repeats in *Salmonella enterica* subsp. *enterica* for typing purposes. J. Clin. Microbiol. **42:**5722–5730.
47. **Sabat, A., J. Krzyszton-Russjan, W. Strzalka, R. Filipek, K. Kosowska, W. Hryniewicz, J. Travis, and J. Potempa.** 2003. New method for typing *Staphylococcus aureus* strains: multiple-locus variable-number tandem repeat analysis of polymorphism and genetic relationships of clinical isolates. J. Clin. Microbiol. **41:**1801–1804.
48. **Said-Salim, B., B. Mathema, and B. N. Kreiswirth.** 2003. Community-acquired methicillin-resistant *Staphylococcus aureus*: an emerging pathogen. Infect. Control Hosp. Epidemiol. **24:**451–455.
49. **Salama, N., K. Guillemin, T. K. McDaniel, G. Sherlock, L. Tompkins, and S. Falkow.** 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. Proc. Natl. Acad. Sci. USA **97:**14668–14673.
50. **Saunders, N. A., A. Underwood, A. M. Kearns, and G. Hallas.** 2004. A virulence-associated gene microarray: a tool for investigation of the evolution and pathogenic potential of *Staphylococcus aureus*. Microbiology **150:**3763–3771.
51. **Shore, A., A. S. Rossney, C. T. Keane, M. C. Enright, and D. C. Coleman.** 2005. Seven novel variants of the staphylococcal chromosomal cassette *mec* in methicillin-resistant *Staphylococcus aureus* isolates from Ireland. Antimicrob. Agents Chemother. **49:**2070–2083.
52. **Shukla, S. K., M. E. Stemper, S. V. Ramaswamy, J. M. Conradt, R. Reich, E. A. Graviss, and K. D. Reed.** 2004. Molecular characteristics of nosocomial and Native American community-associated methicillin-resistant *Staphylococcus aureus* clones from rural Wisconsin. J. Clin. Microbiol. **42:**3752–3757.
53. **Struelens, M. J.** 1996. Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. Clin. Microbiol. Infect. **2:**2–11.
54. **Talaat, A. M., S. T. Howard, W. Hale, R. Lyons, H. Garner, and S. A. Johnston.** 2002. Genomic DNA standards for gene expression profiling in *Mycobacterium tuberculosis*. Nucleic Acids Res. **30:**e104.
55. **Tang, Y. W., G. W. Procop, and D. H. Persing.** 1997. Molecular diagnostics of infectious diseases. Clin. Chem. **43:**2021–2038.
56. **Teixeira, L. A., C. A. Resende, L. R. Ormonde, R. Rosenbaum, A. M. Figueiredo, H. de Lencastre, and A. Tomasz.** 1995. Geographic spread of epidemic multiresistant *Staphylococcus aureus* clone in Brazil. J. Clin. Microbiol. **33:**2400–2404.
57. **Vandenesch, F., T. Naimi, M. C. Enright, G. Lina, G. R. Nimmo, H. Heffernan, N. Liassine, M. Bes, T. Greenland, M. E. Reverdy, and J. Etienne.** 2003. Community-acquired methicillin-resistant *Staphylococcus aureus* carrying Panton-Valentine leukocidin genes: worldwide emergence. Emerg. Infect. Dis. **9:**978–984.
58. **Zaidi, N., K. Konstantinou, and M. Zervos.** 2003. The role of molecular biology and nucleic acid technology in the study of human infection and epidemiology. Arch. Pathol. Lab. Med. **127:**1098–1105.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Microbiological Methods 60 (2005) 247–257
Journal of Microbiological Methods
www.elsevier.com/locate/jmicmeth

Correlation of proteomic and transcriptomic profiles of *Staphylococcus aureus* during the post-exponential phase of growth

Alexander Scherl[a,1], Patrice François[b,*,1], Manuela Bento[b], Jacques M. Deshusses[a], Yvan Charbonnier[b], Véronique Converset[a], Antoine Huyghe[b], Nadia Walter[a], Christine Hoogland[c], Ron D. Appel[a,c,d], Jean-Charles Sanchez[a], Catherine G. Zimmermann-Ivol[a], Garry L. Corthals[a], Denis F. Hochstrasser[a,e], Jacques Schrenzel[b]

[a]*Biomedical Proteomics Research Group, Geneva University Hospitals, Geneva, Switzerland*
[b]*Genomic Research Laboratory, Geneva University Hospital, Geneva, Switzerland*
[c]*Swiss Institute of Bioinformatics, Geneva, Switzerland*
[d]*Medical Informatics Division, Geneva University Hospitals, Geneva, Switzerland*
[e]*Faculty of Science, Geneva University, Geneva, Switzerland*

Received 20 August 2004; received in revised form 17 September 2004; accepted 30 September 2004
Available online 10 November 2004

Abstract

A combined proteomic and transcriptomic analysis of *Staphylococcus aureus* strain N315 was performed to study a sequenced strain at the system level. Total protein and membrane protein extracts were prepared and analyzed using various proteomic workflows including: 2-DE, SDS-PAGE combined with microcapillary LC-MALDI-MS/MS, and multidimensional liquid chromatography. The presence of a protein was then correlated with its respective transcript level from *S. aureus* cells grown under the same conditions. Gene-expression data revealed that 97% of the 2′ 596 ORFs were detected during the post-exponential phase. At the protein level, 23% of these ORFs (591 proteins) were identified. Correlation of the two datasets revealed that 42% of the identified proteins (248 proteins) were amongst the top 25% of genes with highest mRNA signal intensities, and 69% of the identified proteins (406 proteins) were amongst the top 50% with the highest mRNA signal intensities. The fact that the remaining 31% of proteins were not strongly expressed at the RNA level indicates either that some low-abundance proteins were identified or that some transcripts or proteins showed extended half-lives. The most abundant

* Corresponding author. Tel.: +41 22 372 93 38; fax: +41 22 372 98 30.
E-mail address: patrice.francois@genomic.ch (P. François).
[1] Both authors contributed equally to this work.

doi:10.1016/j.mimet.2004.09.017

classes identified with the combined proteomic and transcriptomic approach involved energy production, translational activities and nucleotide transport, reflecting an active metabolism. The simultaneous large-scale analysis of transcriptomes and proteomes enables a global and holistic view of the *S. aureus* biology, allowing the parallel study of multiple active events in an organism.

Keywords: Transcriptome; Microarray; Proteome; Proteins; *Staphylococcus aureus*

1. Introduction

Staphylococcus aureus is a Gram positive bacterium responsible for a broad diversity of infections ranging from minor skin infections to life-threatening infections such as infective endocarditis or sepsis (Lowy, 1998). The high virulence of *S. aureus* isolates is poorly understood and—to a given extent—related to the presence of bacterial proteins belonging to various families such as adhesins, toxins, and exo-enzymes. The expression of the corresponding genes is under control of global regulators and strictly regulated during the growth phases (Novick and Jiang, 2003; Xiong et al., 2004). Additional factors contributing to the pathogenicity of *S. aureus* are related to its ability to rapidly acquire or develop a broad range of antibiotic resistance determinants. Only a few years after the introduction of methicillin as the drug of choice, Barber et al. described the emergence of methicillin-resistant *S. aureus* strains (MRSA) (Barber, 1961). The introduction of new drug classes was again rapidly followed by the isolation of resistant organisms (Chang et al., 2003; Hiramatsu, 2001; Kuroda et al., 2001). *S. aureus* is now considered as a prototype bacterial species responsible for nosocomial infections and MRSA has recently become a major cause of community-acquired infections (Naimi et al., 2003).

The recent sequencing and release of the complete DNA sequence of MRSA strains N315 and Mu50 (Kuroda et al., 2001), followed by that of the community-acquired MRSA strain MW2 (Baba et al., 2002), has opened new avenues for the study of virulence factors and pathogenicity in *S. aureus*. When taken together, genome-wide profiling via transcriptomics and proteomics would allow the comprehensive measurement of the *S. aureus* biological system. The recent evolution towards rapid protein identification has made large-scale proteome analysis possible. This is mainly due to achievements in protein and peptide separation, mass spectrometry techniques, and the combination of these technologies with powerful bioinformatics tools. However, there are still numerous challenges that lie ahead on the way to providing a comprehensive proteomic analysis. The number of proteins detected and identified by the different analytical strategies can be highly divergent. For example, the analysis of hydrophobic proteins remains a technical challenge due to their poor solubility in aqueous buffers. To overcome these limitations and in order to obtain the largest possible picture of expressed proteins, in terms of Open Reading Frames (ORFs), we have performed extensive sample fractionation followed by alternative analysis strategies. These strategies included two-dimensional gel electrophoresis (2-DE), SDS-PAGE, phase-partitioning, multidimensional-liquid chromatography and tandem mass spectrometry (MS/MS). The use of a co-solvent during the isoelectric focusing (Deshusses et al., 2003) allowed the analysis of the most soluble membrane and membrane-associated proteins by 2-DE. Presence of high amount of detergents during SDS-PAGE separation allowed strongly hydrophobic protein fractions enriched after phase-partitioning to be analyzed. Finally, the use of multidimensional-liquid chromatography introduced also a gel-free workflow, making analysis of either very large or very small proteins possible.

We have compared proteomic to transcriptomic profiles of the organism grown under the same conditions and have assessed them in terms of relative abundance and their representation in different functional categories.

2. Materials and methods

2.1. Reagents and chemicals

All chemicals purchased were of the highest purity grade, unless otherwise stated. MilliQ water (Millipore, Bedford, MA) was used for the preparation of all buffers and solvents. Methanol, hydrochloric acid, magnesium chloride, potassium chloride, saccharose and Coomassie brilliant blue R250 were purchased from Merck (Darmstadt, Germany). Acetonitrile and acrylamide:bis solution (37.5:1) were purchased from Biosolve (Valkenswaard, The Netherland). SDS, orthophosphoric acid and 2,2,2-trifluoroethanol (99.0%) were purchased from Fluka (Buchs, Switzerland). Trifluoroacetic acid, α-cyano-4-hydroxycinnamic acid, 1,4-dithioerythritol (DTE), ammonium bicarbonate, potassium chloride, potassium dihydrogenophosphate, iodoacetamide, CHAPS, glycerol, glycine, phosphate buffered saline, porcine trypsin and Tris were from Sigma (St. Louis, MO). IPG strips (4–7 and 3–10) were purchased from Amersham Biosciences (Piscataway, NJ). Agarose and molecular mass markers were purchased from BioRad (Hercules, CA). Muller Hinton broth was obtained from Difco (Detroit, MI), and saccharose from Merck. The murolytic enzyme lysostaphin (Ambicin) was purchased from Applied Microbiology (Tarrytown, NY). dCTP coupled to cyanine dyes were obtained from NEN (Perkin Elmer, Boston, MA, USA).

2.2. Growth conditions and time point

For protein extracts, *S. aureus* strain N315 (Kuroda et al., 2001) was grown with agitation at 37 °C in Mueller Hinton Broth (MHB; 200 mL in 1000-mL flask), as previously described (Scherl et al., 2004). At post-exponential phase ($OD_{540\ nm}$=6 corresponding to $2–3\times10^9$ cells/ml), cells were chilled on ice and harvested by centrifugation at 8000×*g* for 5 min at 4 °C. For preparation of membrane extracts, 20 mL culture aliquots were washed in 1.1 M saccharose-containing buffer (McDevitt et al., 1994), and then suspended in 2 mL aliquots of the same buffer containing 50 μg/mL of the hydrolytic enzyme lysostaphin for 10 min at 37 °C. Protoplasts were recovered after centrifugation (30 min at 8000×*g*) and hypo-osmotic shock was applied in the presence of 10 μg/mL DNase I (Fluka) to decrease the viscosity of the medium. Membrane pellets were obtained after ultracentrifugation at 110,000×*g* for 50 min in a Beckman Optima TLX (Beckman Coulter Intl, Nyon, Switzerland).

2.3. Protein extraction, analysis and identification

Total protein extraction and enriched membrane fraction were prepared using described protocols (Scherl et al., 2004). Membrane protein enrichments were further fractionated using a water/trifluoroethanol/chloroform phase partitioning system (Deshusses et al., 2003). 2-DE gels were loaded with total protein extracts (pI 3.5–10 and 4–7), total membrane extracts (pI 4–7) and soluble membrane extracts (pI 4–7). Spots were excised, *in-gel* digested with trypsin, and analyzed using a 4700 MALDI-TOF/TOF tandem mass spectrometer (Applied Biosystems, Framingham, MA). Proteins from the soluble membrane fraction and from the membrane pellet fraction were also separated by SDS-PAGE. Both gel lanes were sliced in 30 bands and *in-gel* digested with trypsin. Extracted peptides were loaded on a 10-cm long homemade column with an ID of 75 μm, packed with C_{18} reverse phase (YMS-ODS-AQ200, Michrom BioResource, Auburn, CA), and eluted directly on a MALDI target using a homemade spotting robot. For liquid enzymatic digestion, total protein extracts were reduced and alkylated with iodoacetamide in presence of 2 M urea and 0.05% SDS and digested with trypsin. Peptides were concentrated and desalted using an Oasis HLB 1 cm^3 10 mg solid-phase extraction cartridge (Waters, Milford, MA). Biphasic multidimensional liquid chromatography was performed in a 100-μm ID column; first packed with C_{18} reverse phase over 10 cm (at the tip) and then with 5 cm strong cation exchange phase (polysulfoethyl A BMSE05-03, PolyLC, Columbia, MD). Fractions obtained with 0, 10, 20, 40, 80, 150, 300, and 500 mM KCl were spotted on MALDI target plates and analyzed by MALDI-TOF/TOF. Database search and validation were performed with MASCOT 1.8 software, using the same search parameters and validations criteria as already described (Scherl et al., 2004).

2.4. Microarray manufacturing

The microarray was manufactured by in situ synthesis of 8454 long oligonucleotide probes (Agilent, Palo Alto, CA, USA), selected as previously described (Charbonnier et al., 2004). It covers approximately 99.2% of all ORFs as annotated in strains N315, Mu50 (Kuroda et al., 2001), MW2 (Baba et al., 2002) and COL (http://www.tigr.org). Briefly, the microarray contained 8492 staphylococcal specific oligonucleotides. Based on available sequence information 89% are common for the four strains used during the design whereas 11% are strain-specific capture elements. The experimental validation of this array has been described elsewhere (Charbonnier et al., 2004).

2.5. Preparation of labeled nucleic acids

S. aureus strain N315 was grown overnight in 2 mL MHB, as described for proteomics analysis. Total RNA was extracted from 2 mL of cells at $2–3\times10^9$ cells/ml, using the RNeasy kit (Qiagen, Basel, Switzerland) as previously described (Vaudaux et al., 2002). After additional DNase treatment, the absence of remaining DNA traces was evaluated by quantitative PCR (SDS 7700; Applied Biosystems) with assays specific for 16S rRNA and HU genes (Renzoni et al., in press). Batches of 10 μg total *S. aureus* RNA were labeled by Cy-3 dCTP using the SuperScript II (Invitrogen, Basel, Switzerland), following manufacturer's instructions. Labeled cDNA was then purified onto QiaQuick columns (Qiagen).

Purified genomic DNA (DNeasy) of strain N315 was labeled with Cy-5 dCTP using the Klenow fragment of DNA polymerase I (BioPrime, Invitrogen, Carlsbad, CA) and used as a control channel for per chip and per gene normalization, as described elsewhere (Talaat et al., 2002).

2.6. Hybridization and scanning parameters

Cy5-labelled DNA (0.5 μg) (Talaat et al., 2002) and Cy3-labelled cDNA (10 μg) mixture was diluted in 250 μl Agilent hybridization buffer, and hybridized at a temperature of 60 °C for 17 h in a dedicated hybridization oven (Robbins Scientific, Sunnyvale, CA, USA). Slides were washed, dried under nitrogen flow, and scanned (Agilent) using 100% PMT power for both wavelengths. Data were extracted and processed using Feature Extraction™ software (version 6.1.1, Agilent). For expression analysis, local background-subtracted signals were corrected for unequal dye incorporation or unequal load of labeled product. The algorithm consisted of a rank consistency filter and a curve fit using the default LOWESS (locally weighted linear regression) method. Data consisting of three independent biological experiments were expressed as Log_{10} ratios and analyzed using GeneSpring 6.1 (SiliconGenetics, Redwood City, CA, USA). For comparative study of the proteome and transcriptome content, fluorescent signal corresponding to N315 only were considered.

2.7. Bioinformatic tools and sequence analyses

Functional classes of all predicted ORFs were retrieved from the COG database (Tatusov et al., 1997, 2001), publicly available from the internet (http://www.ncbi.nlm.nih.gov/COG/index.html). The signal peptide of the predicted protein sequences was removed using the SignalP V2.0 tool (Nielsen et al., 1997), accessible through the internet (http://www.cbs.dtu.dk/services/SignalP-2.0/). The number of transmembrane segments was then predicted using the TMHMM2.0 tool (Sonnhammer et al., 1998), also accessible through the internet (http://www.cbs.dtu.dk/services/TMHMM-2.0/).

3. Results

3.1. Genome-wide study of S. aureus N315 mRNA levels

Total RNA extracts from three different N315 cultures yielded equivalent amounts of material, generally 10 ± 1 μg per mL of overnight culture. Raw fluorescent signals obtained from microarrays showed limited variability (Supplementary Table 1). Reproducibility yielded correlation coefficients >0.9 (Pearson) when analyzing replicates. By applying a cut-off defined as $2\times$SD obtained from negative control spots (Leonard et al., 2003), approximately 97% of probes covering the 2596 ORFs of N315 genome were considered as displaying positive

fluorescence signals, i.e. detectable transcripts (Supplementary Table 1). A list of these genes and their normalized expression value is shown in Supplementary Table 1 (www.genomic.ch/sup.php).

Functional classification of these genes according to the COG database and their predicted number of transmembrane domains is also indicated in Supplementary Table 1. The median fluorescent intensity values were 0.5, 0.7, 1.1 and 19.6 in the 1st, 2nd, 3rd and 4th quarter, respectively.

3.2. Proteomic analysis

Total protein extracts and membrane protein extracts were analyzed using various separation and identification techniques (summarized in Fig. 1). Total protein extracts were first digested in solution with trypsin, and the resulting peptides analyzed by multidimensional LC-MALDI-MS/MS. This resulted in the identification of 245 unique proteins. Proteins from total extracts were also separated by 2-DE. For these experiments, IPG strips from pI=4–7 and pI=3–10 were used. Protein spots were cut from the gels, digested with trypsin and analyzed by MALDI-MS/MS. This resulted in the identification of 211 unique proteins.

Membrane protein extracts were separated by 2-DE (pI=4–7). The resulting spots were digested and analyzed by MALDI-MS/MS. This led to the identification of only 38 unique proteins. In order to

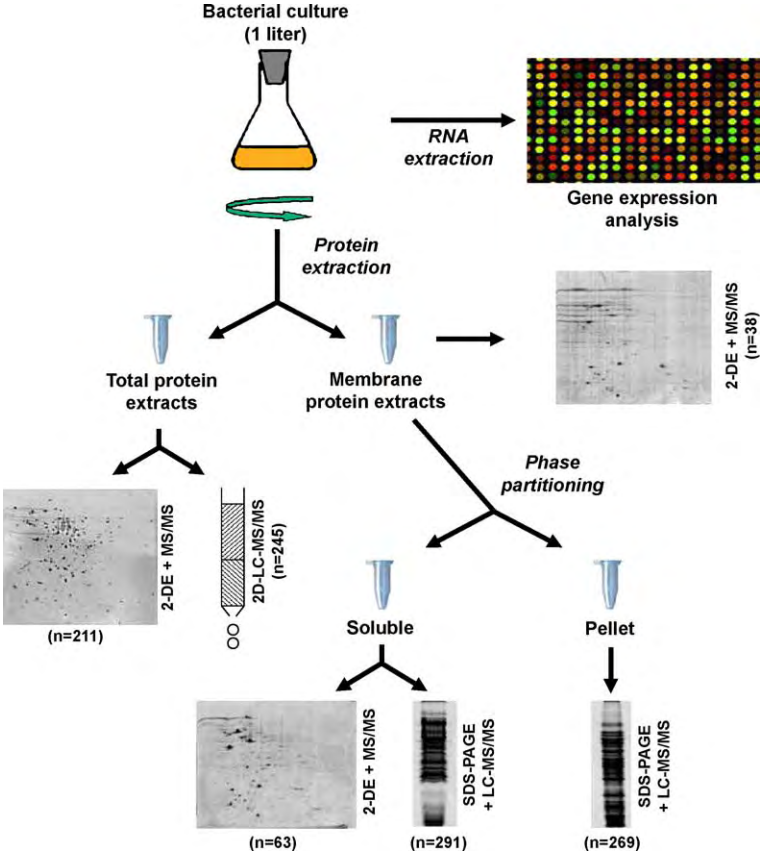


Fig. 1. Protein and RNA sample preparation with analysis steps. Bacteria were grown until post-exponential phase ($OD_{540\ nm}$=6 corresponding to 2–3×10^9 cells/ml). Total bacterial RNA was labeled, hybridized and analyzed on a custom genome-wide oligoarray. Total protein extracts were analyzed with 2-DE LC-MS/MS or multidimensional LC-MS/MS. Total membrane extracts were either analyzed by 2-DE/LC-MS/MS or subjected to a phase partitioning procedure. Soluble membrane extracts were analyzed with 2-DE/LC-MS/MS or 1-DE/LC-MS/MS. Insoluble membrane extracts were analyzed by 1-DE/LC-MS/MS. Numbers in parenthesis refer to the number of unique proteins identified by each specific workflow.

improve the detection of membrane proteins, the membrane fraction was further fractionated using a phase extraction system with water/trifluoroethanol/chloroform. The soluble fraction was analyzed using the 2-DE workflow, allowing the identification of 63 unique proteins. In addition, both the trifluoroethanol soluble phase and the protein pellet were separated by 1-DE, digested and analyzed by LC-MALDI-MS/MS. This resulted in the identification of 269 unique proteins from the pellet and 291 unique proteins from the soluble phase.

Combined together, these proteomic approaches resulted in the identification of 591 unique proteins (23% of all ORFs deduced from nucleotide sequence). For those proteins, both the sample fraction and the analytical method used for their identification are provided in Supplementary Table 2, together with their number of predicted transmembrane domains. Two unpredicted proteins were also identified: *S. aureus* N315 gene SA2420.1 (located between SA2420 and SA2421) and the N-terminus of SA2285 (ORF numbers are shown in Supplementary Table 2).

3.3. Comparison of protein and transcript profiles

The total number of identified protein (591) corresponds to approximately 23% of the genome. Comparison of protein and mRNA profiles reveals that all identified proteins were detected at the mRNA level excepted 3 proteins encoded by *S. aureus* N315 plasmid (Kuroda et al., 2001), namely SAP010, SAP018 and SAP029, corresponding to ORFs not present on the microarray.

The median fluorescence signals recorded for genes identified by the combined proteomic approaches (n=567)[2] was 2.6 fold higher than that observed for all other protein-coding genes (n=2038) described in N315 genome.

A comparison of the protein list with the 50% most abundant transcripts, as determined by their fluorescence signal intensities, showed an overlap of 406 proteins (69%). This value decreased to 248 proteins (42%) when considering only the 25% most expressed genes, thus reflecting a limited correlation between transcript abundance and relative amounts of proteins. Indeed, the remaining 31% of identified proteins corresponded to the lowest-level transcribed messengers. Fig. 2 represents the identified proteins as a function of their corresponding mRNA fluorescence signals. Individual data resulting from correlation between detected transcripts and identified proteins are summarized in Supplementary Table 3.

To identify the biological processes related to the expressed genes and protein content, all ORFs from Supplementary Table 3 were grouped into 19 functional classes, according to the COG classification. The relative proportion of the expressed genes and proteins in these functional classes is depicted in Fig. 3, as a function of their relative mRNA signal levels. This functional classification indicates that the most abundant classes identified by the combined proteomic and transcriptomic approaches involved energy production (COG group C), translational activities (COG J), and nucleotide transport (COG F). In contrast, proteins involved in lipid metabolism (COG I), defense mechanisms (COG V), and cell-cycle control (COG D) were only marginally identified (Fig. 3). The proportion of expressed proteins remained comparable across all functional classes, besides two. Genes involved in transport and metabolism (COG H) appeared to be more represented by abundant transcripts. On the contrary, genes involved in translation (COG J) appeared to be overrepresented in the more abundant mRNAs.

Transcripts involved in regulatory processes such as *sarA*, *sarR* and homologues, *sarH1*, *saeR*, *hit*, *vraR*, *sigB*, and *rot* were also found at the protein level. Among other proteins identified during the late exponential phase of growth, lipases, nucleases, and the IgG-binding protein A were present, possibly contributing to the virulence of the strain (Greene et al., 1996; Moreillon et al., 1995; Switalski et al., 1993).

The number of transmembrane (TM) segments of all predicted proteins suggests that 25% have at least one TM segment (Supplementary Table 1). However, combined proteomics approaches revealed that only 29% of proteins that were predicted to contain no TM segment could be detected; 24% of proteins were detected when predicted to contain one TM segment;

[2] The difference between the number of characterized peptides (591) and the list of matching hits (567) is related to newly detected ORFs and to the absence of plasmid-encoded genes on the microarray.

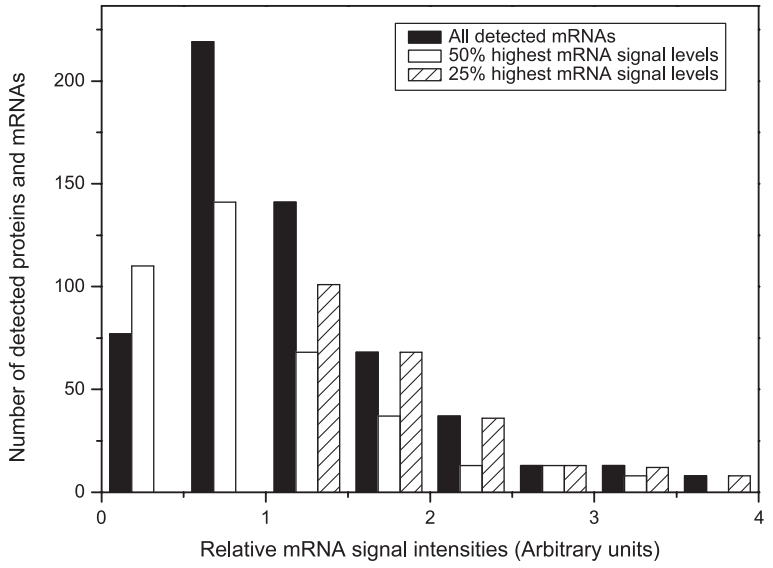


Fig. 2. Percentage of *S. aureus* proteins detected as a function of their relative mRNA signal intensity. The number of identified proteins was 567, 406, and 248, when considering 100% (■), 50% (□) or 25% (▨) of the highest mRNA fluorescence signals, respectively.

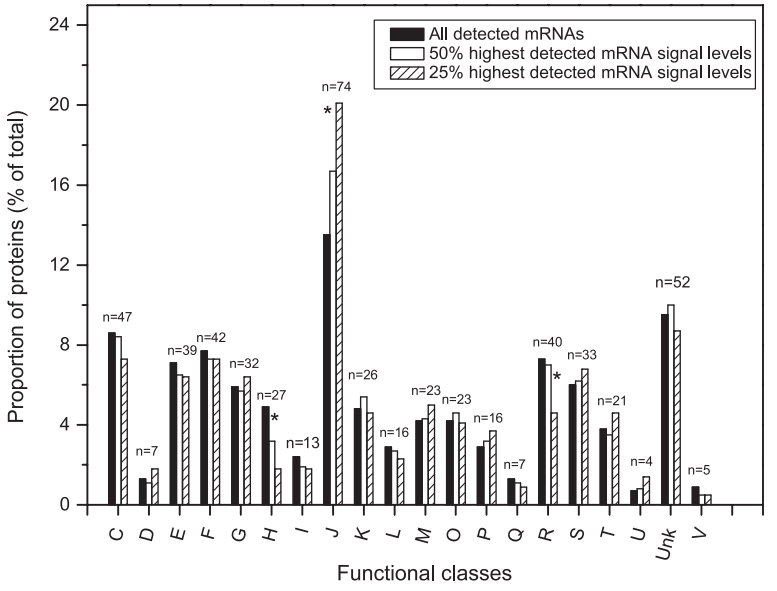


Fig. 3. Repartition of ORFs detected at the mRNA and protein level according to the COG classification for functional classes. The number of identified proteins was 567, 406, and 248; when considering 100% (■), 50% (□) or 25% (▨) of the highest mRNA fluorescence signals, respectively. Functional classes are: C=energy production and conversion; D=cell division and chromosome partitioning; E=amino acid transport and metabolism; F=nucleotide transport and metabolism; G=carbohydrate transport and metabolism; H=coenzyme metabolism; I=lipid metabolism; J=translation, ribosomal structure and biogenesis; K=transcription; L=DNA replication, recombination and repair; M=cell envelope biogenesis, outer membrane; N=cell motility; O=post-translational modification, protein turnover, chaperones; P=inorganic ion transport and metabolism; Q=secondary metabolites biosynthesis, transport and catabolism; R=general function prediction; S=function unknown; T=signal transduction mechanisms; U=secretion; Unk=function unknown; V=defense mechanism. Number of proteins for each category is indicated on top of the corresponding columns. *Categories showing variable profile (see also Discussion).

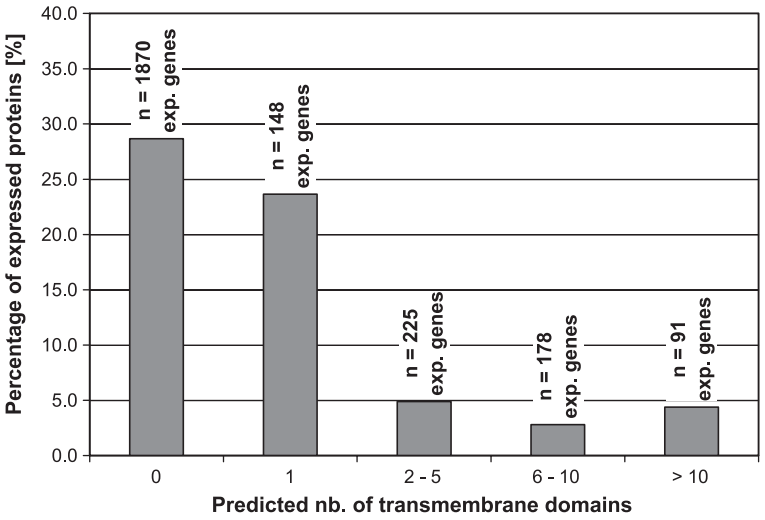


Fig. 4. Relationship between the number of predicted transmembrane segments and the identified proteins. Total number of expressed transcripts is indicated for each category of predicted transmembrane-containing proteins.

5% between 2 and 5 TM segments; 3% between 6 and 10 TM segments, and finally 4% to contain more than 10 TM segments (Fig. 4). Calculations from Supplementary Table 2 show that 179 proteins were identified in both membranes and total extracts, whereas 140 were unique to the total extracts and 272 to the membrane extracts.

4. Discussion

Combined transcriptomic and proteomic studies have already been reported and contributed to identify events taking place in specific metabolic pathways (Lee et al., 2003; Mader et al., 2002a,b), during growth phases or stringent response (Eymann et al., 2002). Global large-scale analyses were recently performed in *Escherichia coli* (Corbin et al., 2003) and *Bacillus anthracis* (Liu et al., 2004). In the latter study, 744 proteins (15% of the predicted ORFs) were identified from mature spores, using multidimensional LC-MS/MS. Global gene expression performed during the phase preceding the formation of mature spores provided a global picture of the sporulation mechanisms.

For an organism whose genome sequence has been determined, analysis of complex mixtures of tryptic peptides by tandem mass-spectrometry (MS/MS) provides a powerful means of determining its protein content. In this study, total protein extracts prepared simultaneously permitted the identification of approximately 23% of the theoretical proteome, deduced from the nucleotidic sequence of the strain N315 (Kuroda et al., 2001) genome. A comparison of the identified proteins with the 23% most abundant transcripts revealed that the majority of the identified proteins were highly transcribed, an observation in accordance with a comparable study on *E. coli* (Corbin et al., 2003) which revealed that the overwhelming majority of identified proteins (n=1147 proteins, 27% of the predicted ORFs) were detected at the transcript level.

In the proteomics approach used here, absolute quantification of proteins was not possible. However, a substantial proportion of proteins were detected in 2-DE gels of total extracts (see Supplementary Tables 2 and 3) suggesting that these proteins were abundant during the observed growth-phase. The observed lack of correlation between transcript abundance and identified proteins might therefore be related to different protein turnover and/or RNA stability. Another hypothesis relies on solubility of proteins in the buffers used in the different proteomic workflows: some highly expressed genes probably encode for highly insoluble proteins that are difficult to analyze even in the presence of detergents and still underrepresented in this study.

Relative RNA abundance is a function of neosynthesis and decay. However, mRNA stability is not a predictor of RNA abundance (Bernstein et al., 2002). In these experiments, ribosomal proteins (COG class J) represented 13% of the total proteome of *S. aureus* N315 (55 proteins identified). In our study this group constituted more than 20% of the proteome when considering the 25% most transcribed genes. This observation is in accordance with recent reports showing that mRNAs encoding for ribosomal proteins have generally slower decay rates than other mRNAs (Bernstein et al., 2002; Wang et al., 2002).

Classification by functional categories revealed that gene products involved in the formation of the translation complex and in the general metabolism constituted the largest groups of identified proteins. Minor groups were composed by defense and cell division mechanisms. These observations are consistent with a bacterial population that has essentially stopped dividing at that growth phase.

Protein analysis at steady-state revealed virulence factors known to be expressed during the post-exponential phase of growth such as *spa, clfA* and γ-haemolysin and several proteases (Novick and Jiang, 2003; Wolz et al., 1996). In contrast, we were not able to identify ClfB or Coa, two proteins reported to be synthesized only during the exponential phase of growth and that are rapidly degraded by metalloproteases or exported to the growth medium (McAleese et al., 2001; Ni Eidhin et al., 1998). Another important population recovered both at the protein and mRNA levels contained numerous regulators or effectors such as SarA, SarR and its homologue sarH1, SaeR, VraR, SigB, Rot. These factors have been reported to be transcribed after the exponential phase, thus in accordance with our study design (Manna and Cheung, 2001; Novick and Jiang, 2003).

A clear over-representation of proteins that contained no transmembrane domain was observed here. Using standard 2-DE technique, we identified only 5 proteins (<1%) predicted to contain transmembrane domains. The fact that (multiple) transmembrane proteins are difficult to solubilize, separate and digest with endopeptidases is particularly well documented for 2-DE experiments (Deshusses et al., 2003; Molloy et al., 2001). However, extraction with organic solvents or carbonate buffer combined with non-2-DE workflows has enabled high recovery of such hydrophobic proteins (Blonder et al., 2002, 2004; Ferro et al., 2000, 2002). In this study, we used combined workflows adapted for both soluble and hydrophobic proteins. Protein analysis of the membrane fractions led to the identification of a subset of the 55 transmembrane proteins. This population of 55 identified proteins, containing at least one (predicted) transmembrane domain corresponds to 10% of the *S. aureus* proteome instead of a theoretical value of 25%, considering the whole predicted proteome (Supplementary Table 1). It is important to stress that proteins that do not contain transmembrane domains may still be classified as membrane proteins, because they can be 'attached' to the cellular membrane via a linker, for example. Therefore, the membrane protein enrichment used in this study led not only to the identification of proteins with transmembrane domains, but also increased the overall yield of proteins without transmembrane domains. In addition, these strongly hydrophobic proteins, probably underrepresented, could partly contribute to the limited correlation observed between transcripts abundance and identified proteins.

Combined proteomics and transcriptomic studies will contribute to global understanding of biological systems. Even if present technologies used in proteomics workflows do not permit the identification of all expressed proteins, mainly for solubility and quantitative aspects, molecular effectors of several biological pathways are represented. This study enabled the identification of such effectors involved in energy production, translational activities and nucleotide transport. Similar experiments will be required grown under different conditions, to analyze poorly represented categories such as cell division or lipid metabolism. Our results dealing with specific points such as cell regulation during bacterial life cycle (Bernstein et al., 2002; Tao et al., 1999), provided evidence that biologically relevant information can be extracted from such large-scale explorations. Comparative assessment and integration of both genome/proteome-wide techniques, especially when quantitative, will improve our knowledge about bacterial growth, virulence and relation with its environment.

Acknowledgments

This work was supported by grants from the Swiss National Science Foundation PP00B-103002/1 (JS) and the Fondation pour Recherches Médicales (Gene Expression in Health and Diseases Program).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.mimet.2004.09.017.

References

Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K., Oguchi, A., Nagai, Y., Iwama, N., Asano, K., Naimi, T., Kuroda, H., Ciu, L., Yamamoto, K., Hiramatsu, K., 2002. Genome and virulence determinants of high virulence community-acquired MRSA. Lancet 359 (9320), 1819–1827.

Barber, M., 1961. Methicillin-resistant staphylococci. J. Clin. Pathol. 14, 385–393.

Bernstein, J.A., Khodursky, A.B., Lin, P.H., Lin-Chao, S., Cohen, S.N., 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. Proc. Natl. Acad. Sci. U. S. A. 99 (15), 9697–9702.

Blonder, J., Goshe, M.B., Moore, R.J., Pasa-Tolic, L., Masselon, C.D., Lipton, M.S., Smith, R.D., 2002. Enrichment of integral membrane proteins for proteomic analysis using liquid chromatography-tandem mass spectrometry. J. Proteome Res. 1 (4), 351–360.

Blonder, J., Conrads, T.P., Yu, L.R., Terunuma, A., Janini, G.M., Issaq, H.J., Vogel, J.C., Veenstra, T.D., 2004. A detergent- and cyanogen bromide-free method for integral membrane proteomics: application to *Halobacterium purple* membranes and the human epidermal membrane proteome. Proteomics 4 (1), 31–45.

Chang, S., Sievert, D.M., Hageman, J.C., Boulton, M.L., Tenover, F.C., Downes, F.P., Shah, S., Rudrik, J.T., Pupp, G.R., Brown, W.J., Cardo, D., Fridkin, S.K., 2003. Infection with vancomycin-resistant *Staphylococcus aureus* containing the vanA resistance gene. N. Engl. J. Med. 348 (14), 1342–1347.

Charbonnier, Y., Gettler, B.M., Francois, P., Lew, D.P., Schlegel, W., Schrenzel, J., 2004. A generic approach for the design of whole-genome oligoarrays validated for genotyping, deletion mapping and gene expression analysis on *Staphylococcus aureus*. Nat. Methods (in revision).

Corbin, R.W., Paliy, O., Yang, F., Shabanowitz, J., Platt, M., Lyons Jr., C.E., Root, K., McAuliffe, J., Jordan, M.I., Kustu, S., Soupene, E., Hunt, D.F., 2003. Toward a protein profile of *Escherichia coli*: comparison to its transcription profile. Proc. Natl. Acad. Sci. U. S. A. 100 (16), 9232–9237.

Deshusses, J.M., Burgess, J.A., Scherl, A., Wenger, Y., Walter, N., Converset, V., Paesano, S., Corthals, G.L., Hochstrasser, D.F., Sanchez, J.C., 2003. Exploitation of specific properties of trifluoroethanol for extraction and separation of membrane proteins. Proteomics 3 (8), 1418–1424.

Eymann, C., Homuth, G., Scharf, C., Hecker, M., 2002. *Bacillus subtilis* functional genomics: global characterization of the stringent response by proteome and transcriptome analysis. J. Bacteriol. 184 (9), 2500–2520.

Ferro, M., Seigneurin-Berny, D., Rolland, N., Chapel, A., Salvi, D., Garin, J., Joyard, J., 2000. Organic solvent extraction as a versatile procedure to identify hydrophobic chloroplast membrane proteins. Electrophoresis 21 (16), 3517–3526.

Ferro, M., Salvi, D., Riviere-Rolland, H., Vermat, T., Seigneurin-Berny, D., Grunwald, D., Garin, J., Joyard, J., Rolland, N., 2002. Integral membrane proteins of the chloroplast envelope: identification and subcellular localization of new transporters. Proc. Natl. Acad. Sci. U. S. A. 99 (17), 11487–11492.

Greene, C., Vaudaux, P.E., Francois, P., Proctor, R.A., McDevitt, D., Foster, T.J., 1996. A low-fibronectin-binding mutant of *Staphylococcus aureus* 879R4S has Tn918 inserted into its single fnb gene. Microbiology 142 (Pt. 8), 2153–2160.

Hiramatsu, K., 2001. Vancomycin-resistant *Staphylococcus aureus*: a new model of antibiotic resistance. Lancet, Infect. Dis. 1 (3), 147–155.

Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M., Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A., Mizutani-Ui, Y., Takahashi, N.K., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K., Hirakawa, H., Kuhara, S., Goto, S., Yabuzaki, J., Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H., Hiramatsu, K., 2001. Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. Lancet 357 (9264), 1225–1240.

Lee, J.H., Lee, D.E., Lee, B.U., Kim, H.S., 2003. Global analyses of transcriptomes and proteomes of a parent strain and an L-threonine-overproducing mutant strain. J. Bacteriol. 185 (18), 5442–5451.

Leonard, E.E., Takata, T., Blaser, M.J., Falkow, S., Tompkins, L.S., Gaynor, E.C., 2003. Use of an open-reading frame-specific *Campylobacter jejuni* DNA microarray as a new genotyping tool for studying epidemiologically related isolates. J. Infect. Dis. 187 (4), 691–694.

Liu, H., Bergman, N.H., Thomason, B., Shallom, S., Hazen, A., Crossno, J., Rasko, D.A., Ravel, J., Read, T.D., Peterson, S.N., Yates III, J., Hanna, P.C., 2004. Formation and composition of the *Bacillus anthracis* endospore. J. Bacteriol. 186 (1), 164–178.

Lowy, F.D., 1998. *Staphylococcus aureus* infections. N. Engl. J. Med. 339 (8), 520–532.

Mader, U., Antelmann, H., Buder, T., Dahl, M.K., Hecker, M., Homuth, G., 2002a. *Bacillus subtilis* functional genomics: genome-wide analysis of the DegS-DegU regulon by transcriptomics and proteomics. Mol. Genet. Genomics 268 (4), 455–467.

Mader, U., Homuth, G., Scharf, C., Buttner, K., Bode, R., Hecker, M., 2002b. Transcriptome and proteome analysis of *Bacillus*

subtilis gene expression modulated by amino acid availability. J. Bacteriol. 184 (15), 4288–4295.

Manna, A., Cheung, A.L., 2001. Characterization of sarR, a modulator of sar expression in *Staphylococcus aureus*. Infect. Immun. 69 (2), 885–896.

McAleese, F.M., Walsh, E.J., Sieprawska, M., Potempa, J., Foster, T.J., 2001. Loss of clumping factor B fibrinogen binding activity by *Staphylococcus aureus* involves cessation of transcription, shedding and cleavage by metalloprotease. J. Biol. Chem. 276 (32), 29969–29978.

McDevitt, D., Francois, P., Vaudaux, P., Foster, T.J., 1994. Molecular characterization of the clumping factor (fibrinogen receptor) of *Staphylococcus aureus*. Mol. Microbiol. 11 (2), 237–248.

Molloy, M.P., Phadke, N.D., Maddock, J.R., Andrews, P.C., 2001. Two-dimensional electrophoresis and peptide mass fingerprinting of bacterial outer membrane proteins. Electrophoresis 22 (9), 1686–1696.

Moreillon, P., Entenza, J.M., Francioli, P., McDevitt, D., Foster, T.J., Francois, P., Vaudaux, P., 1995. Role of *Staphylococcus aureus* coagulase and clumping factor in pathogenesis of experimental endocarditis. Infect. Immun. 63 (12), 4738–4743.

Naimi, T.S., LeDell, K.H., Como-Sabetti, K., Borchardt, S.M., Boxrud, D.J., Etienne, J., Johnson, S.K., Vandenesch, F., Fridkin, S., O'Boyle, C., Danila, R.N., Lynfield, R., 2003. Comparison of community- and health care-associated methicillin-resistant *Staphylococcus aureus* infection. JAMA 290 (22), 2976–2984.

Ni Eidhin, D., Perkins, S., Francois, P., Vaudaux, P., Hook, M., Foster, T.J., 1998. Clumping factor B (ClfB), a new surface-located fibrinogen-binding adhesin of *Staphylococcus aureus*. Mol. Microbiol. 30 (2), 245–257.

Nielsen, H., Engelbrecht, J., Brunak, S., Von Heijne, G., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 10 (1), 1–6.

Novick, R.P., Jiang, D., 2003. The staphylococcal saeRS system coordinates environmental signals with *agr* quorum sensing. Microbiology 149 (Pt. 10), 2709–2717.

Renzoni, A., Francois, P., Li, D., Kelley, W.L., Lew, D., Vaudaux, P., Schrenzel, J., 2004. Modulation of fibronectin adhesins and other virulence factors in a teicoplanin-resistant derivative of methicillin-resistant *Staphylococcus aureus*. Antimicrob. Agents Chemother. 48 (8), 2958–2965.

Scherl, A., Francois, P., Converset, V., Bento, M., Burgess, J.A., Sanchez, J.C., Hochstrasser, D.F., Schrenzel, J., Corthals, G.L., 2004. Nonredundant mass spectrometry: a strategy to integrate mass spectrometry acquisition and analysis. Proteomics 4 (4), 917–927.

Sonnhammer, E.L., Von Heijne, G., Krogh, A., 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. 6, 175–182.

Switalski, L.M., Patti, J.M., Butcher, W., Gristina, A.G., Speziale, P., Hook, M., 1993. A collagen receptor on *Staphylococcus aureus* strains isolated from patients with septic arthritis mediates adhesion to cartilage. Mol. Microbiol. 7 (1), 99–107.

Talaat, A.M., Howard, S.T., Hale, W., Lyons, R., Garner, H., Johnston, S.A., 2002. Genomic DNA standards for gene expression profiling in *Mycobacterium tuberculosis*. Nucleic Acids Res. 30 (20), e104.

Tao, H., Bausch, C., Richmond, C., Blattner, F.R., Conway, T., 1999. Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. J. Bacteriol. 181 (20), 6425–6440.

Tatusov, R.L., Koonin, E.V., Lipman, D.J., 1997. A genomic perspective on protein families. Science 278 (5338), 631–637.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., Koonin, E.V., 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. 29 (1), 22–28.

Vaudaux, P., Francois, P., Bisognano, C., Kelley, W.L., Lew, D.P., Schrenzel, J., Proctor, R.A., McNamara, P.J., Peters, G., von Eiff, C., 2002. Increased expression of clumping factor and fibronectin-binding proteins by *hemB* mutants of *Staphylococcus aureus* expressing small colony variant phenotypes. Infect. Immun. 70 (10), 5428–5437.

Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., Brown, P.O., 2002. Precision and functional specificity in mRNA decay. Proc. Natl. Acad. Sci. U. S. A. 99 (9), 5860–5865.

Wolz, C., McDevitt, D., Foster, T.J., Cheung, A.L., 1996. Influence of agr on fibrinogen binding in *Staphylococcus aureus* Newman. Infect. Immun. 64 (8), 3142–3147.

Xiong, Y.Q., Bayer, A.S., Yeaman, M.R., Van Wamel, W., Manna, A.C., Cheung, A.L., 2004. Impacts of sarA and agr in *Staphylococcus aureus* strain Newman on fibronectin-binding protein A gene expression and fibronectin adherence capacity in vitro and in experimental infective endocarditis. Infect. Immun. 72 (3), 1832–1836.

References

Antwerpen,M.H., Schellhase,M., Ehrentreich,F., Bier,F., Witte,W., and bel,U. (2007) DNA microarray for detection of antibiotic resistance determinants in Bacillus anthracis and closely related Bacillus cereus. *Molecular and Cellular Probes* 21: 152-160.

Aziz,R.K. and Kotb,M. (2008) Rise and persistence of global M1T1 clone of Streptococcus pyogenes. *Emerg Infect Dis* 14: 1511-1517.

Baker,G.C., Smith,J.J., and Cowan,D.A. (2003) Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* 55: 541-555.

Baratti-Mayer,D., Pittet,B., Montandon,D., Bolivar,I., Bornand,J.E., Hugonnet,S. et al. (2003) Noma: an "infectious" disease of unknown aetiology. *The Lancet Infectious Diseases* 3: 419-431.

Barmes,D.E., Enwonwu,C.O., Leclercq,M.H., Bourgeois,D., and Falkler,W.A. (1997) The need for action against oro-facial gangrene (noma). *Trop Med Int Health* 2: 1111-1114.

Bischoff,S.R., Tsai,S., Hardison,N.E., York,A.M., Freking,B.A., Nonneman,D. et al. (2008) Identification of SNPs and INDELS in swine transcribed sequences using short oligonucleotide microarrays. *BMC Genomics* 9: 252.

Bisno,A.L. and Stevens,D.L. (1996) Streptococcal Infections of Skin and Soft Tissues. *N Engl J Med* 334: 240-246.

Blackwood,C.B., Oaks,A., and Buyer,J.S. (2005) Phylum- and class-specific PCR primers for general microbial community analysis. *Appl Environ Microbiol* 71: 6193-6198.

Bolivar,I., Stadelmann,B., Baratti-Mayer,D., Mombelli,A., and And the Gesnoma study group . Bacterial diversity in oral samples of children in Niger with acute Noma, acute necrotizing gingivitis, and healthy controls. 2009.
Ref Type: Generic

Bourgeois,D.M. and Leclercq,M.H. (1999) The World Health Organization initiative on noma. *Oral Dis* 5: 172-174.

Brodie,E.L., DeSantis,T.Z., Parker,J.P.M., Zubietta,I.X., Piceno,Y.M., and Andersen,G.L. (2007) Urban aerosols harbor diverse and dynamic bacterial populations. *Proceedings of the National Academy of Sciences* 104: 299-304.

Cargill,M., Altshuler,D., Ireland,J., Sklar,P., Ardlie,K., Patil,N. et al. (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 22: 231-238.

Charbonnier,Y., Gettler,B., François,P., Bento,M., Renzoni,A., Vaudaux,P. et al. (2005) A generic approach for the design of whole-genome oligoarrays, validated for genomotyping, deletion mapping and gene expression analysis on Staphylococcus aureus. *BMC Genomics* 6: 95.

Chou,C.C., Chen,C.H., Lee,T.T., and Peck,K. (2004) Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression. *Nucleic Acids Res* 32: e99.

Claridge,M.F., Dawah,H.A., Wilson,M.R., and Association,S. (1997) *Species: The Units of Biodiversity*. Springer.

Coleman,W.B. and Tsongalis,G.J. (2005) *Molecular diagnostics*. Humana Press.

Colombo,A.P., Boches,S.K., Cotton,S.L., Goodson,J.M., Kent,R., Haffajee,A.D. et al. (2009) Comparisons of subgingival microbial profiles of refractory periodontitis, severe periodontitis, and periodontal health using the human oral microbe identification microarray. *J Periodontol* 80: 1421-1432.

Cutler,D.J., Zwick,M.E., Carrasquillo,M.M., Yohn,C.T., Tobin,K.P., Kashuk,C. et al. (2001) High-throughput variation detection and genotyping using microarrays. *Genome Res* 11: 1913-1925.

DeSantis,T.Z., Hugenholtz,P., Larsen,N., Rojas,M., Brodie,E.L., Keller,K. et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72: 5069-5072.

DeSantis,T.Z., Brodie,E.L., Moberg,J.P., Zubieta,I.X., Piceno,Y.M., and Andersen,G.L. (2007) High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microb Ecol* 53: 371-383.

Drancourt,M., Roux,V., Fournier,P.E., and Raoult,D. (2004) rpoB gene sequence-based identification of aerobic Gram-positive cocci of the genera Streptococcus, Enterococcus, Gemella, Abiotrophia, and Granulicatella. *J Clin Microbiol* 42: 497-504.

Duggan,D.J., Bittner,M., Chen,Y., Meltzer,P., and Trent,J.M. (1999) Expression profiling using cDNA microarrays. *Nat Genet* 21: 10-14.

Eckburg,P.B., Lepp,P.W., and Relman,D.A. (2003) Archaea and their potential role in human disease. *Infect Immun* 71: 591-596.

Enwonwu,C.O. (2007) Ruminations on the causation of noma. *Stomatologie* 104: 43–48.
Enwonwu,C.O., Falkler,W.A., and Idigbe,E.O. (2000) Oro-facial gangrene (noma/cancrum oris): pathogenetic mechanisms. *Crit Rev Oral Biol Med* 11: 159-171.

Enwonwu,C.O., Falkler,W.A., Idigbe,E.O., and Savage,K.O. (1999) Noma (cancrum oris): questions and answers. *Oral Dis* 5: 144-149.

Enwonwu,C.O., Falkler,W.A., and Phillips,R.S. (2006) Noma (cancrum oris). *Lancet* 368: 147-156.

Falkler,W.A., Enwonwu,C.O., and Idigbe,E.O. (1999a) Isolation of Fusobacterium necrophorum from cancrum oris (noma). *The American Journal of Tropical Medicine and Hygiene* 60: 150-156.

Falkler,W.A., Enwonwu,C.O., and Idigbe,E.O. (1999b) Microbiological understandings and mysteries of noma (cancrum oris). *Oral Dis* 5: 150-155.

Fan,J.B., Chen,X., Halushka,M.K., Berno,A., Huang,X., Ryder,T. et al. (2000) Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res* 10: 853-860.

Fish,D.J., Horne,M.T., Brewood,G.P., Goodarzi,J.P., Alemayehu,S., Bhandiwad,A. et al. (2007) DNA multiplex hybridization on microarrays and thermodynamic stability in solution: a direct comparison. *Nucleic Acids Res* 35: 7197–7208.

Francois,P., Charbonnier,Y., Jacquet,J., Utinger,D., Bento,M., Lew,D. et al. (2006) Rapid bacterial identification using evanescent-waveguide oligonucleotide microarray classification. *J Microbiol Methods* 65: 390-403.

Francois,P., Harbarth,S., Huyghe,A., Renzi,G., Bento,M., Gervaix,A. et al. (2008) Methicillin-resistant Staphylococcus aureus, Geneva, Switzerland, 1993-2005. *Emerg Infect Dis* 14: 304-307.

Francois,P., Huyghe,A., Charbonnier,Y., Bento,M., Herzig,S.+., Topolski,I. et al. (2005) Use of an automated multiple-locus, variable-number tandem repeat-based method for rapid and high-throughput genotyping of Staphylococcus aureus isolates. *J Clin Microbiol* 43: 3346-3355.

Frandsen,E.V., Reinholdt,J., Kjeldsen,M., and Kilian,M. (1995) In vivo cleavage of immunoglobulin A1 by immunoglobulin A1 proteases from Prevotella and Capnocytophaga species. *Oral Microbiol Immunol* 10: 291-296.

Garzoni,C., Francois,P., Huyghe,A., Couzinet,S., Tapparel,C., Charbonnier,Y. et al. (2007) A global view of Staphylococcus aureus whole genome expression upon internalization in human epithelial cells. *BMC Genomics* 8: 171.

Gharbia,S.E., Haapasalo,M., Shah,H.N., Kotiranta,A., Lounatmaa,K., Pearce,M.A., and Devine,D.A. (1994) Characterization of Prevotella intermedia and Prevotella nigrescens isolates from periodontic and endodontic infections. *J Periodontol* 65: 56-61.

Grasselli,E., François,P., Gutacker,M., Gettler,B., Benagli,C., Convert,M. et al. (2008) Evidence of horizontal gene transfer between human and animal commensal Escherichia coli strains identified by microarray. *FEMS Immunol Med Microbiol*.

Halgren,R.G., Fielden,M.R., Fong,C.J., and Zacharewski,T.R. (2001) Assessment of clone identity and sequence fidelity for 1189 IMAGE cDNA clones. *Nucleic Acids Res* 29: 582-588.

Halushka,M.K., Fan,J.B., Bentley,K., Hsie,L., Shen,N., Weder,A. et al. (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet* 22: 239-247.

Harbarth,S., François,P., Shrenzel,J., Fankhauser-Rodriguez,C., Hugonnet,S., Koessler,T. et al. (2005) Community-associated methicillin-resistant Staphylococcus aureus, Switzerland. *Emerg Infect Dis* 11: 962-965.

Harmsen,D., Rothgänger,J., Frosch,M., and Albert,J. (2002) RIDOM: Ribosomal Differentiation of Medical Micro-organisms Database. *Nucleic Acids Res* 30: 416-417.

He,Z., Gentry,T.J., Schadt,C.W., Wu,L., Liebich,J., Chong,S.C. et al. (2007) GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J* 1: 67-77.

He,Z., Wu,L., Li,X., Fields,M.W., and Zhou,J. (2005) Empirical establishment of oligonucleotide probe design criteria. *Appl Environ Microbiol* 71: 3753-3760.

Hong,B.X., Jiang,L.F., Hu,Y.S., Fang,D.Y., and Guo,H.Y. (2004) Application of oligonucleotide array technology for the rapid detection of pathogenic bacteria of foodborne infections. *J Microbiol Methods* 58: 403-411.

Horning,G.M. (1996) Necotizing gingivostomatitis: NUG to noma. *Compend Contin Educ Dent* 17: 951-4, 956, 957.

Hsiao,C.R., Huang,L., Bouchara,J.P., Barton,R., Li,H.C., and Chang,T.C. (2005) Identification of Medically Important Molds by an Oligonucleotide Array. *J Clin Microbiol* 43: 3760-3768.

Huang,A., Li,J.W., Shen,Z.Q., Wang,X.W., and Jin,M. (2006) High-throughput identification of clinical pathogenic fungi by hybridization to an oligonucleotide microarray. *J Clin Microbiol* 44: 3299-3305.

Hughes,T.R., Mao,M., Jones,A.R., Burchard,J., Marton,M.J., Shannon,K.W. et al. (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* 19: 342-347.

Huyghe,A., Francois,P., Charbonnier,Y., Tangomo-Bento,M., Bonetti,E.J., Paster,B.J. et al. (2008) Novel microarray design strategy to study complex bacterial communities. *Appl Environ Microbiol* 74: 1876-1885.

Huyghe,A., Francois,P., Tangomo,M., Girard,M., Baratti-Mayer,D., Bolivar,I. et al. (2009) Microarray analysis of the microflora of noma lesions. *Submitted*.

Isenbarger,T.A., Finney,M., Rìos-Velàzquez,C., Handelsman,J., and Ruvkun,G. (2008) Miniprimer PCR, a New Lens for Viewing the Microbial World. *Appl Environ Microbiol* 74: 840–849.

Islam,Z., van Belkum,A., Wagenaar,J.A., Cody,A.J., de Boer,A.G., Tabor,H. et al. (2009) Comparative genotyping of Campylobacter jejuni strains from patients with Guillain-Barré syndrome in Bangladesh. *PLoS ONE* 4: e7257.

Jaing,C., Gardner,S., McLoughlin,K., Mulakken,N., egria-Hartman,M., Banda,P. et al. (2008) A Functional Gene Array for Detection of Bacterial Virulence Elements. *PLoS ONE* 3: e2163.

Kane,M.D., Jatkoe,T.A., Stumpf,C.R., Lu,J., Thomas,J.D., and Madore,S.J. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res* 28: 4552-4557.

Kassinen,A., Krogius-Kurikka,L., Mäkivuokko,H., Rinttilä,T., Paulin,L., Corander,J. et al. (2007) The fecal microbiota of irritable bowel syndrome patients differs significantly from that of healthy subjects. *Gastroenterology* 133: 24-33.

Köck,R., Harlizius,J., Bressan,N., Laerberg,R., Wieler,L., Witte,W. et al. (2009) Prevalence and molecular characteristics of methicillin-resistant Staphylococcus aureus (MRSA) among pigs on German farms and import of livestock-related MRSA into hospitals. *Eur J Clin Microbiol Infect Dis*.

Koessler,T., Francois,P., Charbonnier,Y., Huyghe,A., Bento,M., Dharan,S. et al. (2006) Use of Oligoarrays for Characterization of Community-Onset Methicillin-Resistant Staphylococcus aureus. *J Clin Microbiol* 44.

Kothapalli,R., Yoder,S., Mane,S., and Loughran,T. (2002) Microarray results: how accurate are they? *BMC Bioinformatics* 3: 22.

Kuhbacher,T., Ott,S.J., Helwig,U., Mimura,T., Rizzello,F., Kleessen,B. et al. (2006) Bacterial and fungal microbiota in relation to probiotic therapy (VSL#3) in pouchitis. *Gut* 55: 833-841.

Leaw,S.N., Chang,H.C., Barton,R., Bouchara,J.P., and Chang,T.C. (2007) Identification of Medically Important Candida and Non-Candida Yeast Species by an Oligonucleotide Array. *J Clin Microbiol* 45: 2220-2229.

Leinberger,D.M., Schumacher,U., Autenrieth,I.B., and Bachmann,T.T. (2005) Development of a DNA Microarray for Detection and Identification of Fungal Pathogens Involved in Invasive Mycoses. *J Clin Microbiol* 43.

Leonard,E.E., Takata,T., Blaser,M.J., Falkow,S., Tompkins,L.S., and Gaynor,E.C. (2003) Use of an open-reading frame-specific Campylobacter jejuni DNA microarray as a new genotyping tool for studying epidemiologically related isolates. *J Infect Dis* 187: 691-694.

Lepp,P.W., Brinig,M.M., Ouverney,C.C., Palm,K., Armitage,G.C., and Relman,D.A. (2004) Methanogenic Archaea and human periodontal disease. *Proc Natl Acad Sci U S A* 101: 6176-6181.

Leski,T.A., Lin,B., Malanoski,A.P., Wang,Z., Long,N.C., Meador,C.E. et al. (2009) Testing and validation of high density resequencing microarray for broad range biothreat agents detection. *PLoS ONE* 4: e6569.

Li,F. and Stormo,G.D. (2001) Selection of optimal DNA oligos for gene expression arrays. *Bioinformatics (Oxford, England)* 17: 1067-1076.

Lindsay,J.A., Moore,C.E., Day,N.P., Peacock,S.J., Witney,A.A., Stabler,R.A. et al. (2006) Microarrays reveal that each of the ten dominant lineages of Staphylococcus aureus has a unique combination of surface-associated and regulatory genes. *J Bacteriol* 188: 669-676.

Loy,A., Maixner,F., Wagner,M., and Horn,M. (2007) probeBase--an online resource for rRNA-targeted oligonucleotide probes: new features 2007. *Nucleic Acids Res* 35: D800-D804.

Maidak,B.L., Cole,J.R., Lilburn,T.G., Parker,C.T., Saxman,P.R., Farris,R.J. et al. (2001) The RDP-II (Ribosomal Database Project). *Nucleic Acids Res* 29: 173-174.

Manichanh,C., Rigottier-Gois,L., Bonnaud,E., Gloux,K., Pelletier,E., Frangeul,L. et al. (2005) Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* gut.

Martens,M., Weidner,S., Linke,B., de Vos,P., Gillis,M., and Willems,A. (2007) A prototype taxonomic microarray targeting the rpsA housekeeping gene permits species identification within the rhizobial genus Ensifer. *Syst Appl Microbiol* 30: 390-400.

Mégevand,C., Gervaix,A., Heininger,U., Berger,C., Aebi,C., Vaudaux,B. et al. (2009) Molecular epidemiology of the nasal colonization by methicillin-susceptible Staphylococcus aureus in Swiss children. *Clin Microbiol Infect*.

Neumann,A.P. and Rehberger,T.G. (2009) MLST Analysis Reveals a Highly Conserved Core Genome among Poultry Isolates of Clostridium septicum. *Anaerobe*.

Nübel,U., Schmidt,P.M., Reiss,E., Bier,F., Beyer,W., and Naumann,D. (2004) Oligonucleotide microarray for identification of Bacillus anthracis based on intergenic transcribed spacers in ribosomal DNA. *FEMS Microbiol Lett* 240: 215-223.

Palacios,G., Quan,P.l., Jabado,O.J., Conlan,S., Hirschberg,D.L., Liu,Y. et al. (2007) Panmicrobial oligonucleotide array for diagnosis of infectious diseases. *Emerg Infect Dis* 13: 73-81.

Palmer,C., Bik,E.M., Eisen,M.B., Eckburg,P.B., Sana,T.R., Wolber,P.K. et al. (2006) Rapid quantitative profiling of complex microbial populations. *Nucleic Acids Res* 34: e5.

Palmer,C., Bik,E.M., Digiulio,D.B., Relman,D.A., and Brown,P.O. (2007) Development of the Human Infant Intestinal Microbiota. *PLoS Biol* 5: e177.

Paster,B.J., Falkler Jr,W.A., Enwonwu,C.O., Idigbe,E.O., Savage,K.O., Levanos,V.A. et al. (2002) Prevalent bacterial species and novel phylotypes in advanced noma lesions. *J Clin Microbiol* 40: 2187-2191.

Peplies,J., Glöckner,F.O., and Amann,R. (2003) Optimization strategies for DNA microarray-based detection of bacteria with 16S rRNA-targeting oligonucleotide probes. *Appl Environ Microbiol* 69: 1397-1407.

Perreten,V., Vorlet-Fawer,L., Slickers,P., Ehricht,R., Kuhnert,P., and Frey,J. (2005) Microarray-based detection of 90 antibiotic resistance genes of gram-positive bacteria. *J Clin Microbiol* 43: 2291-2302.

Phillips,R.S., Enwonwu,C.O., and Falkler,W.A. (1995) Pro- versus anti-inflammatory cytokine profile in African children with acute oro-facial noma (cancrum oris, noma). *Eur Cytokine Netw* 16: 70-77.

Pitondo-Silva,A., Minarini,L.A.R., Camargo,I.L.B.C., and Darini,A.L.c. (2009) Clonal relationships determined by multilocus sequence typing among enteropathogenic Escherichia coli isolated in Brazil. *Can J Microbiol* 55: 672-679.

Porreca,G.J., Zhang,K., Li,J.B., Xie,B., Austin,D., Vassallo,S.L. et al. (2007) Multiplex amplification of large sets of human exons. *Nat Meth* 4: 931-936.

Pozhitkov,A., Noble,P.A., Domazet-Loso,T., Nolte,A.W., Sonnenberg,R., Staehler,P. et al. (2006) Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Res* 34: e66.

Price,S.B. and McCallum,R.E. (1986) Enhancement of Bacteroides intermedius growth by Fusobacterium necrophorum. *J Clin Microbiol* 23: 22-28.

Relogio,A., Schwager,C., Richter,A., Ansorge,W., and Valcarcel,J. (2002) Optimization of oligonucleotide-based DNA microarrays. *Nucl Acids Res* 30: e51.

Renesto,P., Rovery,C., Schrenzel,J., Leroy,Q., Huyghe,A., Li,W. et al. (2008) Rickettsia conorii transcriptional response within inoculation eschar. *PLoS ONE* 3: e3681.

Rhee,S.K., Liu,X., Wu,L., Chong,S.C., Wan,X., and Zhou,J. (2004) Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl Environ Microbiol* 70: 4303-4317.

SantaLucia,J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighborΓÇëthermodynamics. *Proc Natl Acad Sci U S A* 95: 1460-1465.

Saunders,N.A., Underwood,A., Kearns,A.M., and Hallas,G. (2004) A virulence-associated gene microarray: a tool for investigation of the evolution and pathogenic potential of Staphylococcus aureus. *Microbiology* 150: 3763-3771.

Schena,M., Shalon,D., Davis,R.W., and Brown,P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467-470.

Sergeev,N., Distler,M., Vargas,M., Chizhikov,V., Herold,K.E., and Rasooly,A. (2006) Microarray analysis of Bacillus cereus group virulence factors. *J Microbiol Methods* 65: 488-502.

Slots,J. (1986) Bacterial specificity in adult periodontitis. A summary of recent work. *J Clin Periodontol* 13: 912-917.

Slots,J., Bragd,L., Wikström,M., and Dahlén,G. (1986) The occurrence of Actinobacillus actinomycetemcomitans, Bacteroides gingivalis and Bacteroides intermedius in destructive periodontal disease in adults. *J Clin Periodontol* 13: 570-577.

Srour,M.L., Watt,B., Phengdy,B., Khansoulivong,K., Harris,J., Bennett,C. et al. (2008) Noma in Laos: Stigma of Severe Poverty in Rural Asia. *Am J Trop Med Hyg* 78: 539-542.

Suzuki,S., Ono,N., Furusawa,C., Kashiwagi,A., and Yomo,T. (2007) Experimental optimization of probe length to increase the sequence specificity of high-density oligonucleotide microarrays. *BMC Genomics* 8: 373.

Taylor,E., Cogdell,D., Coombes,K., Hu,L., Ramdas,L., Tabor,A. et al. (2001) Sequence verification as quality-control step for production of cDNA microarrays. *BioTechniques* 31: 62-65.

Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A., Peterson,S., Wessels,M.R. et al. (2002) Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae. *Proc Natl Acad Sci U S A* 99: 12391-12396.

Tiquia,S.M., Wu,L., Chong,S.C., Passovets,S., Xu,D., Xu,Y., and Zhou,J. (2004) Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. *BioTechniques* 36: 664-70, 672, 674.

Tomazinho,L.F. and vila-Campos,M.J. (2007) Detection of Porphyromonas gingivalis, Porphyromonas endodontalis, Prevotella intermedia, and Prevotella nigrescens in chronic endodontic infection. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 103: 285-288.

Troesch,A., Nguyen,H., Miyada,C.G., Desvarenne,S., Gingeras,T.R., Kaplan,P.M. et al. (1999) Mycobacterium species identification and rifampin resistance testing with high-density DNA probe arrays. *J Clin Microbiol* 37: 49-55.

van Winkelhoff,A.J., Carlee,A.W., and de Graaff,J. (1985) Bacteroides endodontalis and other black-pigmented Bacteroides species in odontogenic abscesses. *Infect Immun* 49: 494-497.

Varshney,A.K., Mediavilla,J.R., Robiou,N., Guh,A., Wang,X., Gialanella,P. et al. (2009) Diverse enterotoxin gene profiles among clonal complexes of Staphylococcus aureus isolates from the Bronx, New York. *Appl Environ Microbiol* 75: 6839-6849.

Volokhov,D., Rasooly,A., Chumakov,K., and Chizhikov,V. (2002) Identification of Listeria Species by Microarray-Based Assay. *J Clin Microbiol* 40: 4720-4728.

Wang,D.G., Fan,J.B., Siao,C.J., Berno,A., Young,P., Sapolsky,R. et al. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280: 1077-1082.

Wang,D., Urisman,A., Liu,Y.T., Springer,M., Ksiazek,T.G., Erdman,D.D. et al. (2003) Viral Discovery and Sequence Recovery Using DNA Microarrays. *PLoS Biology* 1: e2.

Wang,Z., Daum,L.T., Vora,G.J., Metzgar,D., Walter,E.A., Canas,L.C. et al. (2006) Identifying influenza viruses with resequencing microarrays. *Emerg Infect Dis* 12: 638-646.

Wennström,J.L., Dahlén,G., Svensson,J., and Nyman,S. (1987) Actinobacillus actinomycetemcomitans, Bacteroides gingivalis and Bacteroides intermedius: predictors of attachment loss? *Oral Microbiol Immunol* 2: 158-162.

Wertz,J.E., Goldstone,C., Gordon,D.M., and Riley,M.A. (2003) A molecular phylogeny of enteric bacteria and implications for a bacterial species concept. *J Evol Biol* 16: 1236-1248.

Woese,C.R. (1987) Bacterial evolution. *Microbiol Rev* 51: 221-271.

Wong,R.S.Y. and Chow,A.W. (2002) Identification of enteric pathogens by heat shock protein 60 kDa (HSP60) gene sequences. *FEMS Microbiol Lett* 206: 107-113.

Zhu,X.Y., Zhong,T., Pandya,Y., and Joerger,R.D. (2002) 16S rRNA-Based Analysis of Microbiota from the Cecum of Broiler Chickens. *Appl Environ Microbiol* 68: 124-137.

Zoetendal,E.G., Cheng,B., Koike,S., and Mackie,R.I. (2004) Molecular microbial ecology of the gastrointestinal tract: from phylogeny to function. *Curr Issues Intest Microbiol* 5: 31-47.

Zwick,M., Mcafee,F., Cutler,D., Read,T., Ravel,J., Bowman,G. et al. (2004) Microarray-based resequencing of multiple Bacillus anthracis isolates. *Genome Biol* 6: R10.