METAGENOMIC ANALYSIS OF PERIODONTAL BACTERIA ASSOCIATED WITH GENERALIZED AGGRESSIVE PERIODONTITIS

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the
Graduate School of The Ohio State University

By
Khaled Altabbaei

Graduate Program in Dentistry

The Ohio State University

2016

Master’s Examination Committee:

Dr. Purnima Kumar, Advisor

Dr. Pooja Maney

Dr. John Walters
Copyright by
Khaled Altabtbaei
2016
ABSTRACT

**Background:** Generalized aggressive periodontitis (GAP) is a microbial disease with distinct clinical profile to the more clinically prevalent chronic periodontitis. Until recently, the capability to study the microbiome of the disease has not been possible.

**Methods:** 17 patients with GAP were enrolled. Patients were excluded if they had uncontrolled systemic disease, pregnancy, or had received antibiotic therapy in the last 3 months. Paper point samples were taken from the deep and shallow sulci of the affected patients. Genomic DNA was isolated, sheared, size selected and sequenced using Illumina. Sequences were filtered, and functionally annotated using the MG-RAST pipeline. The intra- and inter-group functional profiles were compared between the deep and shallow pockets, as well as with healthy individuals and with patients with chronic periodontitis (CP), using Wald test and FDR (DESeq2).

**Results:** Statistical analysis revealed no significant differences in the functional diversity between the shallow and deep pockets in generalized aggressive periodontitis, thus establishing that globalized functional capacity in GAP. Deep sites in GAP have higher functional capability than healthy sites, especially in regards to chemotaxis and motility.
Finally, the functional capability of generalized aggressive periodontitis is distinct from chronic periodontitis.

**Conclusion:** Distinct functional differences are found between health and disease in generalized aggressive periodontitis. These pockets, despite looking clinically healthy, have the same functional capacity as the deeper pockets in those patients. Moreover, these functional capability are distinct from the disease state in chronic periodontitis.
Dedicated to
My family; close and far
ACKNOWLEDGEMENTS

I wish to thank Dr. Purnima Kumar for giving me the opportunity to work on this project. Her presence, guidance and availability has made it possible for me to complete this thesis.

I also want to thank my committee members, Dr. John D. Walters and Dr. Pooja Maney. Their advice, and contribution was crucial to the development of this work.

I also wish to mention that this project could not have worked without the precious help of Dr. Shareef Dabdoub, Dr. Matt Mason, and Dr. Sukirth Ganesan. They have made it possible for me to conduct the laboratory manipulation, and formulate the data into a cogent thesis.

I want to thank as well all the members of the Ohio State Periodontology Department. Patient recruitment would have been nearly impossible without their vigilance and assistance. I would also like to thank Dr. Laura Porra for her help with recruiting patients in LSU.

Finally, I would like to thank my family and friends for their constant support.
VITA

March 9, 1985 … Born: Kuwait city, Kuwait

2009 … Obtained B. D. M.: University of Kuwait

FIELD OF STUDY

Major Field: Dentistry
TABLE OF CONTENTS

ABSTRACT .......................................................................................................................... ii
ACKNOWLEDGEMENTS................................................................................................. iv
VITA .................................................................................................................................... vi
FIELDS OF STUDY .......................................................................................................... vi
LIST OF FIGURES .......................................................................................................... viii
LIST OF TABLES ............................................................................................................ ix
CHAPTER 1: INTRODUCTION ....................................................................................... 1
CHAPTER 2: MATERIALS AND METHODS ................................................................. 17
CHAPTER 3: RESULTS ..................................................................................................... 22
CHAPTER 4: DISCUSSION .............................................................................................. 27
CHAPTER 5: CONCLUSION ............................................................................................ 29
REFERENCES ................................................................................................................ 30
LIST OF FIGURES

Figure 1: Logfold change in genes between GAP and health………………………….. 42
Figure 2: PCA comparison between GAP and healthy patients…………………………43
Figure 3: Comparison of GAP and healthy patients for the statistically significant genes as a fraction of their total .................................................................43
Figure 4: Carbohydrates – functional capability of GAP vs Health……………………49
Figure 5: Membrane transport – functional capability of GAP vs Health………………50
Figure 6: Phages, prophages, transposable elements, plasmids – functional capability of GAP vs Health ........................................................................................................... 50
Figure 7: Logfold change in genes between deep and shallow sites in GAP………….. 51
Figure 8: PCA comparison between shallow and deep sites in GAP………………….. 52
Figure 9: Logfold change in genes between GAP and CP…………………………… 53
Figure 10: PCA comparison between GAP and CP……………………………………..54
Figure 11: Comparison of GAP and CP patients for the statistically significant genes as a fraction of their total .................................................................54
LIST OF TABLES

Table 1: Demographic data................................................................. 35

Table 2: Core metagenome in GAP......................................................... 36

Table 3: Composition of statistically significant genes in GAP and Health as a percentage of their whole......................................................... 44

Table 4: Composition of statistically significant genes in GAP and CP as a percentage of their whole......................................................... 55
Periodontitis

Periodontitis is a microbially-driven inflammatory disease that is characterized by destruction of the supporting tissues of the teeth (periodontium). According to the 2009-2012 NHANES data, 47% of adults in the US suffer from periodontitis, with 8.9% suffering from severe forms of the disease. The disease is diagnosed based on clinical findings of inflammation, pocketing and attachment loss, and hence, it is possible that multiple diseases with the same clinical presentation, i.e., inflammation and destruction of the periodontium, may have multifactorial etiologies.

In 1999, the American Academy of Periodontology, based on the available knowledge, grouped periodontitis into two categories based on clinical, familial and immunological presentation: chronic, and aggressive. Despite the fact that each category may be composed of a heterogeneous group of diseases, the distinction between chronic and aggressive periodontitis is important clinically, since they respond differently to therapy, especially, non-surgical therapy. Further, the long-term prognosis of the affected teeth, and dental implants placed in these patients differs between the two disease phenotypes.
**Aggressive Periodontitis**

Aggressive periodontitis is a disease that is classified based on evidence of familial aggregation and a clinical presentation of rapid attachment loss in systemically healthy young adults. Patients with this disease typically demonstrate significant destruction of supporting structures that is not consistent with amount of microbial deposits. It is sub-classified into localized aggressive periodontitis (LAP) and generalized aggressive periodontitis (GAP). LAP affects the first molar/incisor regions with interproximal attachment loss on at least 2 permanent teeth, one of which is a first molar, with involvement of no more than 2 teeth other than the first molars and incisors. GAP patients have at least 3 permanent teeth affected by interproximal attachment loss other than the first molars and incisors.²

Localized aggressive periodontitis usually occurs at the time of puberty, while generalized aggressive periodontitis generally occurs in people under the age of 30. At the time of reclassification, evidence suggested that there is a robust serum antibody response to the infecting agents, in contrast to individuals with generalized aggressive periodontitis, who have poor serum response. Both conditions were thought to be associated with elevated proportions of *Aggregatibacter* (previously *Actinobacillus*) *actinomycetemcomitans*, and in some populations, *Porphyromonas gingivalis*. Phagocyte abnormalities and hyperresponsive macrophages were also associated with the two disease entities.²
Epidemiology of Aggressive Periodontitis

In the United States, as part of a comprehensive nationwide survey of school-age children, approximately 14,013 adolescents (Grade 8-12) were examined. It was estimated that aggressive periodontitis occurs in 0.4% in adolescents 13-15 years old, and 0.8% in adolescents 16-19 years old, establishing an increasing prevalence with age. Based on that survey, Löe and Brown estimated that there are 70,000 and 17,000 adolescents suffering from LAP and GAP, respectively. Of those that are 14-17 years old age, 0.53% had localized aggressive periodontitis, and 0.13% had generalized aggressive periodontitis. In a follow-up study, 91 subjects were re-examined 6 years later, Brown found that the areas with pre-existing attachment loss continued to have attachment loss, and some unaffected teeth started developing attachment loss. 35% of those with localized aggressive periodontitis had progressed to become generalized, while 62% had stayed the same. Of those that had generalized aggressive periodontitis, 82% continued to have the same diagnosis.

Globally, similar studies have been conducted, revealing similar statistics. In 14-17 year old Dutch population, the prevalence of aggressive periodontitis was estimated to be 0.1-0.2%. A Swiss study in 19-20 years old found the prevalence to be 0.13%. In the UK, it was estimated to be 0.1-0.2%. Studies in Chile and Brazil found the prevalence to be 0.32% and 5.5%, respectively. In Asia, studies in school-aged children in Turkey and Iran found the prevalence to be 0.6% and 0.13%, respectively. Finally, in a Sudanese children population, the prevalence was found to be 3.4%.
Demographics of Aggressive Periodontitis

It was estimated that the odds ratio for detecting localized aggressive periodontitis in a 17 years old US adolescent compared to a 14 years old to be 3.8. There is also evidence that the disease is more prevalent in males than in females, 0.78% compared to 0.52%, respectively. In those that are affected by GAP, the male to female ratio was 4.3:1. Similar result from a 3-year longitudinal Brazilian study corroborate the increase prevalence of aggressive periodontitis at 13 and 16 years old.

There is also a predilection for the disease to occur in young people of African origin. In the national survey in the United States, the prevalence was highest in Black Americans (2.6%), followed by Hispanics (0.5%), while it was the lowest in Caucasians (0.06%). This distinct difference was found in other studies worldwide. In the UK, 1% of those of Afro-Caribbean origins were affected compared to Asians (0.3%) and Caucasians (0.1%). A Sudanese study has found that there is a higher prevalence of aggressive periodontitis in children of African origins (6%) than those of Afro-Arab origins (2.3%).

Aggressive and Chronic periodontitis - Clinical Difference

Both conditions, in essence, encompass a variety of diseases that cause an immune-inflammatory response against a dysbiotic microbiome. Several factors can affect this inflammation, such as genetic factors, smoking, and systemic diseases that modify these immunologic disorders.

The main clinical feature that distinguishes the two variations of the disease is the rate of attachment loss. Chronic periodontitis is characterized by the slow to moderate
progression of periodontal destruction, while aggressive periodontitis is characterized by the rapid destruction of the periodontium.\textsuperscript{15} This difference in the rate of progression is presumably due to a myriad of factors that modify the inflammatory response, for example: the genetic makeup and the inflammatory response of the patient, the microbial antigens within in the microbiome, and environmental factors.\textsuperscript{16} A secondary clinical feature that has been used to distinguish between the two diseases is the amount of local factors such as plaque and calculus compared to the attachment loss. Aggressive periodontitis generally has few local factors compared to the attachment loss, while patients with chronic periodontitis demonstrate an extensive amount of plaque and calculus. However, the amount of local factors is not always distinct between the two entities, as some patients with aggressive periodontitis, especially those with generalized aggressive periodontitis, present with extensive local factors.\textsuperscript{12} This can lead to confusion as to whether the patient exhibits chronic or generalized aggressive forms of the disease.

**Periodontal health: A Polymicrobial Condition**

In periodontally healthy pockets, a biofilm exists that has a small number of microorganisms with a limited range of microorganisms. Each one of these microorganisms come with its own set of genetic composition and phenotypic expression. These microorganisms, through quorum sensing, have their phenotypic expression either enhanced or suppressed depending on its neighboring bacteria. Some species also have the ability to shelter the microbiome from foreign insults by producing the antimicrobial resistant agents such as b-lactamase, which reduces the fluctuations that could occur in the biofilm.\textsuperscript{17}
This allows the bacteria to work together in consortium to confer properties and functions that no single cell is able to fully perform on its own, much like a multicellular organism. A well-documented example is the co-dependence of the Veillonellae species on the lactate that is produced by the Streprococcus species as food source. This biofilm exists in harmony with the host, and possibly incur beneficial effects on the host such as create a barrier for non-indigenous pathogens to colonize the biofilm, inhibit the overgrowth of opportunistic microorganisms, mature the host immune system to recognize pathogenic microorganisms versus those associated with health (Reviewed - Kumar 2005\textsuperscript{18}). This microbiome is stable in its composition, and in some people, remains stable over time, even in cases of changes in the oral environment such as ingestion of food.\textsuperscript{19,20}

Bacterial Identification Methods

Much of the knowledge regarding the bacterial etiology of periodontitis has been derived using culture-based and targeted molecular methods. Culturing creates a bias in the identification of etiologic agents, since it promotes the growth of certain kinds of microorganisms in preference to others. While this type of identification methodology is important in monoinfections, it is less important in polymicrobial diseases. The metabolic pathways expressed in planktonic microorganisms are different than when they are part of a biofilm. For example, the genomic expression of \textit{Porphyromonas gingivalis} (P.g.) in its planktonic state is 18\% different than that when it is in biofilm. Moreover, some fastidious microorganisms, such as spirochetes, require other microorganisms before colonization and cannot be grown solitarily.\textsuperscript{21} It is also currently not possible to culture
every species, even in combination with other known species. A method that circumvents some of these issues is DNA-DNA hybridization. This method uses the properties of hybridization of the DNA double helix to identify the microorganism, where one strand from a specific organism is used as a probe that would hybridize with the complementary strand of the same species if it were present in the biofilm. This method allows for relatively fast identification a wide variety of microorganisms without the need for culturing. This method obviously requires prior knowledge of the target species in a community and requires stringent hybridization conditions to keep heterogeneous hybridization from other species low. Based on that, it becomes obvious that it still does not address the variety of microorganisms that are still not known, and its usage still requires knowledge on what is in the microbiome before testing it.

As technology progressed, it became possible to identify the species in the microbiome through open-ended methods. A highly conserved region of prokaryotic and archaeal DNA that codes for rDNA, hence termed as 16S rDNA, has been used to identify the constituents of the microbiome. By comparing this region in unknown bacteria to databases of bacteria with known 16s rDNA regions, it is possible to identify the bacteria and archaea in an open ended fashion. Once the organism is identified, its functional capability can be inferred. Currently, this method is a relatively cheap and fast method to identify the bacteria in the microbiome with moderate resolution. Since this method uses the highly conserved portion of the prokaryotic gene and as such identifies the genera/species that the bacteria is, it does not take into account mutations or horizontal gene transfers that may have occurred. This is especially true in some
genera/species where the species that have recently reclassified, notably those belonging to the families *Actinomyces*, and *Enterobacteriaceae*.\textsuperscript{22,23}

Advances in the sequencing technology has made it possible to sequence the entire genome of the bacterial community, which is known as the metagenome. Since the entire genome of the microbiome is analyzed, the identification of the bacterial constituents can be done based on similarity of long sequences of DNA as opposed to only the 16s rDNA sequences. This provides higher identification accuracy at the bacterial species and subspecies level.\textsuperscript{24} Moreover, the DNA sequences can be used to identify the genes and their actual pathogenic potential.

**Microbiology of Periodontitis at the Community Level**

The theory that there are specific microbial organisms that cause periodontitis was proposed in 1976 by Loesche et al (reviewed by Rosier et al., Frontiers in Cellular and Infection Microbiology). Socransky et al grouped bacteria based on their presence in healthy individuals or individuals with chronic periodontitis, that is, bacteria that are health compatible or associated with disease. His group has observed that some clusters of bacteria appear to be mostly associated with either health or disease. A distinct complex made of *Porphyromonas gingivalis*, *Tannerella forsythia*, and *Treponema denticola*, which they called the red complex, was mostly found in the patients with chronic periodontitis. Moreover, some clusters of bacteria were found to be rarely associated with other clusters of bacteria. For example, the green and yellow complexes were less commonly associated with the members of the more pathogenic complexes. This lead to the speculation that the presence of certain groups of bacteria could lead to
making the biofilm less hospitable to other bacteria. Furthermore, another pathogenic complex was found to be closely associated with the red complex, albeit less so with disease, which was termed as the orange complex. It was speculated that the presence of the orange complex is needed for the colonization of the red complex, and as such show a temporal relationship in colonization as well as the co-dependence of different species.\textsuperscript{25} Open-ended microbial identification methods have also corroborated this finding.\textsuperscript{26,27} Further research has led to a relatively small group of microorganisms that are associated with disease: \textit{P. gingivalis, P. endodontalis T. forsythia, A. actinomycetemcomitans, Prevotella intermedia, P. melaninogenica, P. denticola, P. nigrescens, P. corporis, P. disiens, Fusobacterium nucleatum, Parvimonas micra, Eikenella corrodens, Capnocytophaga gingivalis, Treponema denticola, T. socranskii, T. maltophilum, Treponema sp. Smibert-3, T. lecithinolyticum, Treponema putidum sp. nov, Eubacterium nodatum, E. saphenum, Campylobacter rectus, Filifactor alocis, Cryptobacterium curtum, Mogibacterium timidum, Peptostreptococcus magnus, Slackia exigua, Enterococcus faecalis, Escherichia coli and Bartonella sp.}\textsuperscript{28}

**Microbiota associated with GAP**

Our knowledge of aggressive periodontitis is limited due to its comparatively rare prevalence, which leads to difficulties in garnering a large cohort of patients for study.\textsuperscript{29} Moreover, due to changes in disease identification criteria over the last six decades, studies conducted done prior to the 1999 reclassification did not take into consideration the difference between localized and generalized aggressive periodontitis, and therefore translation of those findings to be applicable to current disease classification is
These early studies grouped both aggressive disease entities together, or simply did not identify whether the sampled patients have either form. Finally, direct comparison of different results is difficult due to evidence that different isolation methods, probing/sampling sequence, number of sampled teeth, and paper point size could change the microbial profile of the sample. That being said, previous efforts have shed light on the microbiota present in GAP.

A pilot investigation by Laksmana examined the 16s rDNA of 2 subjects with GAP and identified 208 species/phylotypes, with 129 species/phylotypes that were shared by both patients. 28-42 species represented 90% of all subgingival bacteria in each sample. *F. nucleatum, Prevotella spp., Porphyromonas gingivalis, A. actinomycetemcomitans* were identified in both samples.

Faveri looked at the 16s rDNA sequences of the microbiota in 10 patients with GAP. 57% of the microbiome were found to be from species have not been previously cultivated. From these samples, 110 species-level taxa were detected, with 70 of them being prevalent in the majority of individuals. The most prevalent genera were *Selemonas* and *Streptococcus*. Other commonly identified genera in these subjects included *Eubacterium* and *Peptostreptococcus*, Members of the phyla *Spirocheta* and *Actinobacteria* were also identified. Interestingly, none of the red complex pathogens, *F. nucleatum*, or *A. actinomycetemcomitans* were found in any of the samples.

**CP and GAP: Microbial difference**

In 2002, a meta-analysis by Mombelli looked at the microbial difference between aggressive and chronic periodontitis in studies that specifically compared chronic and
aggressive periodontitis (both forms) based on the prevalence and frequency of detection of the five species—Porphyromonas gingivalis, Prevotella intermedia, Aggregatibacter actinomycetemcomitans, Tannerella forsythia, and Campylobacter rectus. The studies included in this meta-analysis used a variety of methods to detect the microorganisms, such as culturing, DNA-probes, and indirect immunofluorescence. It was found that the presence or absence of these microorganisms did not discriminate between the two disease entities. Since then, a few studies have explicitly compared the microbiomes of the two disease entities. Since there is evidence that the microbiome is specific to the race of the individual, the following studies will be presented depending on the geographic region.

**Europe**

In Germany, Riep (GAP=44, CP=46; 16s rDNA) found that A.a., P.g., P.i., T.f., Treponemas denticola-like, treponema lecithinolyticum, Campylobacter rectus, Capnocytophaga ochracea, Fusobacterium spp., and F. nucleatum. Significantly more Treponema lecithinolyticum found in GAP vs CP. The rest were not significantly different. In the UK (CP=183, GAP=84; PCR), Nibali found no association with the two clinical pictures and the detection of A.a. And/or P.g.  

**South America**

In Brazil, Rescala et al compared the abundances of 40 different species using DNA-DNA hybrization, and found no difference between generalized aggressive and chronic periodontitis (GAP=17,CP=20). The same conclusion was reached by Lourenço
et al after analyzing 16S rDNA sequencing of 24 GAP and 35 CP patients. Comparisons of all species have found that detection of Aa, Cardiobacterium hominis, Peptostreptococcaceae sp., P. alactolyticus, and absence of Fretibacterium spp., Fusobacterium naviforme/Fusobacterium nucleatum ss vincentii and Granulicatella adiacens/Granulicatella elegans were associated with GAP compared to CP.\textsuperscript{41}

In a Chilean study (GAP=6, CP=17) that used microscopic identification of 8 different species (\textit{A. actinomyecetemcomitans, P. gingivalis, P. intermedia/nigrescens, E. corrodens, F. nucleatum, Capnocytophaga sp., C. rectus, P. micra}), Gajardo found that C. rectus was isolated significantly more frequently and in larger numbers in aggressive periodontitis compared to chronic periodontitis.\textsuperscript{42} Casarin (GAP= 40, CP=28) found more \textit{A. actinomyecetemcomitans} and \textit{P. gingivalis} in GAP compared to CP.\textsuperscript{43}

\textbf{Asia}

In Chinese populations, Li et al (GAP=10,CP=10) sequenced the 16S rDNA of bacteria in GAP patients and compared them to a first-degree relatives that have CP. GAP patients had lower taxonomic diversity than patients with CP. The ten most abundant shared species between the two conditions were \textit{Porphyromonas gingivalis, Porphyromonas endodontalis, Tannerella forsythia, Treponema medium, Leptotrichia hofstadii, Treponema denticola, Prevotella intermedia, and Prevotella loescheii}, representing only 44.12% and 19.35% of the total constituents of the microbiome in CP and GAP respectively. The unique highly abundant species in GAP were \textit{Treponema denticola, Prevotella intermedia, Prevotella loescheii, Selenomonas infelix, and Prevotella tannerae}, each having an abundance of less than 1%. On the other hand, the
unique highly abundant species in CP were *Leptotrichia wadei* (2.82%), *Fusobacterium canifelinum* (2.48%), *Corynebacterium matruchotii* (1.16%), and *Prevotella scotosa* (0.7%). They also identified a higher abundance of *P. gingivalis* in GAP (35.88%) compared to CP (11.26%). *A. actinomyecetemcomitans* was not present in any of the GAP samples. The constituents of the microbiome of those with aggressive periodontitis were compared to both others with the same condition, and to their relatives. Interestingly, GAP people were closer in their microbial profile to their first-degree relatives with chronic periodontitis than to others with the same diagnosis of aggressive periodontitis.\(^{44}\)

In another Chinese study, Liu (GAP = 57, CP = 73;) quantified the levels of *Tannerella fosythia*, *Porphyromonas gingivalis* and *Aggregatibacter actinomycetemcomitans* using real-time PCR. *Porphyromonas gingivalis* was found in higher frequency and quantity in CP compared to GAP. *Aggregatibacter actinomycetemcomitans* was detected more in GAP but the difference was not statistically significant.\(^{45}\)

In Japan, Tomita (GAP=20, CP=20) quantified *Aggregatibacter actinomycetemcomitans*, *Porphyromonas gingivalis*, *Tannerella forsythia* using real-time PCR. None of the target bacteria was detected in healthy individuals. Moreover, no significant difference was found in the prevalence or abundance of the 3 pathogens between the two diseased groups. That being said, *Tannerella forsythia* was found 4 folds higher in CP than in GAP.\(^{46}\)
The microbiome is both individualized and universal

To understand the limitations of the previous studies, it would be helpful to examine these investigations in light of our current understanding of the microbiome associated with health and disease.

A major shift in our collective understanding of the microbiome occurred due to The Human Metagenome Project (HMP) which was a large multicenter study that investigated the microbiome of 242 healthy individuals across 18 body sites (supra- and subgingival plaque being among them). Both functional and organismal diversity were analyzed. It was found that at the human community level, each body site is usually inhabited by a few signature taxa, which makes the plurality of the inhabitants of that body site. These clades would constitute, on average, anywhere from 17% to 84% of the inhabitants of that body site. The most abundant taxa in supragingival plaque was found to be from the Actinomyces taxa, while the Prevotella dominates the subgingival plaque. Less dominant taxa were considered to be highly personalized, both at the body site level, and among different people.47,48 One explanation to this diversity has been confirmed by ethnicity studies, which found that in the oral cavity, the bacterial profile of a person is highly individualized, with only 2% bacteria being present in all ethnicities. Nevertheless, certain clusters of bacteria were found to be highly specific to certain ethnicities (74-91% specificity).37

Functionally, it was found that across body parts, the pathways in the microbial community metagenome where much less variable than the organismal diversity. Nevertheless, each body part had specific and highly abundant general metabolic pathways that can characterize the healthy function of that body part. While a relatively
few genes were specifically present or absent in any one body habitat, over two thirds were differentially abundant in the body habitats. These pathways appear to be relatively stable over time.\textsuperscript{47,48} The diversity of the microbial constituents along with the relative stability of the functional pathways could be due to the interchangeability of the microbial community members, each with the capability of providing the same functional pathway to the collective microbiome.\textsuperscript{49}

The gradual disruption of the healthy profile of the microbiome is termed dysbiosis.\textsuperscript{50} The current theory is that certain microorganisms in low abundance, "keystone species", are able to change the commensal microbial environment, and as such prepare the microbiome to be more virulent. Experimental evidence suggests that, in the oral cavity, P. gingivalis is a candidate keystone species.\textsuperscript{51} Some explanation of the dysbiotic transformation of the microbiome has been recently published by Yost. In a prospective study, it was found that the stable sites in individuals with periodontitis have higher microbial activity than the sites in periodontally healthy individuals, even though they both were clinically similar in appearance. Moreover, it found that the members of the aforementioned yellow complex, \textit{Streptococcus oralis}, \textit{S. intermedius}, \textit{S. mitis}, as well as \textit{Veillonella parvula} and \textit{Pseudomonas fluorescens} were found to be actively transcribing virulence factors.\textsuperscript{27} This suggests that the whole biofilm becomes virulent as opposed to only specific pathogens in it.

Based on the available evidence, we hypothesized that the microbiome associated with generalized aggressive periodontitis would differ significantly from both that associated with health and that of chronic periodontitis. We tested this hypothesis with the following specific aims:
1. Aim 1: To investigate whether generalized aggressive periodontitis is associated with a site-specific or global dysbiosis in the microbiome.

2. Aim 2: To compare the functional diversity of the periodontal microbiome in subjects with chronic and aggressive periodontitis.
Study population and study design

The research protocol was a cross-sectional study. The recruitment of patients occurred in two different locations; The Ohio State University, and Louisiana State University, as two separate studies.

In the Ohio State University, patients that were receiving nonsurgical periodontal therapy as a part of their treatment in the Graduate Periodontology Clinics of The Ohio State University were selected from April 2014 to February 2016.

Inclusion criteria

Patients recruited in the research project had to be at least 8 years of age, be able and willing to provide consent for the study (or obtain assent from guardian), and be diagnosed with Generalized Aggressive Periodontitis according to the criteria of the 1999 American Academy of Periodontology workshop.

Data from twenty-five healthy never-smokers (attachment loss ≤ 1, probing pocket depths ≤ 3) and twenty-five never-smokers with generalized moderate to severe chronic periodontitis (attachment loss ≥ 5, probing pocket depths ≥ 5) were used as the comparison groups.
Exclusion criteria

Patients were excluded from the research project if they were pregnant, had taken antibiotics in the last 3 months, were medicated with immunosuppressant medications or bisphosphonates, or presented with signs of uncontrolled systemic disease or diabetes.

Informed consent was obtained from every participating patient before entering the study. The study protocol and the informed consent forms were approved by The Ohio State Institutional Review Board (2014H0020).

Study design

Patients were seen at the time of scaling and root planing and consent obtained prior to any clinical procedure. At that time, the study protocol was explained, the investigator determined the patient's eligibility, the patient was given the opportunity to ask questions, and written informed consent form was obtained. Demographic and medical information was obtained from the patient and the chart.

During the visit, an assessment of clinical indices was initially conducted, including gingival index (GI, Loe and Silness\textsuperscript{52}), plaque index (PI, Silness and Loe\textsuperscript{52}). After isolation with cotton rolls, supragingival plaque and calculus were carefully removed using curettes. Following this step, paper points were carefully inserted in each tooth sulcus of the pocket investigated. In all cases, paper points were left in the site for 15-30 seconds, and immediately placed in 100\(\mu\)L of RNAlater and temporarily stored in ice for the duration of the appointment. Special attention was given to keep the paper points isolated from any saliva, and for them to be transferred to the RNAlater media as
quickly as possible to avoid contamination. Each patient had 3 sites with deep probing depth (GAPD) and 3 sites with shallow probing depth (GAPH) sampled (PPD ≥ 3 mm).

After the visit, the samples were transferred from ice to a controlled freezer (-80°C) and left in this location until the next step.

**DNA isolation**

At this step, the paper points were removed from the freezer and placed in a new sterile 1.5 mL collection tube. 200 µL of PBS was added and the samples were agitated for 30 minutes to allow for a maximum of the DNA to transfer from the paper points to the liquid. After this step, the paper points were removed and placed in a small punctured collection tube, which was then placed into a 1.5 mL tube and centrifuged to separate the DNA from the paper points.

180 µL of ATL buffer and 40 µL of proteinase K were added to the eluent and vortexed for 15 seconds. The samples were incubated in a water bath at a temperature of 56°C for a minimum of 2 hours. 200 µL of AL buffer was then added, the mix was agitated for 15 seconds and heated for 10 minutes in a dry bath at a temperature of 70°C. 200 µL of 100% ethanol was added to the mix, which was then agitated for 15 seconds.

The mix was then added to a QIAamp Spin Column and centrifuged for 1 minute. The filtrate was then discarded and 500 µL of AW1 washing solution was added to the spin column, centrifuged for 1 minute and the filtrate was again discarded. 500 µL of AW2 washing solution was then added to the same spin column, centrifuged for 3 minutes and the filtrate was discarded along with the 2.0mL collection tube.
The spin column was seated in a new 1.5 mL collection tube, 50 µL TE buffer with EDTA added to the filter incubated at room temperature for 5 minutes and centrifuged for 1 minute. The eluent was quantified in a NanoSpec ND 1000 Spectrophotometer and 100ng of DNA was used for whole genome shotgun sequences.

**Illumina Whole-Genome Sequencing**

With the help of the Illumina MiSeq paired-end platform (Illumina Inc., San Diego, California) located in a commercial facility (Molecular Research LP, Shallowater, Texas), Multiplex bacterial tag-encoded Whole Genome Shotgun Sequencing was performed. An Nextera DNA Sample Preparation Kit (Illumina) was then used according to manufacturer's instructions to complete library generation. Genomic DNA was then sheared enzymatically to obtain an average fragment size of 500 base pairs. Pooled libraries (12pM) were loaded to a 600 Cycles v3 Reagent cartridge (Illumina) and the sequencing was performed on Miseq (illumina).

**Metagenomic analysis**

The sequences were trimmed and filtered, and submitted electronically to the MG-RAST pipeline for quality processing and functional analysis. Sequences were functionally assigned by the MG-RAST server (Argonne National labs). Functional potential was compared between groups using KEGG (Kyoto Encyclopedia of Genes and Genomes) and SEED hierarchies.
**Statistical analysis**

The statistical analysis for functional potential comparisons were conducted using the Wald-test for inter- and intra-species diversity, and adjustment of the p-values through FDR (False Discovery Rates) methods using the statistical package DESeq2 within R.\textsuperscript{53}
CHAPTER 3

Results

Clinical data

From April 2014 to February 2016, 17 patients completed the study at The Ohio State University and Louisiana State University. The demographic data for these 17 patients, as well as the demographic data of 25 samples of chronic periodontitis (CP) can be found in appendix I.

GAP - Core Metagenome in Diseased Sites

A total of 8998 genes were identified at sites with deep periodontal pocket depths in subjects with GAP. 57.3% (5159 genes) can be considered as part of the core metagenome of this disease, based on their presence in 80% of the samples (appendix B).

The largest category of the core metagenome was related to metabolism of carbohydrates, which occupies 14.4% of genes (748 genes). Within this category, utilization and synthesis of mono-, di-, oligo-, and polysaccharides constituted 32% of the carbohydrate metabolism genes, while 23.1% of the genes were related to central carbohydrate metabolism pathways such as the TCA cycle, pyruvate metabolism, and glycolysis and gluconeogensis. Utilization of sugar alcohols such as ethanolamine,
glycerols, and mannitol, and fermentation of sugars to lactate, acetone, and ethanol occupied 17.9% of this category.

Utilization of amino acids and proteins contributed to 14.2% of the core metagenome. 41.2% of these genes were related to protein biosynthesis, degradation, processing, folding, and modification, while the utilization, synthesis and degradation of the proteinogenic amino acids lysine, threonine, methionine, cysteine, alanine, serine, and glycine constituted 20% of the metagenome of that category.

The third most abundant category of genes (12.2%) are clusters with hypothesized functions based functional coupling, but whose functions are not known (clustering-based subsystems).

**GAP vs Health**

9621 genes were identified in healthy patients and deep sites of patients with GAP. Of those genes, 4062 genes were differentially abundant in either condition (Wald test, p>0.05). After p-value adjustment for multiple comparison testing, 2027 were significantly more abundant in GAP, compared to 1519 genes in healthy patients. This suggests qualitative functional diversity in GAP compared to health (figure 1-2). 1388 (68.47%) genes were in the core metagenome of GAP.

When the genes that are significantly abundant in the two states are compared (figure 3), the majority of them were related to the central metabolic functions in the microbiome such as carbohydrate, protein and amino acid metabolism. Within this group, the largest contributor in both conditions were related to carbohydrates, which contribute 15.1% and 14.6% to the total metagenome in GAP and in healthy sites, respectively, with
the genes related to the central carbohydrate metabolism and its auxiliary pathways being the most significant contributors (figure 4). The total functional capability for carbohydrate utilization is depleted compared to health (268 GAP, 305 in health), especially in regards to utilization of di and oligosaccharides. The capability for fermentation, and metabolism of one-carbon such the case of methanogenesis, however, is enriched in GAP. The second largest contributor to the central metabolic functions is protein metabolism, which contribute to 14.7% and 16% in GAP and healthy, respectively.

Of the genes that are related to the virulence lifestyle, some genes in GAP appear to have an overall enrichment of both quality and quantity. For example: membrane transport (figure 5), a more diverse genetic makeup exists in GAP, in terms of quality (healthy: membrane secretion type II, IV, ABC transporters. GAP: membrane secretion type II, VI, VIII, ABC transporter, protein translocation across cytoplasmic membrane). Likewise, with regards to motility, there are 63 differentially enriched genes were found in GAP that were related to flagella, all of which are from 0.9-5.8 logfold change, compared to 1 gene in health, which is concerned with bacterial chemotaxis. Iron acquisition and metabolism, surprisingly, is more diverse in health compared to diseased sites (54 unique genes vs 46, abundance higher in health than in disease). Finally, in some categories, such as in phages, there is a distinct shift in function between the two conditions (figure 6), such as the enrichment of transposable elements in GAP compared to health (15 vs 0; in diseased sites; the majority of which are related to conjugative transposon in the order Bacteroidales) while 35 unique genes were found in health with regards to phages, prophages, compared to 15 in the diseased sites.
GAPD- GAPH - Global disease

10183 genes have been identified in patients with generalized aggressive periodontitis in both shallow and deep sites. When the shallow and deep sites were compared, 1725 genes were differentially abundant in either condition (Wald test, p>0.05), which represents %28.72 of the genes. When the p-value was adjusted for multiple comparison analysis, 1 gene only, belonging to the clustering based subsystem, reached statistical significance. As such, the microbiome in the deep and shallow sites is considered as functionally identical (figure 7).

GAP vs CP

A total of 9741 genes were identified in sites with disease in subjects with GAP and CP. Of those genes, 1982 were differentially abundant in either condition (Wald test, p>0.05), which represents 20.3% of the genes. After p-value adjustment for multiple comparison testing, 333 were significantly more abundant in GAP, compared to 761 genes in CP patients. (figures 9-10). 247 of those genes are part of the core metagenome of GAP, representing 74.1% of the identified abundant genes in this disease.

Of the genes that are significantly abundant in either condition (figure 11), the majority of them were related to the central metabolic functions in the microbiome such as carbohydrate, protein and amino acid metabolism. The largest contributor to the differences in their metagenome in both conditions is, again, the processes related to the metabolic potential for the two disease states. Genes related to carbohydrates are slightly
more depleted in GAP, while protein metabolism is slightly more enriched in GAP compared to CP.

Of the genes that are related to virulence, CP had a more diverse genetic capability than GAP. The production capability of cell wall and capsule components in CP is more diverse than in GAP (CP 77 vs. GAP 29); more gram positive and negative cell wall components, and capsular and extracellular polysaccharides are found in CP than in GAP. Moreover, membrane transport genes (35 in CP, 8 in GAP) are more abundant in CP, especially those related ECF class transporters, and ATP-dependent efflux. In relation to phages and transposable elements (39 in CP, 7 in GAP), GAP can be characterized by its enrichment of genes related to integrons, and the aforementioned conjugative transposons in Bacteriodales. In CP, phage capsid proteins, entry and exit, and packaging are more abundant. The capability of GAP to resist stresses is depleted compared to CP (GAP - 10 genes, CP - 13 genes). Capability of GAP to resist oxidative, osmotic, acid, and heat/cold challenges is depleted compared to CP. However, periplasmic stress response is enhanced in GAP compared to CP. The resistance to antibiotics and toxic compounds is more diverse in CP than in GAP. CP has the capability to produce more proteins related to zinc resistance, and fluoroquinolones, and beta- lactamase, while GAP is more capable to produce genes related to methicillin resistance, and multidrug resistance efflux pumps. Finally, the capability of the GAP metagenome for motility and chemotaxis is enriched compared to CP (42 in GAP, 2 in CP), with GAP being capable to produce factors related to the aforementioned bacterial chemotaxis, flagellar motility, flagellum.
CHAPTER 4
Discussion

The functional potential of the subgingival microbiome, and its shift from health and disease has been examined by several investigators recently. These studies have focused on the functional differences in health and disease, and after treatment. It is assumed from those studies that the diseased state is that of chronic periodontitis. To our knowledge, this is the first study that looks at the functional difference between generalized aggressive periodontitis and health, and as well as comparing the generalized aggressive form to the chronic form of periodontitis.

The shift from health states to diseased states is characterized by a change in the diversity in the genetic capability in the microbiome. This shift results from a gain and loss of certain functions that give the diseased microbiome its virulence factors. The microbiome in chronic periodontitis has been described as a global disease in the mouth, with the metagenome in the shallow and the diseased sites being virtually indistinguishable, and that the disease state is due to the overexpression of certain genes in the metagenome. This investigation reveals that generalized aggressive periodontitis is also a globalized condition in the mouth, where the shallow and deep sites are functionally identical. It remains to be examined whether the difference between the shallow and deep sites is also due to the heightened transcription rate in disease.
The transition of the subgingival microbiome from health to GAP results in a microbiome that begins to organize and isolate itself from the outside stressors, most likely toward tissue and intracellular environments. Intercellular invasion of the microbiome in GAP has been confirmed previously through histology\(^\text{57}\), which undercuts the importance of the microbiome in GCF. Moreover, both A.a. and P.g. have the capability of intracellular invasion of buccal epithelial cells\(^\text{58}\), and it remains to be seen whether such intracellular invasion is possible in the cells of the sulcular epithelium. Investigating the abundance of the membrane secretion pathways (II, V, VI) in GAP could describe the molecules by which the microbiome could invade those cells. The matching of these membrane transport system, as well as the other functional genes, with their species would shed light on what the true function of those virulence associated genes are. Interestingly, it has been found that fimbrial biogenesis can be classified as a membrane secretion system type II.\(^\text{59}\)

As mentioned previously, the microbiome in GAP is also characterized by high capacity for motility, and intracellular invasion. In light of this, the capacity of the outer layer of GAP to resist outside stressors, and may be of lesser importance to the survival of the microbiome. This is evident during the comparison between GAP and CP, in which the capability of the disease to resist outside stressors and create a favorable environment in the GCF compartment is depleted compared to CP.
CHAPTER 5

Conclusion

The results of this study shows the distinct functional differences in the state of health and disease in generalized aggressive periodontitis. These pockets, despite looking clinically healthy, have the same functional capacity as the deeper pockets in those patients. This confirms that, just like the chronic variant, generalized aggressive periodontitis is a globalized condition in the mouth.

These results bring interesting findings, both in terms of our understanding of the pathogenesis of the disease, as well as shedding light on novel methods by which diagnosis, and prognosis of the diseases can be done.
REFERENCES


52. Löe H. The gingival index, the plaque index and the retention index systems. JOP. 1967.


Appendix A  
Demographic Data

Table 1 - Demographic data of the sampled patients

<table>
<thead>
<tr>
<th></th>
<th>GAP</th>
<th>CP</th>
<th>Health</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>22.93 ± 7.5</td>
<td>58 ± 2</td>
<td>56 ± 3</td>
</tr>
<tr>
<td>%Male</td>
<td>52</td>
<td>72</td>
<td>89</td>
</tr>
<tr>
<td>PD</td>
<td>6.9 ± 1.1</td>
<td>6.0 ± 1.5</td>
<td>3.2 ± 1.7</td>
</tr>
<tr>
<td>%BOP</td>
<td>100</td>
<td>100</td>
<td>10 ± 4</td>
</tr>
</tbody>
</table>
### Appendix B

Core Metagenome in diseased sites of GAP

Table 2: Core metagenome in GAP

<table>
<thead>
<tr>
<th>Level 1</th>
<th>Level 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino Acids and Derivatives</td>
<td>Lysine, threonine, methionine, and cysteine</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>Arginine; urea cycle, polyamines</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Aromatic amino acids and derivatives</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Branched-chain amino acids</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Alanine, serine, and glycine</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Glutamine, glutamate, aspartate, asparagine; ammonia assimilation</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Histidine Metabolism</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Proline and 4-hydroxyproline</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Other amino acids and derivatives</td>
<td>5</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Central carbohydrate metabolism</td>
<td>173</td>
</tr>
<tr>
<td></td>
<td>Monosaccharides</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>Di- and oligosaccharides</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Other carbohydrates</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Sugar alcohols</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Fermentation</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>One-carbon Metabolism</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>CO2 fixation</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Organic acids</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Polysaccharides</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Aminosugars</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Glycoside hydrolases</td>
<td>7</td>
</tr>
<tr>
<td>Cell Division and Cell Cycle</td>
<td>Other cell division and cell cycle</td>
<td>46</td>
</tr>
<tr>
<td>Cell Wall and Capsule</td>
<td>Capsular and extracellular polysaccharides</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>Gram-Negative cell wall components</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Other cell division and cell cycle</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Gram-Positive cell wall components</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Cell wall of Mycobacteria</td>
<td>6</td>
</tr>
<tr>
<td>Cofactors, Vitamins, Prosthetic</td>
<td>Folate and pterines</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>Tetrapyrroles</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>NAD and NADP</td>
<td>35</td>
</tr>
</tbody>
</table>

continued
<table>
<thead>
<tr>
<th>Groups, Pigments</th>
<th>Other cofactors, vitamins, prosthetic groups</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Quinone cofactors</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Riboflavin, FMN, FAD</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Biotin</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Coenzyme A</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Pyridoxine</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Lipoic acid</td>
<td>3</td>
</tr>
<tr>
<td>DNA Metabolism</td>
<td>DNA repair</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>DNA replication</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>Other DNA metabolism</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>DNA uptake, competence</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>CRISP</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>DNA recombination</td>
<td>4</td>
</tr>
<tr>
<td>Dormancy and Sporulation</td>
<td>Other dormancy and sporulation</td>
<td>18</td>
</tr>
<tr>
<td>Fatty Acids, Lipids, and Isoprenoids</td>
<td>Fatty acids</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Isoprenoids</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Phospholipids</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Other fatty acids, lipids, and isoprenoids</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Triacylglycerols</td>
<td>5</td>
</tr>
<tr>
<td>Iron acquisition and metabolism</td>
<td>Other iron acquisition and metabolism</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>Siderophores</td>
<td>10</td>
</tr>
<tr>
<td>Membrane Transport</td>
<td>Other membrane transport</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>ABC transporters</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Protein translocation across cytoplasmic membrane</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Protein and nucleoprotein secretion system, Type IV</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Protein secretion system, Type II</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Sugar Phosphotransferase Systems, PTS</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Uni- Sym- and Antiporters</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Protein secretion system, Type VI</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Protein secretion system, Type V</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Protein secretion system, Type I</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Protein secretion system, Type III</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Protein secretion system, Type VIII (Extracellular nucleation/precipitation pathway, ENP)</td>
<td>1</td>
</tr>
<tr>
<td>Metabolism of Aromatic Compounds</td>
<td>Peripheral pathways for catabolism of aromatic compounds</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Metabolism of central aromatic intermediates</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Anaerobic degradation of aromatic compounds</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Other metabolism of aromatic compounds</td>
<td>3</td>
</tr>
</tbody>
</table>

continued
<table>
<thead>
<tr>
<th>Category</th>
<th>Sections</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motility and Chemotaxis</td>
<td>Flagellar motility in Prokaryota</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Other motility and chemotaxis</td>
<td>16</td>
</tr>
<tr>
<td>Nitrogen Metabolism</td>
<td>Other nitrogen metabolism</td>
<td>45</td>
</tr>
<tr>
<td>Nucleosides and Nucleotides</td>
<td>Purines</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Pyrimidines</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Other nucleosides and nucleotides</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Detoxification</td>
<td>14</td>
</tr>
<tr>
<td>Phages, Prophages, Transposable elements, Plasmids</td>
<td>Phages, Prophages</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Transposable elements</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Pathogenicity islands</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Plasmid related functions</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Phosphorus Metabolism</td>
<td>Other phosphorus metabolism</td>
<td>44</td>
</tr>
<tr>
<td>Potassium metabolism</td>
<td>Other Potassium metabolism</td>
<td>22</td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>Protein biosynthesis</td>
<td>161</td>
</tr>
<tr>
<td></td>
<td>Protein degradation</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Protein processing and modification</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>Protein folding</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Selenoproteins</td>
<td>15</td>
</tr>
<tr>
<td>Regulation and Cell signaling</td>
<td>Other regulation and cell signaling</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Programmed Cell Death and Toxin-antitoxin Systems</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Quorum sensing and biofilm formation</td>
<td>10</td>
</tr>
<tr>
<td>Respiration</td>
<td>Regulation of virulence</td>
<td>10</td>
</tr>
<tr>
<td>RNA Metabolism</td>
<td>Electron donating reactions</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Other respiration</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Electron accepting reactions</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>ATP synthases</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Sodium Ion-Coupled Energetics</td>
<td>11</td>
</tr>
<tr>
<td>Secondary Metabolism</td>
<td>RNA processing and modification</td>
<td>248</td>
</tr>
<tr>
<td></td>
<td>Transcription</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Other RNA metabolism</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Biologically active compounds in metazoan cell defence and differentiation</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Aromatic amino acids and derivatives</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Bacterial cytostatics, differentiation factors and antibiotics</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Lipid-derived mediators</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2 continued
<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plant Alkaloids</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Stress Response</strong></td>
<td>142</td>
</tr>
<tr>
<td>Oxidative stress</td>
<td>5</td>
</tr>
<tr>
<td>Heat shock</td>
<td>22</td>
</tr>
<tr>
<td>Osmotic stress</td>
<td>20</td>
</tr>
<tr>
<td>Other stress response</td>
<td>16</td>
</tr>
<tr>
<td>Acid stress</td>
<td>5</td>
</tr>
<tr>
<td>Detoxification</td>
<td>14</td>
</tr>
<tr>
<td>Periplasmic Stress</td>
<td>5</td>
</tr>
<tr>
<td>Cold shock</td>
<td>2</td>
</tr>
<tr>
<td><strong>Sulfur Metabolism</strong></td>
<td>46</td>
</tr>
<tr>
<td>Organic sulfur assimilation</td>
<td>21</td>
</tr>
<tr>
<td>Other sulfur metabolism</td>
<td>16</td>
</tr>
<tr>
<td>Inorganic sulfur assimilation</td>
<td>9</td>
</tr>
<tr>
<td><strong>Virulence, Disease and Defense</strong></td>
<td>143</td>
</tr>
<tr>
<td>Resistance to antibiotics and toxic compounds</td>
<td>77</td>
</tr>
<tr>
<td>Other virulence, disease, and defense</td>
<td>39</td>
</tr>
<tr>
<td>Adhesion</td>
<td>9</td>
</tr>
<tr>
<td>Bacteriocins, ribosomally synthesized antibacterial peptides</td>
<td>6</td>
</tr>
<tr>
<td>Detection</td>
<td>6</td>
</tr>
<tr>
<td>Invasion and intracellular resistance</td>
<td>5</td>
</tr>
<tr>
<td>Toxins and superantigens</td>
<td>1</td>
</tr>
<tr>
<td><strong>Clustering-based subsystems</strong></td>
<td>633</td>
</tr>
<tr>
<td>Other clustering-based subsystems</td>
<td>372</td>
</tr>
<tr>
<td>Clustering-based subsystems</td>
<td>22</td>
</tr>
<tr>
<td>Fatty acid metabolic cluster</td>
<td>21</td>
</tr>
<tr>
<td>Biosynthesis of galactoglycans and related lipopolysaccharides</td>
<td>13</td>
</tr>
<tr>
<td>Cell Division</td>
<td>13</td>
</tr>
<tr>
<td>Cytochrome biogenesis</td>
<td>11</td>
</tr>
<tr>
<td>Probably GTP or GMP signaling related</td>
<td>9</td>
</tr>
<tr>
<td>Protein export?</td>
<td>9</td>
</tr>
<tr>
<td>Two related proteases</td>
<td>9</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>8</td>
</tr>
<tr>
<td>Ribosomal Protein L28P relates to a set of uncharacterized proteins</td>
<td>8</td>
</tr>
<tr>
<td>Translation</td>
<td>8</td>
</tr>
<tr>
<td>Isoprenoid/cell wall biosynthesis: PREDICTED UNDECAPRENYL DIPHOSPHATE PHOSPHATASE</td>
<td>7</td>
</tr>
<tr>
<td>Monosaccharides</td>
<td>126</td>
</tr>
<tr>
<td>Lysine, threonine, methionine, and cysteine</td>
<td>101</td>
</tr>
<tr>
<td>proteosome related</td>
<td>6</td>
</tr>
<tr>
<td>Ribosome-related cluster</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 2 continued
<table>
<thead>
<tr>
<th>Function/Cluster</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>TldD cluster</td>
<td>6</td>
</tr>
<tr>
<td>Methylamine utilization</td>
<td>5</td>
</tr>
<tr>
<td>Nucleotidyl-phosphate metabolic cluster</td>
<td>5</td>
</tr>
<tr>
<td>Oxidative stress</td>
<td>5</td>
</tr>
<tr>
<td>Putative associate of RNA polymerase sigma-54 factor rpoN</td>
<td>5</td>
</tr>
<tr>
<td>Tricarboxylate transporter</td>
<td>5</td>
</tr>
<tr>
<td>CRISPRs and associated hypotheticals</td>
<td>4</td>
</tr>
<tr>
<td>DNA polymerase III epsilon cluster</td>
<td>4</td>
</tr>
<tr>
<td>Hypothetical in Lysine biosynthetic cluster</td>
<td>4</td>
</tr>
<tr>
<td>Hypothetical lipase related to Phosphatidate metabolism</td>
<td>4</td>
</tr>
<tr>
<td>Hypothetical protein possible functionally linked with Alanyl-tRNA synthetase</td>
<td>4</td>
</tr>
<tr>
<td>Carotenoid biosynthesis</td>
<td>3</td>
</tr>
<tr>
<td>Phosphate metabolism</td>
<td>3</td>
</tr>
<tr>
<td>recX and regulatory cluster</td>
<td>3</td>
</tr>
<tr>
<td>Sulfatases and sulfatase modifying factor 1 (and a hypothetical)</td>
<td>3</td>
</tr>
<tr>
<td>Three hypotheticals linked to lipoprotein biosynthesis</td>
<td>3</td>
</tr>
<tr>
<td>tRNA sulfuration</td>
<td>3</td>
</tr>
<tr>
<td>Chemotaxis, response regulators</td>
<td>2</td>
</tr>
<tr>
<td>D-tyrosyl-tRNA(Tyr) deacylase (EC 3.1.-.-) cluster</td>
<td>2</td>
</tr>
<tr>
<td>Hypothetical associated with RecF</td>
<td>2</td>
</tr>
<tr>
<td>Probably organic hydroperoxide resistance related hypothetical protein</td>
<td>2</td>
</tr>
<tr>
<td>Probably Ybbk-related hypothetical membrane proteins</td>
<td>2</td>
</tr>
<tr>
<td>Putative GGDEF domain protein related to agglutinin secretion</td>
<td>2</td>
</tr>
<tr>
<td>Putrescine/GABA utilization cluster-temporal,to add to SSs</td>
<td>2</td>
</tr>
<tr>
<td>Pyruvate kinase associated cluster</td>
<td>2</td>
</tr>
<tr>
<td>Recombination related cluster</td>
<td>2</td>
</tr>
<tr>
<td>Shiga toxin cluster</td>
<td>2</td>
</tr>
<tr>
<td>Choline bitartrate degradation, putative</td>
<td>1</td>
</tr>
<tr>
<td>Chromosome Replication</td>
<td>1</td>
</tr>
<tr>
<td>DNA metabolism</td>
<td>1</td>
</tr>
<tr>
<td>Flagella protein?</td>
<td>1</td>
</tr>
<tr>
<td>Hypothetical Related to Dihydroorate Dehydrogenase</td>
<td>1</td>
</tr>
<tr>
<td>Lysine Biosynthesis</td>
<td>1</td>
</tr>
<tr>
<td>Molybdopterin oxidoreductase</td>
<td>1</td>
</tr>
<tr>
<td>Probably Pyrimidine biosynthesis-related</td>
<td>1</td>
</tr>
<tr>
<td>Category</td>
<td>Count</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>409</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Photosynthesis</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 2 continued
Figure 1 – logfold change in genes between deep sites (GAP) and health. Red dots represent significantly abundant genes (P≥0.05; Wald-test, FDR adjustment)
Figure 2 – PCA comparison between GAP and healthy patients.

Figure 3 – comparison of GAP and healthy patients for the statistically significant genes as a fraction of their total. Only genes occupying 1% or more shown here.
**Appendix D**

**GAP and Health**

Table 3: Composition of statistically significant genes in GAP and Health as a percentage of their whole

<table>
<thead>
<tr>
<th>Level1</th>
<th>GAP AVG</th>
<th>GAP STDEV</th>
<th>Health AVG</th>
<th>Health STDEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbohydrates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central carbohydrate metabolism</td>
<td>0.044844058</td>
<td>0.003970112</td>
<td>0.042508453</td>
<td>0.004375016</td>
</tr>
<tr>
<td>One-carbon Metabolism</td>
<td>0.024113997</td>
<td>0.001603548</td>
<td>0.026060976</td>
<td>0.003891372</td>
</tr>
<tr>
<td>Fermentation</td>
<td>0.019988291</td>
<td>0.002224415</td>
<td>0.01489299</td>
<td>0.002974447</td>
</tr>
<tr>
<td>Monosaccharides</td>
<td>0.014385848</td>
<td>0.00193677</td>
<td>0.012600605</td>
<td>0.00226214</td>
</tr>
<tr>
<td>Di- and oligosaccharides</td>
<td>0.014192072</td>
<td>0.002058469</td>
<td>0.016853442</td>
<td>0.002858538</td>
</tr>
<tr>
<td>CO2 fixation</td>
<td>0.007810726</td>
<td>0.000755776</td>
<td>0.006406819</td>
<td>0.001348671</td>
</tr>
<tr>
<td>Organic acids</td>
<td>0.007634439</td>
<td>0.000755776</td>
<td>0.006406819</td>
<td>0.001348671</td>
</tr>
<tr>
<td>Polysaccharides</td>
<td>0.00609234</td>
<td>0.000932249</td>
<td>0.006406819</td>
<td>0.001348671</td>
</tr>
<tr>
<td>Sugar alcohols</td>
<td>0.005872963</td>
<td>0.001005617</td>
<td>0.006406819</td>
<td>0.001348671</td>
</tr>
<tr>
<td>Aminosugars</td>
<td>0.005061295</td>
<td>0.000927526</td>
<td>0.006406819</td>
<td>0.001348671</td>
</tr>
<tr>
<td>Glycoside hydrolases</td>
<td>0.001481311</td>
<td>0.000340809</td>
<td>0.006406819</td>
<td>0.001348671</td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein biosynthesis</td>
<td>0.085312912</td>
<td>0.005327484</td>
<td>0.10457927</td>
<td>0.010107436</td>
</tr>
<tr>
<td>Protein degradation</td>
<td>0.030011005</td>
<td>0.00166249</td>
<td>0.025073906</td>
<td>0.002690724</td>
</tr>
<tr>
<td>Protein processing and modification</td>
<td>0.015550648</td>
<td>0.001005617</td>
<td>0.006391105</td>
<td>0.001122468</td>
</tr>
<tr>
<td>Protein folding</td>
<td>0.008317158</td>
<td>0.000755776</td>
<td>0.006406819</td>
<td>0.001348671</td>
</tr>
<tr>
<td>Selenoproteins</td>
<td>0.00813191</td>
<td>0.001424451</td>
<td>0.006728577</td>
<td>0.00160069</td>
</tr>
<tr>
<td>RNA Metabolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA processing and modification</td>
<td>0.097489905</td>
<td>0.008195182</td>
<td>0.11290185</td>
<td>0.009901477</td>
</tr>
<tr>
<td>Transcription</td>
<td>0.009766972</td>
<td>0.000594068</td>
<td>0.011504962</td>
<td>0.001419734</td>
</tr>
<tr>
<td>Cofactors, Vitamins, Prosthetic Groups, Pigments</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Folate and pterines</td>
<td>0.026394084</td>
<td>0.000924003</td>
<td>0.027608376</td>
<td>0.001813022</td>
</tr>
<tr>
<td>Tetrapyrroles</td>
<td>0.024234636</td>
<td>0.00131993</td>
<td>0.018390289</td>
<td>0.003150241</td>
</tr>
<tr>
<td>Riboflavin, FMN, FAD</td>
<td>0.007997049</td>
<td>0.001121664</td>
<td>0.007887121</td>
<td>0.001343256</td>
</tr>
<tr>
<td>NAD and NADP</td>
<td>0.007576357</td>
<td>0.000779844</td>
<td>0.007262464</td>
<td>0.001046334</td>
</tr>
<tr>
<td>Biotin</td>
<td>0.005495757</td>
<td>0.00100596</td>
<td>0.005517486</td>
<td>0.001762248</td>
</tr>
<tr>
<td>Coenzyme A</td>
<td>0.004911941</td>
<td>0.000408782</td>
<td>0.005412905</td>
<td>0.000877175</td>
</tr>
<tr>
<td>Pyridoxine</td>
<td>0.003462805</td>
<td>0.000390119</td>
<td>0.00312757</td>
<td>0.000896515</td>
</tr>
<tr>
<td>Quinone cofactors</td>
<td>0.000279345</td>
<td>0.00013456</td>
<td>0.000194049</td>
<td>0.000128981</td>
</tr>
<tr>
<td>Amino Acids and Derivatives</td>
<td>0.02987797</td>
<td>0.003406327</td>
<td>0.025664234</td>
<td>0.003571661</td>
</tr>
<tr>
<td>Glutamine, glutamate, aspartate, asparagine;</td>
<td>0.01197363</td>
<td>0.001255435</td>
<td>0.012825373</td>
<td>0.002664051</td>
</tr>
</tbody>
</table>

44
<table>
<thead>
<tr>
<th>Table 3 continued</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ammonia assimilation</td>
<td>0.009194873</td>
<td>0.000753106</td>
<td>0.006795463</td>
<td>0.001595475</td>
</tr>
<tr>
<td>Aromatic amino acids and derivatives</td>
<td>0.00895808</td>
<td>0.01533849</td>
<td>0.00851892</td>
<td>0.001750576</td>
</tr>
<tr>
<td>Alanine, serine, and glycine</td>
<td>0.007503421</td>
<td>0.01759952</td>
<td>0.004457933</td>
<td>0.001850326</td>
</tr>
<tr>
<td>Branched-chain amino acids</td>
<td>0.004344551</td>
<td>0.00675792</td>
<td>0.00357856</td>
<td>0.000990057</td>
</tr>
<tr>
<td>Arginine; urea cycle, polyamines</td>
<td>0.004188339</td>
<td>0.001210911</td>
<td>0.00260821</td>
<td>0.001013093</td>
</tr>
<tr>
<td>Histidine metabolism</td>
<td>0.004102275</td>
<td>0.00017202</td>
<td>0.00097639</td>
<td>0.000518605</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>0.076125463</td>
<td>0.002708525</td>
<td>0.084677841</td>
<td>0.006324063</td>
</tr>
<tr>
<td>Plant-Prokaryote DOE project</td>
<td>0.01071492</td>
<td>0.001510115</td>
<td>0.01266959</td>
<td>0.001139291</td>
</tr>
<tr>
<td>Fatty acid metabolic cluster</td>
<td>0.004576949</td>
<td>0.00066912</td>
<td>0.005054034</td>
<td>0.001046436</td>
</tr>
<tr>
<td>Cell Division</td>
<td>0.000392463</td>
<td>0.000698197</td>
<td>0.004478691</td>
<td>0.001219779</td>
</tr>
<tr>
<td>Clustering-based subsystems</td>
<td>0.000357738</td>
<td>0.000412499</td>
<td>0.003014254</td>
<td>0.000825664</td>
</tr>
<tr>
<td>Probably GTP or GMP signaling related</td>
<td>0.0003314987</td>
<td>0.000842579</td>
<td>0.00438759</td>
<td>0.000818764</td>
</tr>
<tr>
<td>Isoprenoid/cell wall biosynthesis: predicted undecaprenyl diphosphate phosphatase</td>
<td>0.0003147785</td>
<td>0.000355045</td>
<td>0.004608064</td>
<td>0.000871713</td>
</tr>
<tr>
<td>Hypothetical protein possible functionally linked with alanyl-tRNA synthetase</td>
<td>0.0003147785</td>
<td>0.000355045</td>
<td>0.004608064</td>
<td>0.000871713</td>
</tr>
<tr>
<td>Ribosomal Protein L28P relates to a set of uncharacterized proteins</td>
<td>0.0002568641</td>
<td>0.000274656</td>
<td>0.003708942</td>
<td>0.00057511</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>0.0002300682</td>
<td>0.000463371</td>
<td>0.002862629</td>
<td>0.000940578</td>
</tr>
<tr>
<td>Lysine, threonine, methionine, and cysteine</td>
<td>0.0002220197</td>
<td>0.000254432</td>
<td>0.002471574</td>
<td>0.000763966</td>
</tr>
<tr>
<td>Shiga toxin cluster</td>
<td>0.0001959703</td>
<td>0.000367025</td>
<td>0.001588339</td>
<td>0.000663688</td>
</tr>
<tr>
<td>Translation</td>
<td>0.000201549</td>
<td>0.00077505</td>
<td>0.000372676</td>
<td>0.00094178</td>
</tr>
<tr>
<td>Oxidative stress</td>
<td>0.0001912799</td>
<td>0.000460107</td>
<td>0.001918503</td>
<td>0.001215986</td>
</tr>
<tr>
<td>Biosynthesis of galactoglycans and related lipopolysaccharides</td>
<td>0.0001864671</td>
<td>0.000377591</td>
<td>0.001439033</td>
<td>0.000434377</td>
</tr>
<tr>
<td>Recombination related cluster</td>
<td>0.0001733824</td>
<td>0.000527181</td>
<td>0.001131828</td>
<td>0.000799455</td>
</tr>
<tr>
<td>Nucleotidylylphosphate metabolic cluster</td>
<td>0.000170794</td>
<td>0.000731023</td>
<td>0.001065438</td>
<td>0.000463151</td>
</tr>
<tr>
<td>Tricarboxylate transporter</td>
<td>0.0001700011</td>
<td>0.000263534</td>
<td>0.001541607</td>
<td>0.000472263</td>
</tr>
<tr>
<td>Putative associate of RNA polymerase sigma-54 factor rpoN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein export</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Clusters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA Metabolism</td>
<td>DNA repair: 0.026979259, 0.001660073, 0.029487881, 0.004155881</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DNA replication: 0.018505303, 0.001667179, 0.021986398, 0.0002239318</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CRISPRs: 0.001786622, 0.000372756, 0.001763933, 0.000478108</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DNA uptake, competence: 0.001177692, 0.000213462, 0.00073876, 0.000330383</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DNA recombination: 0.000526265, 9.75123E-05, 0.000657016, 0.00033193</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA replication</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRISPRs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA uptake, competence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA recombination</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carotenoid biosynthesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRISPRs and associated hypotheticals</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA polymerase III epsilon cluster</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-tyrosyl-tRNA(Tyr) deacylase (EC 3.1.2.2) cluster</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two related proteases</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytochrome biogenesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monosaccharides</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flagella protein?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sulfatases and sulfatase modifying factor 1 (and a hypothetical)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methylamine utilization</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sarcosine oxidase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>recX and regulatory cluster</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyruvate kinase associated cluster</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probably Ybbk-related hypothetical membrane proteins</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypothetical in Lysine biosynthetic cluster</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypothetical lipase related to Phosphatidate metabolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemotaxis, response regulators</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Putative GGDEF domain protein related to agglutinin secretion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypothetical in Lysine biosynthetic cluster</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Table 3 continued</strong></td>
<td>Gram-Negative cell wall components</td>
<td>0.021624274</td>
<td>0.003235724</td>
<td>0.014704217</td>
</tr>
<tr>
<td></td>
<td>Capsular and extracellular polysaccharides</td>
<td>0.01061627</td>
<td>0.001161904</td>
<td>0.007476728</td>
</tr>
<tr>
<td><strong>Capsule and Capsular and extracellular</strong></td>
<td>Cell wall of Mycobacteria</td>
<td>0.004103474</td>
<td>0.000607401</td>
<td>0.005194657</td>
</tr>
<tr>
<td><strong>Gram-Negative cell wall components</strong></td>
<td>Gram-Positive cell wall components</td>
<td>0.002259662</td>
<td>0.00032047</td>
<td>0.001680193</td>
</tr>
<tr>
<td><strong>0.015989444</strong></td>
<td>0.020033559</td>
<td>0.018446111</td>
<td>0.003544269</td>
<td></td>
</tr>
<tr>
<td><strong>0.014747781</strong></td>
<td>0.01798662</td>
<td>0.015024886</td>
<td>0.001930453</td>
<td></td>
</tr>
<tr>
<td><strong>0.003030232</strong></td>
<td>0.00400685</td>
<td>0.003031151</td>
<td>0.000727639</td>
<td></td>
</tr>
<tr>
<td><strong>0.004103474</strong></td>
<td>0.001161904</td>
<td>0.007476728</td>
<td>0.002035542</td>
<td></td>
</tr>
<tr>
<td><strong>0.004103474</strong></td>
<td>0.000607401</td>
<td>0.005194657</td>
<td>0.000689289</td>
<td></td>
</tr>
<tr>
<td><strong>Fatty Acids, Lipids, and Isoprenoids</strong></td>
<td><strong>Fatty acids</strong></td>
<td>0.015989444</td>
<td>0.020033559</td>
<td>0.018446111</td>
</tr>
<tr>
<td></td>
<td>Isoprenoids</td>
<td>0.014747781</td>
<td>0.01798662</td>
<td>0.015024886</td>
</tr>
<tr>
<td></td>
<td>Phospholipids</td>
<td>0.003030232</td>
<td>0.00400685</td>
<td>0.003031151</td>
</tr>
<tr>
<td><strong>Cell wall of Mycobacteria</strong></td>
<td><strong>Gram-Positive cell wall components</strong></td>
<td>0.002259662</td>
<td>0.00032047</td>
<td>0.001680193</td>
</tr>
<tr>
<td><strong>0.004103474</strong></td>
<td>0.001161904</td>
<td>0.007476728</td>
<td>0.002035542</td>
<td></td>
</tr>
<tr>
<td><strong>0.004103474</strong></td>
<td>0.000607401</td>
<td>0.005194657</td>
<td>0.000689289</td>
<td></td>
</tr>
<tr>
<td><strong>Fatty Acids, Lipids, and Isoprenoids</strong></td>
<td><strong>Fatty acids</strong></td>
<td>0.015989444</td>
<td>0.020033559</td>
<td>0.018446111</td>
</tr>
<tr>
<td></td>
<td>Isoprenoids</td>
<td>0.014747781</td>
<td>0.01798662</td>
<td>0.015024886</td>
</tr>
<tr>
<td></td>
<td>Phospholipids</td>
<td>0.003030232</td>
<td>0.00400685</td>
<td>0.003031151</td>
</tr>
<tr>
<td><strong>Nucleosides and Nucleotides</strong></td>
<td>Purines</td>
<td>0.015676226</td>
<td>0.001373682</td>
<td>0.01863873</td>
</tr>
<tr>
<td></td>
<td>Pyrimidines</td>
<td>0.011598502</td>
<td>0.000866251</td>
<td>0.011300277</td>
</tr>
<tr>
<td></td>
<td>Detoxification</td>
<td>0.002446628</td>
<td>0.000243222</td>
<td>0.002645379</td>
</tr>
<tr>
<td><strong>Membrane Transport</strong></td>
<td>ABC transporters</td>
<td>0.009573092</td>
<td>0.001940831</td>
<td>0.008297966</td>
</tr>
<tr>
<td></td>
<td>Protein translocation across cytoplasmic membrane</td>
<td>0.006745742</td>
<td>0.000647912</td>
<td>0.007602533</td>
</tr>
<tr>
<td></td>
<td>Protein secretion system, Type VI</td>
<td>0.004133769</td>
<td>0.00073233</td>
<td>0.005104805</td>
</tr>
<tr>
<td></td>
<td>Protein secretion system, Type V</td>
<td>0.001975117</td>
<td>0.00125336</td>
<td>0.001577343</td>
</tr>
<tr>
<td></td>
<td>Protein secretion system, Type II</td>
<td>0.001399741</td>
<td>0.000257342</td>
<td>0.001422142</td>
</tr>
<tr>
<td></td>
<td>Uni-Sym- and Antiporters</td>
<td>0.001018326</td>
<td>0.000208007</td>
<td>0.000659921</td>
</tr>
<tr>
<td></td>
<td>Protein secretion system, Type VIII</td>
<td>0.000106762</td>
<td>5.45848E-05</td>
<td>8.19249E-05</td>
</tr>
<tr>
<td></td>
<td>(Extracellular nucleation/precipitation pathway, ENP)</td>
<td>7.59602E-05</td>
<td>6.59455E-05</td>
<td>1.32614E-05</td>
</tr>
<tr>
<td></td>
<td>Protein and nucleoprotein secretion system, Type IV</td>
<td>0.001482428</td>
<td>0.000809105</td>
<td>0.008760606</td>
</tr>
<tr>
<td>Phages, Prophages, Transposable elements, Plasmids</td>
<td>Phages, Prophages</td>
<td>0.010095552</td>
<td>0.00124057</td>
<td>0.011347872</td>
</tr>
<tr>
<td></td>
<td>Transposable elements</td>
<td>0.008781031</td>
<td>0.003047399</td>
<td>0.007065314</td>
</tr>
<tr>
<td></td>
<td>Pathogenicity islands</td>
<td>0.002765709</td>
<td>0.000438444</td>
<td>0.00361287</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>0.00019587</td>
<td>8.3035E-05</td>
<td>3.85911E-05</td>
</tr>
<tr>
<td>Motility and Chemotaxis</td>
<td>Flagellar motility in Prokaryota</td>
<td>0.02120679</td>
<td>0.008809105</td>
<td>0.008760606</td>
</tr>
<tr>
<td>Respiration</td>
<td>Electron donating reactions</td>
<td>0.007731682</td>
<td>0.000735014</td>
<td>0.004712614</td>
</tr>
<tr>
<td></td>
<td>ATP synthases</td>
<td>0.00879688</td>
<td>0.001416407</td>
<td>0.003105329</td>
</tr>
<tr>
<td></td>
<td>Electron accepting reactions</td>
<td>0.004255997</td>
<td>0.000382361</td>
<td>0.004006055</td>
</tr>
<tr>
<td></td>
<td>Sodium Ion-Coupled Energetics</td>
<td>0.003209521</td>
<td>0.000836236</td>
<td>0.00173872</td>
</tr>
<tr>
<td>Virulence, Disease and Defense</td>
<td>Resistance to antibiotics and toxic compounds</td>
<td>0.015575265</td>
<td>0.00316478</td>
<td>0.011116091</td>
</tr>
<tr>
<td></td>
<td>Detection</td>
<td>0.001482428</td>
<td>0.00273957</td>
<td>0.001749597</td>
</tr>
<tr>
<td></td>
<td>Adhesion</td>
<td>0.00088689</td>
<td>0.00183578</td>
<td>0.001078551</td>
</tr>
<tr>
<td></td>
<td>Bacteriocins, ribosomally</td>
<td>0.000525629</td>
<td>0.00034621</td>
<td>0.000303432</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------------------------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td><strong>synthesized antibacterial peptides</strong></td>
<td>2.47586E-05</td>
<td>2.96503E-05</td>
<td>8.31306E-06</td>
<td>2.06734E-05</td>
</tr>
<tr>
<td><strong>Invasion and intracellular resistance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stress Response</strong></td>
<td>0.007715306</td>
<td>0.000832054</td>
<td>0.006847427</td>
<td>0.001702424</td>
</tr>
<tr>
<td>Oxidative stress</td>
<td>0.005245206</td>
<td>0.00693301</td>
<td>0.005103509</td>
<td>0.000558209</td>
</tr>
<tr>
<td>Heat shock</td>
<td>0.001087627</td>
<td>0.00336429</td>
<td>0.000452542</td>
<td>0.000338038</td>
</tr>
<tr>
<td>Periplasmic Stress</td>
<td>0.000433775</td>
<td>0.00222965</td>
<td>0.000283195</td>
<td>0.000175665</td>
</tr>
<tr>
<td>Osmotic stress</td>
<td>4.87474E-05</td>
<td>4.29777E-05</td>
<td>2.61143E-05</td>
<td>4.24376E-05</td>
</tr>
<tr>
<td>Acid stress</td>
<td>3.19721E-05</td>
<td>2.08644E-05</td>
<td>1.31418E-05</td>
<td>2.37941E-05</td>
</tr>
<tr>
<td><strong>Detoxification</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Metabolism of Aromatic Compounds</strong></td>
<td>0.003635265</td>
<td>0.000577297</td>
<td>0.004513986</td>
<td>0.001785778</td>
</tr>
<tr>
<td>Peripheral pathways for catabolism of aromatic compounds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anaerobic degradation of aromatic compounds</td>
<td>0.001883113</td>
<td>0.00366801</td>
<td>0.001006804</td>
<td>0.000590269</td>
</tr>
<tr>
<td>Metabolism of central aromatic intermediates</td>
<td>0.000502541</td>
<td>0.00195357</td>
<td>0.000305844</td>
<td>0.000169873</td>
</tr>
<tr>
<td><strong>Regulation and Cell signaling</strong></td>
<td>0.003091922</td>
<td>0.000394667</td>
<td>0.004052434</td>
<td>0.0009115</td>
</tr>
<tr>
<td>Regulation of virulence</td>
<td>0.002146904</td>
<td>0.00542855</td>
<td>0.002768609</td>
<td>0.000479538</td>
</tr>
<tr>
<td>Quorum sensing and biofilm formation</td>
<td>0.000728094</td>
<td>0.00026332</td>
<td>0.000717803</td>
<td>0.000352078</td>
</tr>
<tr>
<td>Programmed Cell Death and Toxin-antitoxin Systems</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sulfur Metabolism</strong></td>
<td>0.000704346</td>
<td>0.000201049</td>
<td>0.000443267</td>
<td>0.000308787</td>
</tr>
<tr>
<td>Organic sulfur assimilation</td>
<td>4.50508E-05</td>
<td>3.58173E-05</td>
<td>5.13531E-06</td>
<td>2.11229E-05</td>
</tr>
<tr>
<td>Inorganic sulfur assimilation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Secondary Metabolism</strong></td>
<td>0.000646517</td>
<td>0.000174183</td>
<td>0.000365623</td>
<td>0.000269666</td>
</tr>
<tr>
<td>Biologically active compounds in metazoan cell defence and differentiation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Iron acquisition and metabolism</strong></td>
<td>0.00025362</td>
<td>0.000118315</td>
<td>0.000126565</td>
<td>0.000156075</td>
</tr>
</tbody>
</table>
Appendix E
Comparisons between statistically significant genes GAP and Health

Figure 4 – Carbohydrates - Functional capability of GAP vs Health
Figure 5 – Membrane Transport - Functional capability of GAP vs Health

Figure 6 – Phages, prophages, transposable elements, plasmids - Functional capability of GAP vs Health
Appendix F
Comparison of Deep and Shallow Sites in GAP

Figure 7 – logfold change in genes between deep and shallow sites of GAP.
Figure 8 – PCA comparison between deep and shallow sites of GAP.
Appendix G

Comparison Between deep sites in generalized aggressive periodontitis (GAP) and chronic periodontitis (CP)

Figure 9 – logfold change in genes between GAP and CP. Red dots represent significantly abundant genes (P≥0.05; Wald-test, FDR adjustment)
Figure 10 – PCA comparison between GAP and CP

Figure 11 – comparison of GAP and CP patients for the statistically significant genes as a fraction of their total. Only genes occupying 1% or more shown here
### Appendix H
#### GAP and CP

Table 4: Composition of statistically significant genes in GAP and CP as a percentage of their whole level

<table>
<thead>
<tr>
<th>level1</th>
<th>level2</th>
<th>GAP AVG</th>
<th>GAP STDEV</th>
<th>CP AVG</th>
<th>CP STDEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino Acids and Derivatives</td>
<td>Alanine, serine, and glycine</td>
<td>0.016811781</td>
<td>0.002551866</td>
<td>0.017957558</td>
<td>0.00790862</td>
</tr>
<tr>
<td></td>
<td>Arginine; urea cycle, polyamines</td>
<td>0.002994924</td>
<td>0.001309775</td>
<td>0.00146714</td>
<td>0.000862366</td>
</tr>
<tr>
<td></td>
<td>Glutamine, glutamate, aspartate, asparagine;</td>
<td>0.027380124</td>
<td>0.00342813</td>
<td>0.032786306</td>
<td>0.010067709</td>
</tr>
<tr>
<td></td>
<td>ammonia assimilation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Proline and 4-hydroxyproline</td>
<td>0.000494799</td>
<td>0.000285857</td>
<td>0.000205757</td>
<td>0.000246</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>Aminosugars</td>
<td>0.008823208</td>
<td>0.001753347</td>
<td>0.009286076</td>
<td>0.00337454</td>
</tr>
<tr>
<td></td>
<td>CO2 fixation</td>
<td>0.019331364</td>
<td>0.004169874</td>
<td>0.022196035</td>
<td>0.006340913</td>
</tr>
<tr>
<td></td>
<td>Central carbohydrate metabolism</td>
<td>0.112919407</td>
<td>0.019285304</td>
<td>0.117162934</td>
<td>0.024697852</td>
</tr>
<tr>
<td></td>
<td>Di- and oligosaccharides</td>
<td>0.01014923</td>
<td>0.004304208</td>
<td>0.007899753</td>
<td>0.00481574</td>
</tr>
<tr>
<td></td>
<td>Fermentation</td>
<td>0.0107546</td>
<td>0.00252022</td>
<td>0.014391671</td>
<td>0.00734407</td>
</tr>
<tr>
<td></td>
<td>Glycoside hydrolases</td>
<td>0.001496555</td>
<td>0.000795289</td>
<td>0.00137072</td>
<td>0.001441574</td>
</tr>
<tr>
<td></td>
<td>Monosaccharides</td>
<td>0.03823996</td>
<td>0.003469442</td>
<td>0.038689722</td>
<td>0.011020533</td>
</tr>
<tr>
<td></td>
<td>One-carbon Metabolism</td>
<td>0.003708689</td>
<td>0.001927946</td>
<td>0.003641143</td>
<td>0.002304363</td>
</tr>
<tr>
<td></td>
<td>Organic acids</td>
<td>0.003708689</td>
<td>0.001927946</td>
<td>0.003641143</td>
<td>0.002304363</td>
</tr>
<tr>
<td></td>
<td>Polysaccharides</td>
<td>0.016652365</td>
<td>0.003601603</td>
<td>0.015451392</td>
<td>0.005054959</td>
</tr>
<tr>
<td></td>
<td>Sugar alcohols</td>
<td>0.010261465</td>
<td>0.004874919</td>
<td>0.007699423</td>
<td>0.003412058</td>
</tr>
<tr>
<td>Cell Wall and Capsule</td>
<td>Capsular and extracellular polysaccharides</td>
<td>0.025817302</td>
<td>0.003787072</td>
<td>0.027960525</td>
<td>0.007716975</td>
</tr>
<tr>
<td></td>
<td>Gram-Negative cell wall components</td>
<td>0.061490558</td>
<td>0.006856763</td>
<td>0.062749194</td>
<td>0.015237252</td>
</tr>
<tr>
<td></td>
<td>Gram-Positive cell wall components</td>
<td>0.011846134</td>
<td>0.001995115</td>
<td>0.012320872</td>
<td>0.003402864</td>
</tr>
<tr>
<td>Clustering-based subsystems</td>
<td>Biosynthesis of galactoglycans and related</td>
<td>0.011936844</td>
<td>0.002473013</td>
<td>0.015237226</td>
<td>0.006331885</td>
</tr>
<tr>
<td></td>
<td>lipopolysaccharides</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Flagella protein?</td>
<td>0.002362247</td>
<td>0.001215368</td>
<td>0.001919129</td>
<td>0.001498986</td>
</tr>
<tr>
<td></td>
<td>Isoprenoid/Cell Wall Biosynthesis: Predicted</td>
<td>0.008644211</td>
<td>0.001588574</td>
<td>0.009287494</td>
<td>0.00293061</td>
</tr>
<tr>
<td></td>
<td>Under Diphosphate Phosphatase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Methylamine utilization</td>
<td>0.001329569</td>
<td>0.000722251</td>
<td>0.000900077</td>
<td>0.000841129</td>
</tr>
<tr>
<td></td>
<td>Monosaccharides</td>
<td>0.00339391</td>
<td>0.000573355</td>
<td>0.002349261</td>
<td>0.00167521</td>
</tr>
<tr>
<td></td>
<td>Nucleotide-l-phosphate metabolic cluster</td>
<td>0.007967289</td>
<td>0.002420347</td>
<td>0.004123584</td>
<td>0.003533657</td>
</tr>
<tr>
<td></td>
<td>Oxidative stress</td>
<td>0.001222253</td>
<td>0.000608455</td>
<td>0.000774187</td>
<td>0.000761161</td>
</tr>
<tr>
<td></td>
<td>Probably Ybbk-related hypothetical membrane</td>
<td>0.006245956</td>
<td>0.001535779</td>
<td>0.007921814</td>
<td>0.003058107</td>
</tr>
<tr>
<td></td>
<td>proteins</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pyruvate kinase associated cluster</td>
<td>0.001274076</td>
<td>0.000731821</td>
<td>0.001127575</td>
<td>0.00198221</td>
</tr>
<tr>
<td></td>
<td>Sarcosine oxidase</td>
<td>0.001329569</td>
<td>0.000722251</td>
<td>0.000900077</td>
<td>0.000841129</td>
</tr>
<tr>
<td></td>
<td>Two related proteases</td>
<td>0.00333991</td>
<td>0.000573355</td>
<td>0.002349261</td>
<td>0.001675721</td>
</tr>
<tr>
<td>Metabolism Category</td>
<td>Example Reactions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cofactors, Vitamins, Prosthetic Groups, Pigments</td>
<td>Biotin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coenzyme A</td>
<td>0.022412161 0.004968859 0.026659132 0.007551213</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Folate and pterines</td>
<td>0.002493842 0.002149406 0.002169685 0.001516988</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAD and NADP</td>
<td>0.010176162 0.002642448 0.010122399 0.004626405</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetrapyrroles</td>
<td>0.007368583 0.002676761 0.007535568 0.003345923</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA Metabolism</td>
<td>DNA repair</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.024734015 0.002788197 0.026930177 0.007840287</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA replication</td>
<td>0.037804116 0.009820986 0.045326213 0.012094898</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatty Acids, Lipids, and Isoprenoids</td>
<td>Fatty acids</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.051110388 0.012819317 0.058459793 0.015722662</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isoprenoids</td>
<td>0.002057585 0.000893951 0.001467604 0.001439486</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phospholipids</td>
<td>0.009926292 0.00258945 0.010254213 0.003192975</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Membrane Transport</td>
<td>ABC transporters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.001316719 0.000491111 0.001058022 0.001165975</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Protein translocation across cytoplasmic membrane</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.015608357 0.002479108 0.019971265 0.004616321</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metabolism of Aromatic Compounds</td>
<td>Peripheral pathways for catabolism of aromatic compounds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.022412161 0.004968859 0.026659132 0.007551213</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>Plant-Prokaryote DOE project</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.026182749 0.003720247 0.0194559 0.003502271</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motility and Chemotaxis</td>
<td>Flagellar motility in Prokaryota</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.092899263 0.03674605 0.067483554 0.029712652</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nucleosides and Nucleotides</td>
<td>Purines</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.017076271 0.00355883 0.016238503 0.005697313</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pyrimidines</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.026715262 0.004111198 0.029892828 0.007776249</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phages, Prophages, Transposable elements, Plasmids</td>
<td>Phages, Prophages</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.000878286 0.000564623 0.000171477 0.000279514</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transposable elements</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.003571297 0.00136222 0.002794901 0.002086374</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.001685765 0.001293703 0.000954185 0.001005108</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein Metabolism</td>
<td>Protein biosynthesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.010995244 0.001057812 0.008827522 0.004305189</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Protein degradation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.019534178 0.003753906 0.015807317 0.00343408</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Protein folding</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.006166044 0.002607962 0.005702778 0.00474944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Protein processing and modification</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.003571538 0.000947042 0.003322424 0.00277853</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Selenoproteins</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.014082815 0.003166607 0.012655155 0.005527843</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA Metabolism</td>
<td>RNA processing and modification</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.033049782 0.004924277 0.035454216 0.007766228</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transcription</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.010173089 0.002291217 0.009748663 0.003653734</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulation and Cell signaling</td>
<td>Programmed Cell Death and Toxin-antitoxin Systems</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.001409828 0.000369675 0.000944248 0.000779902</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quorum sensing and biofilm formation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.000296064 0.000555148 0.000183885 0.000574745</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Regulation of virulence</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.006600152 0.001575887 0.00542432 0.002504768</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiration</td>
<td>Electron accepting reactions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.007724947 0.0042065 0.008233325 0.004345862</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Electron donating reactions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.024603275 0.00457795 0.024181997 0.00652924</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sodium Ion-Coupled Energetics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.002853538 0.000892881 0.002861867 0.001621448</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stress Response</td>
<td>Osmotic stress</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.001329569 0.000722251 0.000900077 0.00841129</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oxidative stress</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.004600892 0.002182701 0.002266443 0.002071781</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Periplasmic Stress</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.005953193 0.001875261 0.00598064 0.004041784</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sulfur Metabolism</td>
<td>Organic sulfur assimilation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.000469654 0.000269269 0.000368973 0.000696528</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4 continued
Table 4 continued

<table>
<thead>
<tr>
<th>Virulence, Disease and Defense</th>
<th>Resistance to antibiotics and toxic compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.037538612</td>
</tr>
</tbody>
</table>