AIX-MARSEILLE UNIVERSITE

FACULTE DE MEDECINE DE MARSEILLE

ECOLE DOCTORALE : SCIENCE DE LA VIE ET DE LA SANTE

THESE

Présentée et publiquement soutenue devant
LA FACULTE DE MEDECINE DE MARSEILLE

Le 15 janvier 2016

Par Mme Sandrine PINOS

Née à Saint-Gaudens le 09 octobre 1989

TITRE DE LA THESE:

Evolution génomique chez les bactéries du super phylum Planctomycetes-Verrucomicrobiae-Chlamydia

Pour obtenir le grade de DOCTORAT d'AIX-MARSEILLE UNIVERSITE
Spécialité : Génomique et Bioinformatique

Membres du jury de la Thèse:

Pr Didier RAOULT .....Directeur de thèse
Dr Pierre PONTAROTTI .....Co-directeur de thèse
Pr Gilbert GREUB .....Rapporteur
Dr Pascal SIMONET.....Rapporteur

Laboratoires d'accueil
Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes – UMR CNRS 6236, IRD 198
I2M - UMR CNRS 7373 - EBM

Avant propos

Le format de présentation de cette thèse correspond à une recommandation de la spécialité Maladies Infectieuses et Microbiologie, à l'intérieur du Master de Sciences de la Vie et de la Santé qui dépend de l'Ecole Doctorale des Sciences de la Vie de Marseille. Le candidat est amené à respecter des règles qui lui sont imposées et qui comportent un format de thèse utilisé dans le Nord de l'Europe permettant un meilleur rangement que les thèses traditionnelles. Par ailleurs, la partie introduction et bibliographie est remplacée par une revue envoyée dans un journal afin de permettre une évaluation extérieure de la qualité de la revue et de permettre à l'étudiant de le commencer le plus tôt possible une bibliographie exhaustive sur le domaine de cette thèse. Par ailleurs, la thèse est présentée sur article publié, accepté ou soumis associé d'un bref commentaire donnant le sens général du travail. Cette forme de présentation a paru plus en adéquation avec les exigences de la compétition internationale et permet de se concentrer sur des travaux qui bénéficieront d'une diffusion internationale.

Professeur Didier RAOULT

Sommaire

Résumé

La compréhension de l'évolution des génomes est un des enjeux clé de la biologie actuelle. Cela induit l'identification des événements génétiques affectant les génomes au cours du temps, mais aussi la détermination des relations entre les génomes, les phénotypes et l'environnement des organismes.

Nous avons rédigé une revue de la littérature consacrée à la contribution de la génomique dans la compréhension de la diversité, de l'évolution et des causes génétiques de phénotypes d'un super-phylum bactérien appelé super-phylum PVC (pour Planctomycetes, Verrucomicrobiae et Chlamydiae). Ces bactéries proviennent d'environnements variés et présentent des caractéristiques phénotypiques particulières, dont certaines ne sont retrouvées que dans ce groupe. Les analyses génomiques ont révélé la grande diversité de ces espèces, mais ont aussi permis de reconstruire l'évolution de leurs génomes et dans certains cas, d'expliquer l'apparition de phénotypes particuliers. Plusieurs caractéristiques morphologiques ou physiologiques restent cependant mal expliquées, que ce soit leur origine ou

leur évolution, entre autre à cause du manque de données disponibles. Une partie de notre travail était consacré à l'étude de l'évolution et de l'impact de la présence d'un plan cellulaire particulier chez les bactéries PVC. Ce plan cellulaire est sujet a différentes interprétations et induirait la compartimentation des cellules en deux régions distinctes dont l'une contiendrait le nucléoside. Les résultats obtenus semblent indiquer que cette caractéristique n'induit pas une protection des génomes bactériens vis à vis des transferts de gènes horizontaux, comme on pourrait le supposer, sa fonction reste donc pour l'instant indéfinie. En revanche les observations microscopiques réalisées sur deux espèces ont permis de mieux appréhender l'évolution de ce plan cellulaire. Nous avons, de plus, détecté une contribution de l'environnement concernant la sélection des gènes transférés. Il semblerait que les gènes transférés soient en effet sélectionnés selon leurs fonctions par les différents environnements.

Nos travaux ont donc permis d'améliorer la compréhension des relations entre l'évolution, les phénotypes et l'environnement, en particulier chez les bactéries du super-phylum PVC, fournissant ainsi de nombreuses pistes de travail qui pourront être approfondies dans le futur.

Abstract

The comprehension of genomes evolution is a key issue of modern biology. This induces the identification of genetic events occurred in history of genomes and also the determination of relations between the genomes, phenotypes and environments. We wrote a review dedicated to the genomic contribution in comprehension of diversity, evolution and genetic causes of phenotypic features, in a bacterial super-phylum named PVC (for Planctomycetes, Verrucomicrobiae and Chlamydiae). These bacteria are distributed in varied environments and present specific phenotypic characteristics, whom some of them are identified only in this group. Genomic analyzes revealed the important diversity of these species and allow also to reconstruct the genomes evolution and, in some cases, to explain the presence of specific phenotypes. Due to the lack of information, it is difficult to define the origins and evolution of all specific phenotypes. One part of our work was dedicated to the study of evolution and impact of one of this phenotype, the special cell plan detected in PVC bacteria. This original cell plan is subject to different interpretations and induces the compartmentalization of cells in two different regions, whom one containing

the nucleoid. Our results indicate that this feature has probably no role in the protection of bacterial genomes against horizontal genes transfers, so, its function is still unknown. Microscopic observations of two species from PVC super-phylum permit to better understand the evolution of the special cell plan. The environment seems to contribute in the genomes evolution, by selection of genes transferred. Genes transferred are probably selected according to their functions by the different environments.

Our works allowed to improve the knowledge about relations between evolution, genomes, phenotypes and environment, especially in bacteria from PVC super-phylum, providing new information for future works.

Introduction

La génomique est aujourd'hui un domaine en pleine expansion, fortement favorisée par des méthodes de séquençage à haut débit de plus en plus perfectionnées, elle donne accès à un nombre croissant de données concernant l'ensemble des génomes. L'exploitation de toutes ces informations nécessite le développement de méthodes d'analyse et de comparaison des génomes diverses et variées qui révèlent l'impressionnante diversité des espèces et souches bactériennes (1-3). Cette diversité au sein des bactéries est liée à la grande plasticité de leurs génomes, fruits des combinaisons complexes entre de multiples événements génétiques ayant eu lieu au cours du temps. En effet, chaque génome est une mosaïque sans cesse réorganisée de différentes d'ADN transmis verticalement mais aussi horizontalement, perdus ou dupliqués (4). L'identification des événements génétiques est donc un enjeu important dans la compréhension de la complexité des génomes. La reconstruction de l'évolution des génomes bactériens qui en découle, permet non seulement de comprendre d'où proviennent ces génomes mais aussi d'identifier les causes de

caractéristiques physiologiques ou morphologiques observées chez les bactéries.

Au début des années 2000, de nombreuses bactéries ou phyla restaient encore mal connues du fait de la difficulté de les cultiver, la génomique a donc permis l'étude de ces différentes espèces. Un grand nombre de bactéries du super-phylum PVC (*Planctomycetes, Verrucomicrobiae* et *Chlamydiae*) ont donc été révélées ces dernières années, telles que de nombreuses *Verrucomicrobiae* (5-7) ou les phyla mineurs *Lentisphaera, Poribacteria, WWE2* ou *Omnitrophica* (8) qui n'ont attisés que très récemment l'intérêt des chercheurs. Ces données ont ainsi permis de mieux appréhender l'évolution complexe de ce super-phylum. En effet, les bactéries PVC, assez variées au niveau des habitats et des modes de vies, présentent des caractéristiques phénotypiques particulières, dont quelques une sont spécifiques à ce super-phylum (9-11). L'évolution des génomes est donc un élément clé pour mieux comprendre la présence de ces différentes caractéristiques.

L'objectif principal de ce travail de thèse était donc de mieux appréhender l'évolution des génomes et de déterminer leurs relations avec les caractéristiques phénotypiques et les environnements des bactéries. Nous nous sommes intéressés en particulier au super-phylum PVC puisqu'il comporte des bactéries aux environnements et aux phénotypes variées, ce qui en fait un bon modèle pour l'étude de l'évolution des génomes. Dans un premier temps, la revue présente les différents travaux de génomiques consacrés à ces bactéries. Ces travaux ont démontré la grande diversité des bactéries de ce super-phylum, menant au séquençage de 240 génomes (principalement *Planctomycetes, Verrucomicrobiae* et *Chlamydiae*) mais aussi à l'identification de nombreux fragments de séquences provenant de bactéries encore non caractérisées dans les banques de données. Quatre publications ont été consacrées à la détection des événements génétiques, et plus particulièrement aux transferts latéraux de matériel génétiques, chez les bactéries PVC. La synthèse des informations obtenues permet ainsi de dessiner l'évolution du super-phylum. La génomique apporte aussi des informations intéressantes concernant les origines et l’évolution des caractéristiques phénotypiques particulières des bactéries du super-phylum

PVC, en particulier a propos de la présence d'une hypothétique membrane intra-cytoplasmique chez certaines bactéries. Cette hypothétique membrane intra-cytoplasmique, identifiée chez les *Planctomycetes*, certaines *Verrucomicrobiae*, une *Lentisphaera* et les *Poribacteria* (12-15), entrainerait la compartimentation des bactéries en deux régions distinctes dont l'une contiendrait le nucléoside. La compartimentation de ces bactéries est en fait le thème principal du premier article de cette thèse. La question testée au cours de ce travail étant de savoir si la présence d'une compartimentation chez les bactéries protège leur ADN contre les transferts de gènes, et si oui par quels moyens. Il s'agit ici non plus de déterminer la cause d'une caractéristique phénotypique mais au contraire d'observer les possibles conséquences d'un phénotype sur l'évolution des génomes. Des observations microscopiques ont donc été réalisées sur les bactéries dont le plan cellulaire était inconnu ; Cela a permis de reconstituer de manière plus précise l'évolution de ce caractère au fil du temps. Une méthode d'analyse phylogénomique a ensuite était mise en place afin de détecter les transferts horizontaux de matériel génétique chez les bactéries PVC, particulièrement celles présentant une membrane intra-cytoplasmique. Des bactéries

phylogéntiquement proches ont aussi été analysées afin de pouvoir comparer bactéries compartimentées ou non. Aucune différence significative entre les bactéries compartimentées et les bactéries non compartimentées n'a pu être identifiée, tendant à prouver que la membrane intra-cytoplasmique ne constitue pas une protection efficace contre les transferts de gènes. Certaines observations intéressantes ont cependant pu être faites, telles que le taux important de transferts entre les Chlamydiae et les plantes ou les différences de fonctions des gènes transférés, non pas selon la structure des bactéries mais selon leur environnement. Cette constatation nous a donc amené à élargir notre étude, en nous interpellant sur les possibles effets de l'environnement des bactéries sur l'évolution de leurs génomes, c'est le thème du deuxième article présenté. Les bactéries déjà étudiées ont été triées en fonction de leur environnement et de leurs style de vie (intracellulaire ou extracellulaire, isolée ou en communauté) et différentes caractéristiques propres aux gènes transférés ont été étudiées. La quantité de HGT et la proportion de gènes due à des transferts au sein des génomes bactériens ont été comparées et analysées. Puis nous nous sommes intéressés aux partenaires des transferts de gènes ainsi qu'a leurs

fonctions, dans le but de détecter une possible sélection par l'environnement. Nous avons pu constater que les bactéries intracellulaires présentent de fortes diminutions du nombre de transferts, sauf si elles sont contenues dans des amibes comme cela été déjà prouvé (16 , 17). De plus les bactéries échangent plus avec les bactéries provenant de leur propre environnement, indépendamment des liens phylogénétiques existants. Enfin l'environnement semble bien affecter la sélection des gènes transférés, puisque plusieurs fonctions sont significativement surreprésentées dans les HGT de bactéries provenant de certains environnements comparé aux autres bactéries.

The genomic contribution to the understanding of the PVC super-phylum

Sandrine PINOS[1,2] , Didier RAOULT[2] and Pierre PONTAROTTI[1]

[1] Institute of Mathematic of Marseille - UMR 7373 AMU-CNRS - Thème Evolution Biologique et Modélisation

[2] URMITE (unité de recherche sur les maladies infectieuses et tropicales émergentes) - UM63, AMU-CNRS

* Corresponding author: Sandrine Pinos, PhD candidate;
URMITE (unité de recherche sur les maladies infectieuses et tropicales émergentes) - UM63, AMU-CNRS
Institute of Mathematic of Marseille, 3 place Victor Hugo, 13331 Marseille CEDEX 03, France
Tel.: + 33 413 55 12 16
E-mail: sandrine.pinos@gmail.com

Keywords : Genomic, evolution, PVC bacteria, genetic events

Abstract

An important amount of data was acquired about bacterial genomes since the first genome sequenced in 1995. These data lead to the reconstruction of the evolution of bacterial genomes contents, allowing to understand the diversity and origins of genomes, but also the causes of phenotypical features. The PVC (*Planctomycetes, Verrucomicrobiae, Chlamydiae*) super-phylum contains bacteria with varied environments and lifestyles. These bacteria present several unusual features, inducing a special interest for the evolution of their genomes contents.

The genomic studies on PVC bacteria allow to understand the diversity of these bacteria: 240 genomes have been sequenced since 2007; additionally lots of 16S fragments different from those of bacteria already described in databases were also identified, showing that we know only a very low proportion of all of these bacteria. New phyla were also identified, as *Omnitrophica, WWE2*, ... Among all the genomic studies dedicated to these bacteria, four were dedicated to the different genetic events occured during the evolution. The comparison of these studies permitted to better understand the genomes content evolution in the PVC super-phylum. Some

genomic works also contributed in the debate about the unusual features of PVC bacteria. For example, the detection of outer membrane protein typical from Gram negatives bacteria by sequence similarity is an interesting argument concerning the question of a possible compartmentalization of *Planctomycetes, Verrucomicrobiae, Poribacteria* and *Lentisphaerae*.

Introduction

In the last years, the important increase of the genomic studies, favored by sequencing methods more and more precise, allowed the acquisition of an incredible amount of information. The analysis and comparison of these data reveal the important diversity of the bacteria and allow also the detection of genetic events occurred during bacterial genomes history, leading to the reconstruction of genomes contents evolution. The reconstruction of evolution show that each genome is a mosaic of genetic material inherited vertically and horizontally from varied sources, also affected by gene loss and duplication (1) (Figure 1). Genomic data are also used to answer questions about genetic causes of physiological or morphological features (2,3). Among all bacteria, we chose to work on the super-phylum PVC (*Planctomycets, Verrucomicrobiae and Chlamydiae*), in one hand for its important diversity and recent metagenomic studies and in other hand for its specific features (for example, the gain of C1 enzyme activity (4) or the possible loss of peptidoglycan (5,6,7,8) in some *Verrucomicrobiae*…) (Figure 2).

1) Genomes available and sequencing projects

The PVC superphylum comprises today 240 complete genomes (counting the different strains of a same species) (9) distributed in the 3 main phyla: *Chlamydia* (180 genomes), *Planctomycetes* (28 genomes), *Verrucomicrobiae* (31 genomes) and 1 minor phyla, *Lentisphaerae* (1 genome). The special interest in *Chlamydiae* is related to the pathogenicity of these bacteria. Lots of other sequencing projects are in progress today, some of them concern bacteria of the 3 main phyla and the others are dedicated to bacteria from minor phyla little studied (*Omnitrophia* (OP3), *Poribacteria and WWE2*). The interest of research community in this super-phylum is really recent, the sequencing began in 2007 with 26 phyla sequenced during this year (especially the different strains of *Chlamydia trachomatis*). The year 2012 was important for genomic of PVC bacteria, with 89 genomes sequenced (whose 75 genomes from strains of *C.trachomatis* and *Chlamydia psitacci*). The latest 2 years present less genomes sequenced with only 28 in 2014.

2) The genomes evolution

1. Genomes evolution and extrinsic constraints: the important diversity of PVC bacteria revealed by genomic

Several metagenomic studies have been realized with a special interest in PVC bacteria detection, these works increase our knowledge about the diversity of the bacteria in this super-phylum.

The *Verrucomicrobiae* were identified thanks to 16S RNA in varied environments (water, soils, plants…) : Bergmann et al (10) proved that these bacteria dominate several bacterial communities (grasslands, soils…), it is also supported that *Verrucomicrobiae* compose 10% of soil bacterial communities (11). Freitas et al (12) determined also that *Verrucomicrobiae* are ubiquitous in marine environments. Metagenomic study of the Rhizosphere by Kielak et al in 2009 (13) permitted the detection of several new sequences of 16s RNA related to *Verrucomicrobiae*, proving that the species already known formed only a little proportion of the phylum. Other studies, as research of *Verrucomicrobiae* in extremes environments (14) and freshwater (15), led also to identify numerous new samples of *Verrucomicrobiae* that will be studied in the future.

The habitats of *Planctomycetes* are less varied than *Verrucomicrobiae* but this phylum remain represented in lots of bacterial communities. A study realized by Shu et al in 2008 (16) about *Planctomycetes* in surface seawater and sediments demonstrate that 50% of the sequences identified are different from *Planctomycetes* sequences already known. So a large part of *Planctomycetes* diversity is still unknown. Another works allow the identification of new *Planctomycetes* sequences in varied environments (Bondoso et al (17) for microalgae, Cayrou et al (18) in human gut, Dedysh (19) for acidophilic Planctomycetes, Chouari at al (20) for Wastewater and Brummer (21) et al for fresh water biofilms).

Chlamydia are the most studied bacteria due to their pathogenicity, however, the recent genomic studies focused more on one species of *Chlamydia* than on their diversity in the different environments. Some metagenomic works were realized, for example Pizzeti et al in 2012 (22), in the Tyrrhenian coastal lake, but the proportion of unknown species in *Chlamydia* seems to be less important than in the two previous phyla.

The *Lentisphaera* were not very well-known today and represent only 1 % of the bacteria in oceans. This phylum contains only 4 characterized

species distributed in 3 families (*Lentisphaeraceae, Victivallaceae and Oligosphaeraceae*) (23). Two metagenomic studies, realized by Chouari et al and published respectively in 2003 and 2005 (20, 24), allowed to identify another monophylogenetic bacterial group possibly associated to *Lentisphaera*, the *WWE2* group. In 2010 Limam et al (25) proved the important diversity of this *WWE2* group and two subgroups were determined.

The *Poribacteria* form an independent monophyletic phylum, these bacteria are symbionts of some marine sponges. One of their strain has a genome sequenced to two-thirds today (26) and lots of segments obtained present a high similarities with sequences from *Rhodopirellula baltica* (27).

The phylum *Omnitrophica* (previously known as *OP3*), is only composed of bacteria known by their 16S RNA and three fragments identified by Glockner et al in 2010 (28). These 16S RNA allow to identify 5 subdivisions and they present a rate of divergence inferior to those of other bacteria, inducing a low phylogenetic deep in these bacteria. In the 3 fragments of genomes detected, numerous sequences present a best hits with the sequences from *Deltaproteobacteria*. However, the frequencies of

tri-tetra nucleotides are uniformly distributed along the sequences, so these ortholog of deltapreoteobacterial genes were probably acquired by ancient HGT or are the results of a convergent evolution caused by a similar lifestyle (based on anaerobic respiration in anoxic environment). Some other phyla are possible candidates as members of PVC super-phylum, bacteria *OD1, BRC1* and *NBL-UPA2* (24).

The PVC bacteria are present in an important quantity of environments, so their genomes have probably been affected by lots of variations since the last common ancestor, allowing the adaptation to different conditions. These bacteria represent an important community of bacteria that should studied deeper to better understand activities in microbial communities.

2. Genomes evolution and Intrinsic constraints : genetic events and variations revealed by genomic

The genomic studies allow to detect the genetic events occurred during bacterial history, inducing the reconstruction of the evolution of genomes contents in PVC superphylum (Figure 1). Four works were especially interested in the evolutionary analysis of PVC genomes : first,

study of Kamneva et al in 2010 (29) is a work focused on indel in PVC genomes. This work gives a first sight into the PVC evolution, 17 species are studied. 12% of branches of the phylogenies are under positive selection of indel (more integration than deletions), few branches present also negative selection. A total of 37365 insertions and 53557 deletions was detected, the medium length of indel was 3.77/3.22 amino acids by protein, respectively. The three other papers are dedicated to the detection of different genetic events that occurred during the PVC evolution. A work realized by Kamneva in 2012 (30) discusses about all genetic events, a project of Pinos et al (paper submitted) focuses on the detection of HGT between PVC and other bacteria. Fuschman et al (31) were interested to determine the proportion of proteins in PVC bacteria that are similar to eukaryote or archaea proteins. We describe briefly the content of these studies and determine the common point characterizing the PVC genomes evolution. Kamneva et al used 26 species in order to determine the rates of the different genetic events occurred along the evolution of PVC super-phylum. In their study, Pinos et al worked on 33 species in order to detect HGT between PVC bacteria or their ancestors and the other bacteria. They also used 31 control bacteria from

phyla *Bacteroidetes-Chlorobi* and *Spirochaetes* in order to compare the HGT rate in the different phyla. Fuchsman et al used more diversified genomes with 134 bacterial genomes blasted against 32 genomes of eukaryotes and archaea in order to determine if *Planctomycetes* or *Verrucomicrobiae* exchange more with eukaryotes and archaea than other bacteria (as it was suggested by Glockner et al in 2003) or not (as it was assumed by Staleyi et al in 2006).

Kamneva et al reconstructed the history of PVC bacteria and determined that the common ancestor of these bacteria should count approximately 3106 genes. However, as we seen in the first part, the number of genomes available increased a lot since 2012, so this approximation should be modified.

The evolution of *Planctomycetes* is characterized by the expansion of their genomes whereas the evolution of *Verrucomicrobiae* and *Lentisphaera* is a balance between the expansion and the reduction of genomes. The *Planctomycetes, Lentisphaerae* and *Verrucomicrobiae* have HGT proportions similar to those of control bacteria.

Chlamydial evolution is marked by the reduction of genomes due to a significantly low gain rate compared to other bacteria, especially in the intracellular species isolated from other microorganisms. The gain rate detected is low due to the important decrease of HGT rate in *Chlamydia*, probably related to their physical isolation from other microorganisms. The decrease of genomic material in these bacteria is consistent with previous studies showing the prevalence of the genome reduction in evolution of intracellular bacteria (32, 33).

Some functions of HGT were identified in the two studies, a high level of HGT of *Chlamydia* are dedicated to cellular motility, a high level of HGT of all PVC bacteria are implicated in transduction signal and amino acid metabolism. Kamneva et al also identified an important quantity of HGT implicated in lipid metabolism, membrane transport and carbohydrate metabolism. Pinos et al identified more transferred genes implicated in Transcription, cell wall/envelop biogenesis and posttranslational modifications/proteins turnover.

Concerning the partners of transfers, bacterial recurrent partners were identified in Kamneva and Pinos studies : the *Proteobacteria, Desulfovibrio*

vulgaris, *Sorangium cellulosum* and *Candidatus Solibacter usitatus* are common partners for recent HGT. The ancestor of PVC seems to exchange more with *Acidobacteria* than other bacteria (*Acidobacteria* are mostly detected in soil, so this observation could be an indication of the environment of PVC ancestor). Kamneva et al identified also three important ways of HGT between PVC bacteria, *Pirellula staleyi* with *Blastopirellula marina, Planctomyces maris* (also called *Gimesia maris*) with *Planctomyces brasiliensis* and *Planctomyces limnophilus*. In their results Pinos et al observed an important exchange of genes between *Chlamydia* and the Eukaryotes, especially in ancient nodes. This observation is consistent with the existing literature dedicated to transfers between plants and *Chlamydia* (34,35) that supports a role of Chlamydia in Chloroplast endosymbiosis. Some bacteria present also a significantly higher HGT rate with Archaea than other bacteria, I.pallida, C.flavus, CK.stuttgartiensis and L.araneosa. This important exchange of genes between CK.stuttgartiensis and Archaea has already been detected by Fuchsman et al in 2006.
These studies give us a general idea of genomes content evolution in PVC super-phylum and their interaction with environment. However, the

genomes of PVC is not very well known, we saw in the first part that lots of bacteria need to be sequenced, especially bacteria from minor phyla. Additionally, the genomes of Planctomycetes and Lentisphaera are enough enigmatic, with only 60% to 54% and 49% of protein associated to a specific function, that is really low compared to other bacteria (72% to 65%) (36, 37). These lack induces a limited reliability for the results obtained. So, in the future, it could be interesting to determine the genes functions thanks to biochemical methods. Associated with the results of genomes sequencing projects, these data will give us a new light concerning the evolution of PVC super-phylum.

3) Genomes evolution and Phenotypic features (Figure 2)

The bacteria of the PVC super-phylum show genetic and cellular features unusual for bacteria: a special cell plan in Verrucomicrobiae, Lentisphaera, Poribacteria and Planctomycetes (38, 39), absence of tubulin like protein FtsZ in Planctomycetes and Chlamydia (40, 41), crateiform surface and budding reproduction (42) in Planctomycetes and some specific enzymatic

activities (amonium oxydation in anammox bacteria (43)or C1 enzyme activity in Methylacidiphilales (4)) .

In some cases, the genomic can be used to determine the causes of these different physiological and morphological characteristics. For example, the C1 enzyme activity is due to a transfer between Archaea and some Planctomycetes and Verrucomicrobiae, followed by an degradation of these genes in Planctomycetes leading to a detection of the C1 Enzyme activity only in the Verrucomicrobiea Methylacidiphilum (4) . Another example is the loss of peptidoglycan in some Verrucomicrobiae from Opitutaceae group, this loss was related to the loss of one protein implicated in the pathway of peptidoglycan biosynthesis (5-8). The FtsZ, an essential protein for cellular division, is missing in Planctomycetes and Chlamydiae, indicating a special cellular division. This absence is a consequence of two independent losses of the gene for FtsZ protein (44). However, the origin of some PVC features are not still resolved. For example the genomic causes of a special cell plan in several bacteria is not known. This specific cell plan concerns all the Planctomycetes (39, 45, 46) some of Verrucomicrobiae (47) one Lentisphaera and one Poribacteria (48). It is characterized by the presence of

an intracytoplasmic membrane, inducing the separation of the cytoplasm in two compartments, the pirellulosome inside (with DNA (49)) and the paryphoplasm outside. Some compartmentalized bacteria show a more complex cell plan with another compartment (anamoxosome in Candidatus Kuenenia Stuttgartiensis (43) or a double internal membrane with ribosome (50) in the same compartment that DNA (49) and endocytosis in Gemmata obscuriglobus (51). Kamneva et al in 2012 (30) tried to detect protein related to this cell plan but they detect only one domain DUF1501, present in all compartmentalized bacteria, except CK.stuttgartiensis. So, the correlation between compartmentalization and the presence of DUF1501 domain is not obvious. The evolution of this character is difficult to determined due to the lack of information about cell plan in several bacteria. However, the most probable history is that cell plan appeared in the common ancestor of PVC super-phylum, followed by a losses in Chlamydiae, when these bacteria became isolated from other microorganims (Pinos et al, submitted paper). The question of the compartmentalization in the Planctomycetes, Verrucomicrobiae, *Lentisphaera* and *Poribacteria* is also subject to very different interpretation. Indeed, some people (52, 53)

suggest that these bacteria are Gram negative bacteria, the intracytoplasmic membrane is only a cytoplasmic membrane with more invaginations than in other Gram negatives bacteria and the cytoplasm is an enlarged periplasm. The genomic works realized on these bacteria give some interesting information about this problem and complete the microscopic observations. In 2012, Speth and al (53) researched and identified several proteins similar to outer membrane proteins typical of bacteria Gram negatives in the Planctomycetes and Verrucomicrobiae genomes (22 species studied where they identified between 20 to 70 outer membrane proteins). The detection of these proteins could indicate the presence of a structure similar to external membrane of gram negative bacteria in the bacteria studied. So, the authors hypothesize that bacteria considered as compartmentalized are just gram-negatives bacteria with a special membrane plasticity. This hypothesis is supported by the identification, in *R.baltica*, of a protein similar to transport membrane proteins characteristic to the envelop of bacteria Gram negatives (54). Another argument debated is the presence or absence of peptidoglycan (PG) in the different bacteria of PVC super-phylum. Until 2014, the Planctomycetes were supposed to lack PG and this phenomenon

was considered as an argument to the compartmentalization of bacteria (37). But in 2014, Jeske et al (55) use genomic analysis on 3 Planctomycetes, completed by biochemical and microscopic analyzes, to prove the presence of PG in Planctomycetes, indeed, all the genes essential for PG synthesis were identified in the 3 bacteria. The paper published by Van-Teesling et al in 2015 (56) confirmed the presence of PG in Planctomycetes. As we seen previously, some Verrucomicrobiae are considered as lacking PG synthesis mechanisms (*Verrucomicrobia, Cerasicoccus, Pelagicoccus, Puniceicoccus* and *Coraliomargarita* (5-8) but genomic analysis of these species could be revised in a next future. These information about cell plan in PVC super-phylum are interesting but the absence of the complete genome of *Poribacteria* (a bacteria potentially compartmentalized) and of *Omnitrophica, WWE2* and *Oligosphaerea* (potential bacteria extracellular without special cell plan) added to the lack of microscopic data prevents a clear conclusion on this topic. We hope that the microscopic information about cell plan in these bacteria will be available in the next future. Associated with the genomes whose sequencing is in project, these data will

allow to better understand the origin, evolution and nature of this hypothetical compartmentalization.

Conclusion

In conclusion the genomic provides important information about the bacteria from PVC super-phylum, in terms of phylogenetic diversity, evolution and phenotypic features. However, some information are still missing to reconstruct the complete history of genomes contents evolution in these bacteria.

Competing interests

The authors declare that they have no competing interests

Acknowledgements

This work was supported by the Assistance Publique - Hopitaux de Marseille (Marseille Public University Hospital System).

Bibliography

1.Chiapello, H., Bourgait, I., Sourivong, F., Heuclin, G., Gendrault-Jacquemard, A., Petit, M. A., & El Karoui, M. (2005). Systematic determination of the mosaic structure of bacterial genomes: species backbone versus strain-specific loops. BMC bioinformatics, 6(1), 171.
2.Stabler, R. A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., ... & Wren, B. W. (2009). Comparative genome and phenotypic analysis of Clostridium difficile 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol, 10(9), R102.
3.Ochman, H., & Moran, N. A. (2001). Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science, 292(5519), 1096-1099.
4.Hou S, Makarova KS et al (2008) Complete genome sequence of the extremely acidophilic methanotroph isolate V4, Methylacidiphilum infernorum, a representative of the bacterial phylum Verrucomicrobia. Biol Direct 3:26
5.Yoon J, Matsuo Y et al (2007a) Cerasicoccus arenae gen. Nov., sp. nov., a carotenoid producing marine representative of the family Puniceicoccaceae within the phylum Verrucomicrobia, isolated from marine sand. Int J Syst Evol Microbiol 57(Pt 9):2067 2072
6.Yoon J, Yasumoto-Hirose M et al (2007b) Coraliomargarita akajimensis gen. nov., sp. nov., a novel member of the phylum Verrucomicrobia isolated from seawater in Japan. Int J Syst Evol Microbiol 57(Pt 5):959–963
7.Yoon J, Yasumoto-Hirose M et al (2007c) Pelagicoccus mobilis gen. nov., sp. nov., Pelagicoccus albus sp. nov. and Pelagicoccus litoralis sp. nov., three novel members of sub-division 4 within the phylum Verrucomicrobia, isolated from seawater by in situ cultiv
8.Choo YJ, Lee K et al (2007) Puniceicoccus vermicola gen. nov.,sp. nov., a novel marine bacterium, and description of Puniceicoccaceae fam. nov., Puniceicoccales ord. nov., Opitutaceae fam. nov., Opitutales ord. nov. and Opitutae classis nov. in the phylum Verrucomicrobia. Int J Syst Evol Microbiol 57(Pt 3):532–537
9.https://gold.jgi.doe.gov/
10.Bergmann, G. T., Bates, S. T., Eilers, K. G., Lauber, C. L., Caporaso, J. G., Walters, W. A., ... & Fierer, N. (2011). The under-recognized dominance of Verrucomicrobia in soil bacterial communities. Soil Biology and Biochemistry, 43(7), 1450-1455.
11.Sangwan, P., Kovac, S., Davis, K. E., Sait, M., & Janssen, P. H. (2005). Detection and cultivation of soil Verrucomicrobia. Applied and environmental microbiology, 71(12), 8402-8410.
12.Freitas, S., Hatosy, S., Fuhrman, J. A., Huse, S. M., Welch, D. B. M., Sogin, M. L., & Martiny, A. C. (2012). Global distribution and diversity of marine Verrucomicrobia. The ISME journal, 6(8), 1499-1505.
13.Kielak, A., Rodrigues, J. L., Kuramae, E. E., Chain, P. S., Van Veen, J. A., & Kowalchuk, G. A. (2010). Phylogenetic and metagenomic analysis of Verrucomicrobia in former agricultural grassland soil. FEMS microbiology ecology, 71(1), 23-33.

14.Sharp, C. E., Smirnova, A. V., Graham, J. M., Stott, M. B., Khadka, R., Moore, T. R., ... & Dunfield, P. F. (2014). Distribution and diversity of Verrucomicrobia methanotrophs in geothermal and acidic environments. Environmental microbiology, 16(6), 1867-1878.
15.Chiang, E. (2015). Ecology of Verrucomicrobia in a Freshwater Estuary (Doctoral dissertation, University of Michigan–Ann Arbor).
16.Shu, Q., & Jiao, N. (2008). Different Planctomycetes diversity patterns in latitudinal surface seawater of the open sea and in sediment. The Journal of Microbiology, 46(2), 154-159.
17.Bondoso, J., Balagué, V., Gasol, J. M., & Lage, O. M. (2014). Community composition of the Planctomycetes associated with different macroalgae. FEMS microbiology ecology, 88(3), 445-456.
18.Cayrou, C., Sambe, B., Armougom, F., Raoult, D., & Drancourt, M. (2013). Molecular diversity of the Planctomycetes in the human gut microbiota in France and Senegal. Apmis, 121(11), 1082-1090.
19.Dedysh, S. N., & Kulichevskaya, I. S. (2013). Acidophilic Planctomycetes: Expanding the Horizons of New Planctomycete Diversity. In Planctomycetes: Cell Structure, Origins and Biology (pp. 125-139). Humana Press.
20.Chouari, R., Le Paslier, D., Daegelen, P., Ginestet, P., Weissenbach, J., & Sghir, A. (2003). Molecular evidence for novel planctomycete diversity in a municipal wastewater treatment plant. Applied and environmental microbiology, 69(12), 7354-7363.
21.Brümmer, I. H. M., Felske, A. D. M., & Wagner-Döbler, I. (2004). Diversity and seasonal changes of uncultured Planctomycetales in river biofilms. Applied and environmental microbiology, 70(9), 5094-5101.
22.Pizzetti, I., Fazi, S., Fuchs, B. M., & Amann, R. (2012). High abundance of novel environmental chlamydiae in a Tyrrhenian coastal lake (Lago di Paola, Italy). Environmental microbiology reports, 4(4), 446-452.
23.Cho, J. C. (2014). The Family Lentisphaeraceae. In The Prokaryotes (pp. 705-710). Springer Berlin Heidelberg.
24.Chouari, R., Le Paslier, D., Dauga, C., Daegelen, P., Weissenbach, J., & Sghir, A. (2005). Novel major bacterial candidate division within a municipal anaerobic sludge digester. Applied and environmental microbiology, 71(4), 2145-2153.
25.Limam, R. D., Bouchez, T., Chouari, R., Li, T., Barkallah, I., Landoulsi, A., & Sghir, A. (2010). Detection of WWE2-related Lentisphaerae by 16S rRNA gene sequencing and fluorescence in situ hybridization in landfill leachate. Canadian journal of microbiology, 56(10), 846-852.
26.Siegl, A., Kamke, J., Hochmuth, T., Piel, J., Richter, M., Liang, C., ... & Hentschel, U. (2011). Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. The ISME journal, 5(1), 61-70.
27.Wagner, M., & Horn, M. (2006). The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. Current opinion in biotechnology, 17(3), 241-249.

28.Glöckner, J., Kube, M., Shrestha, P. M., Weber, M., Glöckner, F. O., Reinhardt, R., & Liesack, W. (2010). Phylogenetic diversity and metagenomics of candidate division OP3. Environmental microbiology, 12(5), 1218-1229.
29.Kamneva, O. K., Liberles, D. A., & Ward, N. L. (2010). Genome-wide influence of indel substitutions on evolution of bacteria of the PVC superphylum, revealed using a novel computational method. Genome biology and evolution, 2, 870-886.
30.Kamneva, O. K., Knight, S. J., Liberles, D. A., & Ward, N. L. (2012). Analysis of genome content evolution in PVC bacterial super-phylum: assessment of candidate genes associated with cellular organization and lifestyle. Genome biology and evolution, 4(12), 1375-1390.
31.Fuchsman, C. A., & Rocap, G. (2006). Whole-genome reciprocal BLAST analysis reveals that planctomycetes do not share an unusually large number of genes with Eukarya and Archaea. Applied and environmental microbiology, 72(10), 6841-6844.
32.Merhej, V., Royer-Carenzi, M., Pontarotti, P., & Raoult, D. 2009. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biology direct 4(1):13.
33.Wolf, Y. I., & Koonin, E. V. 2013. Genome reduction as the dominant mode of evolution. Bioessays 35(9):829-837.
34.Subtil, A., Collingro, A., & Horn, M. (2014). Tracing the primordial Chlamydia: extinct parasites of plants?. Trends in plant science 19(1), 36-43.
35.Moustafa, A., Reyes-Prieto, A., & Bhattacharya, D. (2008). Chlamydia has contributed at least 55 genes to Plantae with predominantly plastid functions. PLoS One 3(5), e2205.
36.Jogler, C., Waldmann, J., Huang, X., Jogler, M., Glöckner, F. O., Mascher, T., & Kolter, R. (2012). Planctomycetes comparative genomics: identification of proteins likely involved in morphogenesis, cell division and signal transduction. Journal of Bacteriology, JB-01325.
37.Fuerst, J. A. (2013). The PVC superphylum: exceptions to the bacterial definition?. Antonie Van Leeuwenhoek, 104(4), 451-466.
38.Lee, K. C., Webb, R. I., Janssen, P. H., Sangwan, P., Romeo, T., Staley, J. T., & Fuerst, J. A. (2009). Phylum Verrucomicrobia representatives share a compartmentalized cell plan with members of bacterial phylum Planctomycetes. BMC microbiology, 9(1), 5.
39.Lindsay, M. R., Webb, R. I., Strous, M., Jetten, M. S., Butler, M. K., Forde, R. J., & Fuerst, J. A. (2001). Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. Archives of Microbiology, 175(6), 413-429.
40.Chopra I, Storey C, Falla TJ & Pearce JH (1998) Antibiotics, peptidoglycan synthesis and genomics: the chlamydial anomaly revisited. Microbiology 144: 2673–2678.
41.McCoy AJ & Maurelli AT (2006) Building the invisible wall: updating the chlamydial peptidoglycan anomaly. Trends Microbiol 14: 70–77.
42.Lage, O. M., Bondoso, J., & Lobo-da-Cunha, A. (2013). Insights into the ultrastructural morphology of novel Planctomycetes. Antonie van Leeuwenhoek, 104(4), 467-476.
43.Kuenen, J. G. (2008). Anammox bacteria: from discovery to application. Nature Reviews Microbiology, 6(4), 320-326.

44.Pilhofer, M., Rappl, K., Eckl, C., Bauer, A. P., Ludwig, W., Schleifer, K. H., & Petroni, G. (2008). Characterization and evolution of cell division and cell wall synthesis genes in the bacterial phyla Verrucomicrobia, Lentisphaerae, Chlamydiae, and Planctomycetes and phylogenetic comparison with rRNA genes. Journal of bacteriology, 190(9), 3192-3202.
45.Lindsay, M. R., Webb, R. I., & Fuerst, J. A. (1997). Pirellulosomes: a new type of membrane-bounded cell compartment in planctomycete bacteria of the genus Pirellula. Microbiology, 143(3), 739-748.
46.Fuerst JA, Sagulenko E. 2011. Beyond the bacterium: planctomycetes challenge our concepts of microbial structure and function. Nat. Rev. Microbiol. 9:403–413.
47.Lee, K. C., Webb, R. I., Janssen, P. H., Sangwan, P., Romeo, T., Staley, J. T., & Fuerst, J. A. (2009). Phylum Verrucomicrobia representatives share a compartmentalized cell plan with members of bacterial phylum Planctomycetes. BMC microbiology, 9(1), 5.
48.Fieseler, L., Horn, M., Wagner, M., & Hentschel, U. (2004). Discovery of the novel candidate phylum “Poribacteria” in marine sponges. Applied and environmental microbiology, 70(6), 3724-3732.
49.Fuerst, J. A., & Webb, R. I. (1991). Membrane-bounded nucleoid in the eubacterium Gemmatata obscuriglobus. Proceedings of the National Academy of Sciences, 88(18), 8184-8188.
50.Fuerst, J. A. (2005). Intracellular compartmentation in planctomycetes. Annu. Rev. Microbiol., 59, 299-328.
51.Lonhienne, T. G., Sagulenko, E., Webb, R. I., Lee, K. C., Franke, J., Devos, D. P., ... & Fuerst, J. A. (2010). Endocytosis-like protein uptake in the bacterium Gemmata obscuriglobus. Proceedings of the National Academy of Sciences, 107(29), 12883-12888.
52.Devos, D. P. (2014). PVC bacteria: variation of, but not exception to, the Gram-negative cell plan. Trends in microbiology, 22(1), 14-20.
53.Speth, D. R., van Teeseling, M. C., & Jetten, M. S. (2012). Genomic analysis indicates the presence of an asymmetric bilayer outer membrane in Planctomycetes and Verrucomicrobia. Frontiers in microbiology, 3.
54.Paparoditis, P., Västermark, Å., Le, A. J., Fuerst, J. A., & Saier, M. H. (2014). Bioinformatic analyses of integral membrane transport proteins encoded within the genome of the planctomycetes species, Rhodopirellula baltica. Biochimica et Biophysica Acta (BBA)-Biomembranes, 1838(1), 193-215.
55.Jeske, O., Schüler, M., Schumann, P., Schneider, A., Boedeker, C., Jogler, M., ... & Jogler, C. (2015). Planctomycetes do possess a peptidoglycan cell wall. Nature communications, 6.
56.van Teeseling, M. C., Mesman, R. J., Kuru, E., Espaillat, A., Cava, F., Brun, Y. V., ... & van Niftrik, L. (2015). Anammox Planctomycetes have a peptidoglycan cell wall. Nature communications, 6.

Figures

Figure 1

Model of genomes evolution in one phylum. Genetic events occured during the evolution of these genomes are indicated by the arrows (continue, for genes transfers and dotted, for gene losses). In the modern species, the parts of the genomes that were vertically and horizontally acquired from different sources are differentiated by colors (blue for vertical inheritance and purple, green, pink, blue and yellow for transfers)

Figure 2

Species tree of PVC bacteria. The unusual features of these bacteria are indicated on this phyla : points for genetic events (pink for gain of C1 enzyme activity, brown for loss of peptidoglycan biosynthesis, green for loss of Ftsz activity, yellow for gain of annamox activity) names of phyla colored for the different cell plans (blue for bacteria with unknown cell plan, red for bacteria with special compartmentalized cell plan, black for bacteria with a classical cell plan)

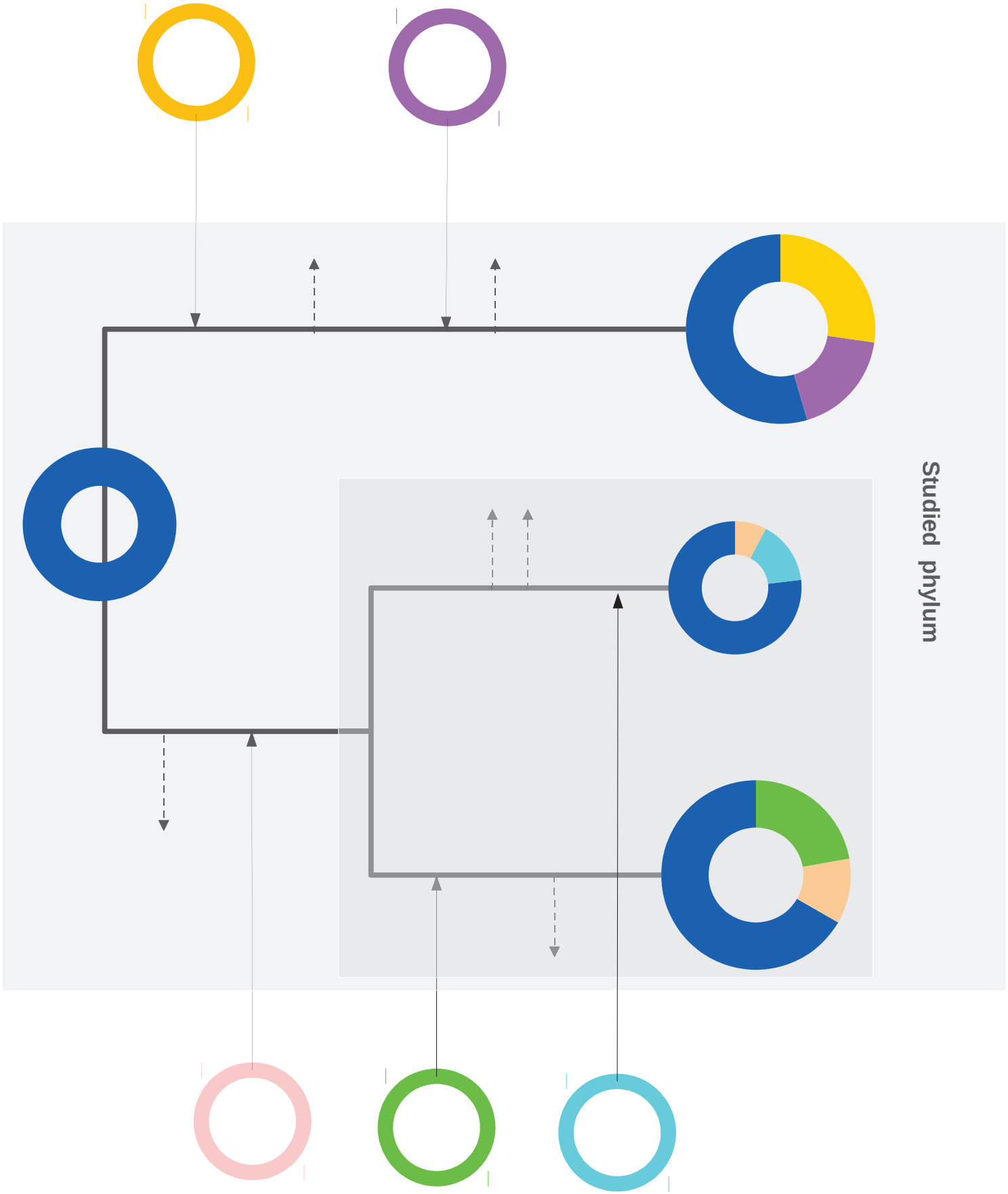


Figure 1

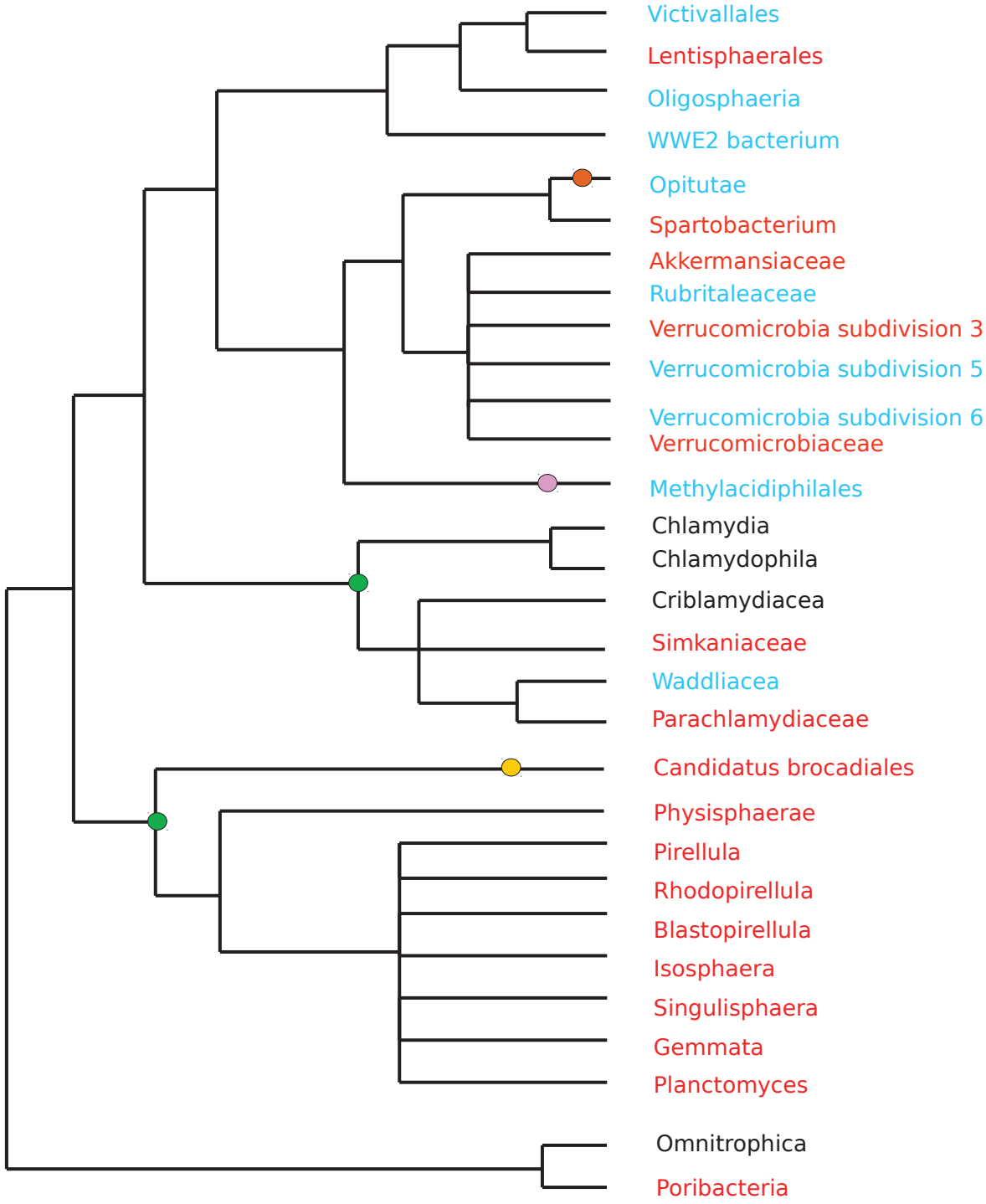

Victivallales
Lentisphaerales
Oligosphaeria
WWE2 bacterium
Opitutae
Spartobacterium
Akkermansiaceae
Rubritaleaceae
Verrucomicrobia subdivision 3
Verrucomicrobia subdivision 5
Verrucomicrobia subdivision 6
Verrucomicrobiaceae
Methylacidiphilales
Chlamydia
Chlamydophila
Criblamydiacea
Simkaniaceae
Waddliacea
Parachlamydiaceae
Candidatus brocadiales
Physisphaerae
Pirellula
Rhodopirellula
Blastopirellula
Isosphaera
Singulisphaera
Gemmata
Planctomyces
Omnitrophica
Poribacteria

Figure 2

Chapitre 1

Le super-phylum PVC est composé de sept phyla comprenant des bactéries avec des modes de vie et des environnements variés. Parmi les différents phyla, quatre (*Planctomycetes, Verrucomicrobiae, Lentisphaera* et *Poribacteria*) comportent des bactéries avec une structure cellulaire inhabituelle chez les bactéries (12-15). Cette structure cellulaire est constituée par une membrane intra-cytoplasmique qui sépare la cellule en deux compartiments distincts (pirellulosome et paryphoplasme) dont l'un contiendrait le nucléoside (18). La présence de cette membrane intra-cytoplasmique reste très controversée et plusieurs études tendent à considérer cette structure cellulaire comme une simple déformation originale de l'enveloppe de bactéries a Gram négatif (19, 20). Qu'elle que soit la réelle nature de cette particularité, sa fonction demeure cependant inconnue. En partant de l'idée de la compartimentation nous avons proposé et testé l'hypothèse suivante : si le nucléoside est bien isolé du cytoplasme par une membrane intra-cytoplasmique, alors il est possible que cette membrane limite le flux d'ADN étranger dans le compartiment interne, diminuant ainsi la fréquence des transferts de

gènes dans l'ADN des organismes compartimentés, en comparaisons des bactéries non compartimentées. Nous avons donc utilisé les génomes de 64 bactéries (33 bactéries PVC et 31 bactéries témoins) afin de déterminer le taux de gènes acquis par transfert dans ces bactéries et comparer les bactéries possédant une membrane intra-cytoplasmique avec les autres bactéries. Afin de mener à bien cette étude nous avons déterminé la structure cellulaire de bactéries pour lesquelles elle était encore inconnue et utilisé les données obtenues pour reconstituer l'évolution de la compartimentation au sein du super-phylum. Une méthode de phylogénomique a été mise au point afin de détecter les transferts de gènes et des analyses statistiques ont été réalisées à partir de ces différents résultats.

Nous avons ainsi pu identifier deux bactéries possiblement compartimentées, en particulier la *chlamydiae Simkania negevensis* qui représente un groupe de *Chlamydia* peu étudié au niveau microscopique. Après reconstitution de l'évolution de la compartimentation chez les bactéries, il semblerait que cette caractéristique soit apparue au niveau de l'ancêtre commun du

super-phylum, cependant le manque d'informations concernant les bactéries PVC issues des phyla mineurs empêche de conclure avec certitude concernant la date d'apparition. La détection et l'analyse des transferts de gènes ne permettent pas de différencier les bactéries compartimentées des non compartimentés, ce qui semble indiquer que la membrane intra-cytoplasmique n'a aucun impact sur le taux de transferts de gènes chez les bactéries. En revanche certains résultats semblent indiquer que le taux de transferts de gènes chez les bactéries pourrait être modulé par le mode vie ou l'environnement des espèces.

Cette article a été soumis récemment au journal Direct Biology.

Compartmentalization in PVC super-phylum:

evolution and impact

Sandrine PINOS[1,2*], Isabelle PAGNIER[1], Pierre PONTAROTTI[2], Didier RAOULT[1]

[1] URMITE (unité de recherche sur les maladies infectieuses et tropicales émergentes) - UM63, AMU-CNRS

[2] Institute of Mathematic of Marseille - UMR 7373 AMU-CNRS - Equipe Evolution Biologique et Modélisation

* Corresponding author: Sandrine Pinos, PhD candidate;
URMITE (unité de recherche sur les maladies infectieuses et tropicales émergentes) - UM63, AMU-CNRS
Institute of Mathematic of Marseille, 3 place Victor Hugo, 13331 Marseille CEDEX 03, France
E-mail: sandrine.pinos@gmail.com

Keywords : Evolution, bacteria, PVC-super phylum, microscopy, compartmentalization

Abstract

Background: The PVC super-phylum gathers bacteria from seven phyla (*Planctomycetes, Verrucomicrobiae, Chlamydiae, Lentisphaera, Poribacteria, OP3, WWE2*) presenting different lifestyles, cell plans and environments. *Planctomyces* and several *Verrucomicrobiae* exhibit a complex cell plan, with an intracytoplasmic membrane inducing the compartmentalization of the cytoplasm into two regions (pirellulosome and paryphoplasm). The evolution and function of this cell plan is always subject to debate. In this work, we hypothesized that it could play a role in protection of the bacterial DNA, especially against Horizontal Genes Transfers (HGT). Therefore, 64 bacterial genomes belonging to 6 different phyla (3 phyla PVC and 3 controls) were studied. We reconstructed the evolution of the cell plan as precisely as possible in each bacteria and their ancestors, thanks to information obtained by bibliographic study and electronic microscopy. We used a strategy based on comparative phylogenomic

in order to determined the part occupied by the horizontal transfers for each studied genomes.

Results: Our results show that the bacteria *Simkania negevensis* (*Chlamydia*) and *Coraliomargarita akajimensis (Verrucomicrobiae*) whose cell plan were unknown before are compartmentalized, as we can see on the micrographies. That is one of the first indication of the presence of intracytoplasmic membrane in a Chlamydia. The proportion of HGT seems to be not related to the cell plan of bacteria, suggesting that compartmentalization does not induce a protection of bacterial DNA against HGT. Conversely, lifestyle of bacteria seems to impact the ability of bacteria to exchange genes.

Conclusions: Our study allows a best reconstruction of the evolution of intracytoplasmic membrane, but this structure seems to have no impact on HGT.

Background

The PVC super-phylum gather seven bacterial phyla *(Planctomycetes, Verrucomicrobiae, Chlamydiae, Lentisphaera, Poribacteria, OP3, WWE2*) [1-4], and comport 37 species of bacteria detected. The monophyly of this super phylum have been discussed a lot in the last years, due to the difficulty to obtain a consensual phylogeny [5-14]. The phylogenetic relations between PVC and other bacteria are a subject still controversial [5, 13]. Recently, it seems that a global consensus was reached, that include all these bacteria in a same super-phylum [2, 5]. This idea is confirmed by the discovery in 2012 and 2014 of a molecular signature conserved in all PVC bacteria [1, 15]. PVC bacteria present an important variety of environments: water, soils, vertebrates, amoeba, insects... [5, 16-21]

The bacteria of this group present different interesting characteristics [11, 22]. Some of these features can be found in other bacteria: some *Planctomycetes* are implicated in carbon and nitrogen cycles [23, 24] or synthesizes special sterols [25], many *Chlamydiae*

are pathogens of mammals [26, 27], some *Chlamydiae* and *Verrucomicrobiae* are symbionts [19]. But PVC present also genetic and cellular features unusual for bacteria: compartmentalization in *Verrucomicrobiae* and *Planctomycetes* [28, 29], absence of tubulin like protein FtZ and of peptidoglycan in cell wall of *Planctomycetes* and *Chlamydia* [30, 31], crateiform surface and budding reproduction [32] in *Planctomycetes*. Among these characteristics, we were specifically interested to the compartmentalization of bacteria. This feature is characterized by a specific cell plan, and concerns all the *Planctomycetes* [29, 33, 34] some of *Verrucomicrobiae* [28] one *Lentisphaera* and one *Poribacteria* [35]. The cytoplasm of bacteria is separated in two compartments by an intracytoplasmic membrane (ICM), the pirellulosome inside (with DNA [36]) and the paryphoplasm outside. This membrane is a lipid bilayer in contact with proteins [29, 32, 33] presenting structural similarities with proteins from eukaryotic membranes as Clathrins [37, 38]. Some compartmentalized bacteria present also a more complex cell plan with presence of another compartment (anamoxosome in

Candidatus Kuenenia Stuttgartiensis [32, 39, 40] or a double internal membrane with ribosome [41] in the same compartment that DNA [36] and endocytosis in *Gemmata obscuriglobus* [42]. Compartmentalization in PVC bacteria is still debatable and three opinions are defended actually : **1-** As some features of PVC bacteria are current in Eukaryota [15], this observation allow some people to assume that the compartmentalization of PVC is the precursor of Eukaryotic nuclei, but this idea is not very popular [34, 43-48]. **2-** Another proposition is that the compartment is structurally and functionally similar to a nucleus but that these two structures appeared independently [49]. **3-** Some people identify the membrane as an invagination of the external membrane |50].

In considering the reality of the existence of compartmentalization in PVC bacteria, it would be interesting to determine the functions of this specific cell plan, because no previous studies could determined this. Here we assume the hypothesis that this membrane could be a protection against Horizontal Gene Transfers (HGT). This hypothesis is based on two main observations:

presence or absence of compartmentalization in the different PVC bacteria and the reconstruction of ancestral states, impact of the internal membrane in Eukaryotes that limits the contact between foreign elements and DNA. We use a strategy of HGT detection to reconstruct the history of HGT during the evolution of PVC bacteria, before and after appearance and disappearance of compartmentalization.

Results

Lifestyle of species and evolution of cell plan

Cell plan and lifestyle (figure 1.a.b.c)

The electron microscopic pictures obtained for the thin sections of *Simkania negevensis* (Figure 1a) and *Coraliomargarita akajimensis* (Figures 1b, 1c), revealed the presence of a potential intracytoplasmic membrane in these 2 bacteria. In both cases, cells contain a pirellulosome and a paryphoplasm separated by the intracytoplasmic membrane. The nucleoid is contained within the pirellulosome. For

S.negevensis, we noticed some differences compared to the other compartmentalized species : We identified the membrane of the phagocytic vacuole, related to the absorption of *Simkania* by an Amoeba, surrounding the Chlamydia. An important presence of internal membranes is also detected, probably the folds of the intracytoplasmic membrane, as it is visible for the G.obscuriglobus in the micrographies and the three dimensional reconstruction published by Santarella-Mellwig et al in 2010 |38] and 2013 [37] respectively. The possible compartmentalization of S.negevensis is very interesting because the cell plan is unknown in a large majority of sympatric Chlamydia. *C.akajimensis* presents a more classical cell plan with an intracytoplasmic membrane close to the external membrane, as it was already observed in some PVC bacteria by Santarella-Mellwig in 2013 [37] and Kuo-Chang Lee and al in 2009 [28]. These observations are really interesting but need to be confirmed by using electron microscopy incorporating high-pressure freezing and cryosubstitution.

Ancestral reconstruction results (Figure 2)

Microscopic observations allow us to better reconstruct the ancestral state of cell plan all along the PVC super-phylum evolution, compared to studies already published. In the tree (Figure 2), we can observe that all allopatric bacteria are not compartmentalized, whereas, before they became isolated, their ancestors should be compartmentalized. All the sympatric PVC bacteria are compartmentalized or supposed to be according cell plan prediction. So, the disappearance of compartmentalization occurred in the same time that bacteria converted to intracellular allopatric lifestyle.

Intracellular allopatric bacteria are isolated from other bacteria, so their genomes are protected against HGT. Extracellular and intracellular sympatric bacteria are more exposed to exchanges because they are in contact with other bacteria. So, it is possible that nuclei disappeared in Intracellular allopatric bacteria because it became useless to avoid HGT, due to isolation of these bacteria.

Relation between cell plan and genomes content (mobilome, HGT)

HGT analysis results (Figure 3)

If we compare allopatric and sympatric bacteria, we detect an important decrease of HGT level after the Chlamydiacea became intracellular allopatric. The rate of recent HGT identified in allopatric bacteria is significantly lower than these of other bacteria (Wilcoxon-Mann-Whitney test : p-value=3.0*10-4). This is probably due to the fact that intracellular bacteria are physically isolated from other microorganisms, which prevents the opportunities of HGT [69]. This agrees with previous studies showing that the predominant evolutionary process in intracellular bacteria is genome reduction leading to small genome sizes [70, 71]. Intracellular bacteria of amoeba with 5.8% of HGT are the exception |72, 73].

The Chlamydia form an interesting group of bacteria, due to their variations in lifestyle and HGT. We focused on this group and we compared the rate of recent HGT between Chlamydiae and the other PVC bacteria. We noticed a low significant difference (t-test : p-value=1.2*10-2). Conversely, if we compare the proportion of HGT

between allopatric and sympatric Chlamydiae, we noticed that sympatric Chlamydia exchange importantly and significantly more with other bacteria than allopatric Chlamydiae (t-test : p-value=5.9*10-3). The results observed for all the bacteria of our set, for PVC bacteria only and for Chlamydia are not modified if we compared only the proportions of genes exchanged with other bacteria and viruses (without sequences transferred with Eukaryotes).

We focused only on sympatric bacteria and compare compartmentalized species with no compartmentalized. The two groups are comparable because the phylogenetic richness and deepness of bacterial groups are similar. We noticed that there is no difference in the level of HGT (recent or ancient) according cell plan. We could imagine that this absence of difference is due to a problem of our strategy but two observations demonstrate the efficiency of our method: *Chlamydia* present an important rate of ancient HGT with eukaryotes, significantly higher than other of bacteria (ANOVA test p-value= 3.3X10-6). This observation is consistent with the

existing literature dedicated to transfers between plants and *Chlamydia* |74, 75] that supports a role of *Chlamydia* in Chloroplast endosymbiosis: *Spirochaetes* seem to exchanges more with *Firmicutes* as it was already detected in previous studies |76, 77]

We observed that intracytoplasmic membrane has no impact on HGT frequency, its function remains unknown. So the disappearance of intracytoplasmic membrane in intracellular isolated bacteria should be related to another cause.

Mobilomes study results

Among the bacterial set, we detected only one element related to mobilomes that seems to vary significantly with cell plan: conjugation genes that are significantly over represented in the genomes of Intracellular allopatric bacteria, and underrepresented in extracellular compartmentalized bacteria . But conjugation genes have other uses in bacteria than conjugation, they play role in flagella and other mechanisms. So the abundance of conjugation genes in bacteria has probably no influence on HGT.

Material and method

Bacteria selection and genomes recovery

We selected Bacteria from 4 phyla of PVC super-phylum: *Planctomycetes, Verrucomicrobia, Lentisphaerae*, and *Chlamydiae.* Bacteria from three phylogenetically close phyla are choose as negative controls, based on a reference tree (this tree was built with 40 markers, comprises 99 bacteria and was published in 2012 by Kamneva et al [5]: *Bacteroidetes, Spirochaetes* and *Chlorobi*. We retrieved the proteomes of all bacteria selected in genomes NCBI database. We also studied the mobilomes of bacteria, in this aim, few bacteria representing each main phyla and different intracellular bacteria were selected as control groups.

Lifestyle determination

The lifestyle of selected bacteria is determined by a bibliographic study of each bacterium. Two types of lifestyles extracellular are known: allopatric or sympatric bacteria. The definition of allopatry and sympatry for bacteria are given in different publications [51, 52] : **Sympatric bacteria** are bacteria living in

community with other microorganisms, they exchange easily genes by lateral transfer through their interactions. **Allopatric bacteria** are living in cells and are not in contact with other microorganisms, they are obligate intracellular bacteria.

Cell plan determination

Cell plans of bacteria are determined thanks to transmission electron microscopy pictures already available in bibliography [28, 29, 34, 35] and microscopic observations of the bacteria whose cell plan is unknown. First, the species *Simkania negevensis* (DSM27360, type strain) was grown within a culture of its amoebal host *A. castellanii.* At H6, and H16 post-infection, cultures were centrifuged at 2000 rpm/min during 10 minutes, and the pellets were fixed for electron microscopy. The second species, *Coraliomargarita akajimensis* (DSM 45221, type strain), is not an obligate intracellular bacteria, and thus, could be cultivated directly into Bacto Marine Broth medium (Difco). Growth occurred after 6 days of cultivation, and when the bacterial suspension reached the exponential growth phase, it was centrifuged at 5500 rpm/min during 30 minutes, and

fixed for electron microscopy.

An aliquot of each culture was kept for DNA extraction, followed by a standard 16S rRNA PCR, in order to confirm the bacterial identification. DNA extraction was performed using the automate EZ1, following manufacturer's instructions (QiaGen). DNA amplification and sequencing reactions were performed with a thermocycler Mastercycler (Eppendorf), using respectively the following sets of primers: FD1 (5' AGAGTTTGATCCTGGCTCAG 3') and RP2 (5' ACGGCTACCTTGTTACGACTT-3'); and 536F (5'-CAGCAGCCGCGGTAATAC-3') and RP2. Sequences were corrected and analyzed using the software CodonCode, and identification was done using the online BLAST program of the NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). For both bacteria, Electron microscopic observations were realized in different steps : for embedding, cells were fixed for 1 hour with glutaraldehyde 2.5% in 0.1M sodium cacodylate buffer and washed three times with a mixture of 0.2M saccharose/0.1M sodium cacodylate. Cells were post-fixed for 1 hour with 1% OsO4 diluted in 0.2M Potassium hexa-

cyanoferrate (III) / 0.1M sodium cacodylate solution. After four 5 min washes with distilled water, cells were gradually dehydrated with ethanol by successive 10 min baths in 30%, 50%, 70%, 96%, 100% and 100% ethanol. Substitution was achieved by successively placing the cells in 25%, 50% and 75% Epon solutions for 15 minutes. Cells were placed for 1 hour in 100% Epon solution and in fresh Epon 100% over-night under vacuum at room-temperature. Polymerization took place with cells in fresh 100% Epon for 24h at 60°. All solutions used above were 0.2µm filtered. Ultrathin 70 nm sections were cut with a UC7 ultramicrotome (Leica) and placed on HR25 300 Mesh Copper/Rhodium grids (TAAB, UK). Sections were stained with Reynolds solution. Electron micrographs were obtained on a Tecnai G20 TEM (FEI) operated at 200 keV equipped with a 4096 x 4096 pixels resolution Eagle camera (FEI)).

Phylogeny and cell plan evolution reconstruction

The phylogenies of PVC bacteria and *Bacteroidetes-Chlorobi-Spirochaeta* were reconstructed thanks to 16S and 23S RNA. We used Mega5 [53] software to perform it (alignment by Muscle algorithm

[54, 55], manual removal of non-conserved positions, tree built with NJ |56] and ML [57] methods, with 150 bootstraps). We checked if tree obtained is congruent with our reference tree for validation. Thanks to information about compartmentalization in bacteria and phylogeny, we reconstructed the ancestral state of cell plan with parsimony method (Mesquite software [58, 59]). The tree obtained allow us to date the different events occurred during evolution of the phyla studied (appearance and disappearance of compartmentalization, conversion to intracellular allopatric or sympatric lifestyle).

Mobilome study

The presence of intracytoplasmic membrane in bacteria can disturb only certain mechanisms of transfers among all of them, leading or not to a decrease of HGT level depending if the deficiency of some mechanisms are offset by the other. The elements studied are those involved in HGT, elements of mobilome or elements allowing transfers. Transfers are realized in four steps described below:

- Transfer of the sequence between donor and acceptor, realized by conjugation (thanks to Conjugation genes and Plasmids) or Transduction (by Phages (complete or incomplete))
- Insertion of the transferred sequence in the genome of the acceptor: mediated by Transposases and Integrases. This process can be stopped by CRISPs.
- Translation of inserted sequence because a sequence not translated is useless and will disappear. The translation of a sequence derived from another organism depends on quantity and diversity of tRNA, more there are various types of tRNA, the more it is possible that one of these tRNA can translate transferred sequence

We retrieve information about these element on different databases.

HGT detection

In order to detect HGT, OrthoMCL [60] was used for construction of orthologous groups. If a gene is absent in Orthologous groups, it means that it is only present in a single species among all those studied, we consider it as a "potential orphan".

These genes were acquired by recent gains (gain specific to one species) and can be either real orphans or recent HGT candidates. Based on specie tree and Orthologous groups, Phylopattern [61] pipeline allowed the detection of genetic events in four steps: Comparison between species tree and each orthologous groups. Detection of missing species in orthologous groups. Reconstruction of ancestral state for each proteins. Identification of two types of genetic events: gains and losses. These gains are ancient and could be a possible ancient HGT, de novo genes, or artifacts. Then we focused on gene gains (ancient and recent) and Blast [62] each sequences against nr database (NCBI). Probable HGT (recent or ancient) are identified thanks to filtering on Blast results: For each blast results, sequences derived from phylum where gene gain was detected were removed (for example, if gain was identified in *Akkermansia muciniphila* (Verrucomicrobiae), all sequences in Blast results derived from Verrucomicrobiae were removed). Sequences with e-value>10-5, coverage<60% or identities<30% were also removed. Then, species of the first 10 sequences remaining in Blast

result were identified. If species identified do not belong to one of the six phyla studied, gene gain is probably an HGT.

For each HGT (recent or ancient), the position of genes transferred were defined thanks to the RAST software [63, 64] in order to detect those that are grouped and could have been transferred in a same event, . Then, we calculated the number of HGT, the percentage of proteins present in each species due to transfers and the rate of pairs base (pb) sequences implicated in transfers compared to all genomes. The length of sequences transferred was also determined.

Statistical analyzes

We realized a comparison of variances between the different groups of bacteria (based on cell plan) for each elements studied (test of normality followed by Kruskal test or ANOVA and Tukey or Nemenuyi test). To check the phylogenetic bias, we realized also statistical analysis with bacterial groups based on the phyla. A Principal component analysis (PCA) and a hierarchical clustering principal component (HCPC) were also realized to define the

clustering of bacteria according the different parameters studied and determine if clustering can be related with phylogeny or cell plan. So, we can see if compartmentalized bacteria present a specific HGT ratio or mobilome structure (we used only elements significantly variable detected by comparison of variance).

Conclusion

The different pictures obtained for Verrucomicrobiae and Chlamydiae allow a best definition of ancestral and modern states of compartmentalization and so, permit to reconstruct intracytoplasmic membrane evolution in PVC bacteria. But this compartmentalization seems have no significant impact on HGT quantity or mechanisms.

Competing interests

The authors declare that they have no competing interests

Author's contributions

PS carried out the design of the study, the strategy elaboration and the collect of data, performed the statistical analysis and interpretation of results and drafted the manuscript. **PI** coordinated and participated in microscopic observations, helped in data interpretation and revised the manuscript. **PP** participated in strategy elaboration, data interpretation and revised the manuscript. **RD** conceived the study, participated in its design and coordination and revised the manuscript.

Acknowledgements

We thank Olivier Chabrol for assistance in computer programming during the elaboration of HT detection strategy and Manuela Royer Carenzi for the assistance in statistical analyzes. We also thank the Xegen company for the assistance in HT detection by using of Phylopattern software. This work was supported by the Assistance Publique - Hopitaux de Marseille (Marseille Public University Hospital System).

Bibliography

1.Gupta, R. S., Bhandari, V., & Naushad, H. S. (2012). Molecular signatures for the PVC clade (Planctomycetes, Verrucomicrobia, Chlamydiae, and Lentisphaerae) of bacteria provide insights into their evolutionary relationships. Frontiers in microbiology, 3.
2.Wagner, M., & Horn, M. (2006). The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. Current opinion in biotechnology, 17(3), 241-249.
3.Cho, J. C., Vergin, K. L., Morris, R. M., & Giovannoni, S. J. (2004). Lentisphaera araneosa gen. nov., sp. nov, a transparent exopolymer producing marine bacterium, and the description of a novel bacterial phylum, Lentisphaerae. Environmental Microbiology, 6(6), 611-621.
4.Siegl, A., Kamke, J., Hochmuth, T., Piel, J., Richter, M., Liang, C., ... & Hentschel, U. (2011). Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. The ISME journal, 5(1), 61-70.
5.Kamneva, O. K., Knight, S. J., Liberles, D. A., & Ward, N. L. (2012). Analysis of genome content evolution in PVC bacterial super-phylum: assessment of candidate genes associated with cellular organization and lifestyle. Genome biology and evolution, 4(12), 1375-1390.
6.Yarza, P., Richter, M., Peplies, J., Euzeby, J., Amann, R., Schleifer, K. H., ... & Rosselló-Móra, R. (2008). The All-Species Living Tree project: A 16S rRNA-based phylogenetic tree of all sequenced type strains. Systematic and applied microbiology, 31(4), 241-250.
7.Griffiths, E., & Gupta, R. S. (2007). Phylogeny and shared conserved inserts in proteins provide evidence that Verrucomicrobia are the closest known free-living relatives of chlamydiae. Microbiology, 153(8), 2648-2654.
8.Brochier, C., & Philippe, H. (2002). Phylogeny: a non-hyperthermophilic ancestor for bacteria. Nature, 417(6886), 244-244.
9.Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N. N., ... & Eisen, J. A. (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. Nature, 462(7276), 1056-1060.
10.Jenkins, C., & Fuerst, J. A. (2001). Phylogenetic analysis of evolutionary relationships of the planctomycete division of the domain bacteria based on amino acid sequences of elongation factor Tu. Journal of molecular evolution, 52(5), 405-418.
11.Pilhofer, M., Rappl, K., Eckl, C., Bauer, A. P., Ludwig, W., Schleifer, K. H., & Petroni, G. (2008). Characterization and evolution of cell division and cell wall synthesis genes in the bacterial phyla Verrucomicrobia, Lentisphaerae, Chlamydiae, and Planctomycetes and phylogenetic comparison with rRNA genes. Journal of bacteriology, 190(9), 3192-3202.
12.Gupta, R. S., & Griffiths, E. (2002). Critical issues in bacterial phylogeny. Theoretical population biology, 61(4), 423-434.
13.Jun, S. R., Sims, G. E., Wu, G. A., & Kim, S. H. (2010). Whole-proteome phylogeny of prokaryotes by feature frequency profiles: An alignment-free method with optimal feature resolution. Proceedings of the National Academy of Sciences, 107(1), 133-138.
14.Ciccarelli, F. D., Doerks, T., Von Mering, C., Creevey, C. J., Snel, B., & Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. science, 311(5765), 1283-1287.
15.Lagkouvardos, I., Jehl, M. A., Rattei, T., & Horn, M. (2014). Signature Protein of the PVC Superphylum. Applied and environmental microbiology, 80(2), 440-445.

16.Buckley, D. H., Huangyutitham, V., Nelson, T. A., Rumberger, A., & Thies, J. E. (2006). Diversity of Planctomycetes in soil in relation to soil history and environmental heterogeneity. Applied and environmental microbiology, 72(7), 4522-4531.
17.Wang, J., Jenkins, C., Webb, R. I., & Fuerst, J. A. (2002). Isolation of Gemmata-like and Isosphaera-like planctomycete bacteria from soil and freshwater. Applied and environmental microbiology, 68(1), 417-422.
18.Sangwan, P., Kovac, S., Davis, K. E., Sait, M., & Janssen, P. H. (2005). Detection and cultivation of soil Verrucomicrobia. Applied and environmental microbiology, 71(12), 8402-8410.
19.Horn, M. (2008). Chlamydiae as symbionts in eukaryotes. Annu. Rev. Microbiol., 62, 113-131.
20.Zwart, G., Crump, B. C., Kamst-van Agterveld, M. P., Hagen, F., & Han, S. K. (2002). Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. Aquatic Microbial Ecology, 28.
21.Bergmann, G. T., Bates, S. T., Eilers, K. G., Lauber, C. L., Caporaso, J. G., Walters, W. A., ... & Fierer, N. (2011). The under-recognized dominance of Verrucomicrobia in soil bacterial communities. Soil Biology and Biochemistry, 43(7), 1450-1455.
22.Ward, N., Staley, J. T., Fuerst, J. A., Giovannoni, S., Schlesner, H., & Stackebrandt, E. (2006). The order planctomycetales, including the genera planctomyces, pirellula, gemmata and isosphaera and the candidatus genera brocadia, kuenenia and scalindua (pp. 757-793). Springer New York.
23.Shu, Q., & Jiao, N. (2008). Different Planctomycetes diversity patterns in latitudinal surface seawater of the open sea and in sediment. The Journal of Microbiology, 46(2), 154-159.
24.Strous, M., Pelletier, E., Mangenot, S., Rattei, T., Lehner, A., Taylor, M. W., ... & Le Paslier, D. (2006). Deciphering the evolution and metabolism of an anammox bacterium from a community genome. Nature, 440(7085), 790-794.
25.Pearson, A., Budin, M., & Brocks, J. J. (2003). Phylogenetic and biochemical evidence for sterol synthesis in the bacterium Gemmata obscuriglobus. Proceedings of the National Academy of Sciences, 100(26), 15352-15357.
26.Bébéar, C., & De Barbeyrac, B. (2009). Genital Chlamydia trachomatis infections. Clinical Microbiology and Infection, 15(1), 4-10.
27.Corsaro, D., & Greub, G. (2006). Pathogenic potential of novel Chlamydiae and diagnostic approaches to infections due to these obligate intracellular bacteria. Clinical Microbiology Reviews, 19(2), 283-297.
28.Lee, K. C., Webb, R. I., Janssen, P. H., Sangwan, P., Romeo, T., Staley, J. T., & Fuerst, J. A. (2009). Phylum Verrucomicrobia representatives share a compartmentalized cell plan with members of bacterial phylum Planctomycetes. BMC microbiology, 9(1), 5.
29.Lindsay, M. R., Webb, R. I., Strous, M., Jetten, M. S., Butler, M. K., Forde, R. J., & Fuerst, J. A. (2001). Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. Archives of Microbiology, 175(6), 413-429.
30.Chopra I, Storey C, Falla TJ & Pearce JH (1998) Antibiotics, peptidoglycan synthesis and genomics: the chlamydial anomaly revisited. Microbiology 144: 2673–2678.
31.McCoy AJ & Maurelli AT (2006) Building the invisible wall: updating the chlamydial peptidoglycan anomaly. Trends Microbiol 14: 70–77.
32.Lage, O. M., Bondoso, J., & Lobo-da-Cunha, A. (2013). Insights into the ultrastructural morphology of novel Planctomycetes. Antonie van Leeuwenhoek, 104(4), 467-476.

33.Lindsay, M. R., Webb, R. I., & Fuerst, J. A. (1997). Pirellulosomes: a new type of membrane-bounded cell compartment in planctomycete bacteria of the genus Pirellula. Microbiology, 143(3), 739-748.
34.Fuerst JA, Sagulenko E. 2011. Beyond the bacterium: planctomycetes challenge our concepts of microbial structure and function. Nat. Rev. Microbiol. 9:403–413.
35.Fieseler, L., Horn, M., Wagner, M., & Hentschel, U. (2004). Discovery of the novel candidate phylum "Poribacteria" in marine sponges. Applied and environmental microbiology, 70(6), 3724-3732.
36.Fuerst, J. A., & Webb, R. I. (1991). Membrane-bounded nucleoid in the eubacterium Gemmatata obscuriglobus. Proceedings of the National Academy of Sciences, 88(18), 8184-8188.
37.Santarella-Mellwig, R., Pruggnaller, S., Roos, N., Mattaj, I. W., & Devos, D. P. (2013). Three-dimensional reconstruction of bacteria with a complex endomembrane system. PLoS biology, 11(5), e1001565.
38.Santarella-Mellwig, R., Franke, J., Jaedicke, A., Gorjanacz, M., Bauer, U., Budd, A., ... & Devos, D. P. (2010). The compartmentalized bacteria of the planctomycetes-verrucomicrobia-chlamydiae superphylum have membrane coat-like proteins. PLoS biology, 8(1), e1000281.
39.Van de Graaf, A. A., Mulder, A., de Bruijn, P., Jetten, M. S., Robertson, L. A., & Kuenen, J. G. (1995). Anaerobic oxidation of ammonium is a biologically mediated process. Applied and environmental microbiology, 61(4), 1246-1251.
40.Strous, M., Fuerst, J. A., Kramer, E. H., Logemann, S., Muyzer, G., van de Pas-Schoonen, K. T., ... & Jetten, M. S. (1999). Missing lithotroph identified as new planctomycete. Nature, 400(6743), 446-449.
41.Fuerst, J. A. (2005). Intracellular compartmentation in planctomycetes. Annu. Rev. Microbiol., 59, 299-328.
42.Lonhienne, T. G., Sagulenko, E., Webb, R. I., Lee, K. C., Franke, J., Devos, D. P., ... & Fuerst, J. A. (2010). Endocytosis-like protein uptake in the bacterium Gemmata obscuriglobus. Proceedings of the National Academy of Sciences, 107(29), 12883-12888.
43.Devos DP,Reynaud EG. 2010. Evolution. Intermediate steps. Science 330:1187–1188.
44.McInerney, J. O., Martin, W. F., Koonin, E. V., Allen, J. F., Galperin, M. Y., Lane, N., ... & Embley, T. M. (2011). Planctomycetes and eukaryotes: a case of analogy not homology. Bioessays, 33(11), 810-817.
45.Fuchsman, C. A., & Rocap, G. (2006). Whole-genome reciprocal BLAST analysis reveals that planctomycetes do not share an unusually large number of genes with Eukarya and Archaea. Applied and environmental microbiology, 72(10), 6841-6844.
46.Budd, A., & Devos, D. P. (2012). Evaluating the evolutionary origins of unexpected character distributions within the bacterial Planctomycetes-Verrucomicrobia-Chlamydiae superphylum. Frontiers in microbiology, 3.
47.Forterre, P. (2011). A new fusion hypothesis for the origin of Eukarya: better than previous ones, but probably also wrong. Research in microbiology, 162(1), 77-91.
48.Fuerst, J. A., & Sagulenko, E. (2012). Keys to eukaryality: planctomycetes and ancestral evolution of cellular complexity. Frontiers in microbiology, 3.
49.Fuerst, J. A. (2013). The PVC superphylum: exceptions to the bacterial definition?. Antonie Van Leeuwenhoek, 104(4), 451-466.
50.Devos, D. P. (2014). PVC bacteria: variation of, but not exception to, the Gram-negative cell plan. Trends in microbiology, 22(1), 14-20.

51.Georgiades, K. (2012). Genomics of epidemic pathogens. Clinical Microbiology and Infection, 18(3), 213-217.
52.Merhej, V., Notredame, C., Royer-Carenzi, M., Pontarotti, P., & Raoult, D. (2011). The rhizome of life: the sympatric Rickettsia felis paradigm demonstrates the random transfer of DNA sequences. Molecular biology and evolution, 28(11), 3213
53.Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28: 2731-2739.
54.Edgar RC (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput". Nucleic Acids Research 32 (5): 1792–97. doi:10.1093/nar/gkh340.
55.Edgar RC (2004). "MUSCLE: a multiple sequence alignment method with reduced time and space complexity". BMC Bioinformatics 5 (1): 113.
56.Saitou, N., and Nei, M. (1987), « The neighbor-joining method: a new method for reconstructing phylogenetic trees », Mol. Biol. Evol., 4(4):406-425.
57.Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In Selected Papers of Hirotugu Akaike (pp. 199-213). Springer New York.
58.Maddison, W. P. and D.R. Maddison. 2015. Mesquite: a modular system for evolutionary analysis. Version 3.04
59.MESQUITE : http://mesquiteproject.org consulted in june and july 2016
60.Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., Shanmugam, D., Roos, D. S. and Stoeckert, C. J. Using OrthoMCL to Assign Proteins to OrthoMCL-DB Groups or to Cluster Proteomes Into New Ortholog Groups Current Protocols in Bioinformatics. 2011 35:6.12.1–6.12.19.
61.Gouret, P., Thompson, J. D., & Pontarotti, P. (2009). PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC bioinformatics, 10(1), 298.
62.Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of molecular biology, 215(3), 403-410.
63.Glass, E. M., Wilkening, J., Wilke, A., Antonopoulos, D., & Meyer, F. (2010). Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. Cold Spring Harbor Protocols, 2010(1), pdb-prot5368.
64.Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., ... & Edwards, R. A. (2008). The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC bioinformatics, 9(1), 386.
65.Wu, S., Zhu, Z., Fu, L., Niu, B., & Li, W. (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC genomics, 12(1), 444.
66.Galperin, M. Y., Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2014). Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic acids research, gku1223.
67.Tatusov, R. L., Koonin, E. V., & Lipman, D. J. (1997). A genomic perspective on protein families. Science, 278(5338), 631-637.
68.Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., ... & Finn, R. D. (2014). The InterPro protein families database: the classification resource after 15 years. Nucleic acids research, gku1243.
69.Hirt, R. P., Alsmark, C., & Embley, T. M. Lateral gene transfers and the origins of the eukaryote proteome: a view from microbial parasites. Current opinion in microbiology. 2015; 23:155-162.
70.Merhej, V., Royer-Carenzi, M., Pontarotti, P., & Raoult, D. Massive comparative genomic

analysis reveals convergent evolution of specialized bacteria. Biology direct. 2009; 4(1):13.
71.Wolf, Y. I., & Koonin, E. V. Genome reduction as the dominant mode of evolution. Bioessays. 2013; 35(9), 829-837.
72.Molmeret, M., Horn, M., Wagner, M., Santic, M., & Kwaik, Y. A. Amoebae as training grounds for intracellular bacterial pathogens. Applied and environmental microbiology. 2005; 71(1), 20-28.
73.Moliner, C., Fournier, P. E., & Raoult, D. Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. FEMS microbiology reviews. 2010; 34(3):281-294
74.Subtil, A., Collingro, A., & Horn, M. (2014). Tracing the primordial Chlamydiae: extinct parasites of plants?. Trends in plant science, 19(1), 36-43.
75.Facchinelli, F., Colleoni, C., Ball, S. G., & Weber, A. P. (2013). Chlamydia, cyanobiont, or host: who was on top in the ménage à trois?. Trends in plant science, 18(12), 673-679.
76.Viswanathan, V. K. (2012). Spirochaetes and their twisted ways. Gut microbes, 3(5), 399-400.
77.Bellgard MI, Wanchanthuek P, La T, Ryan K, Moolhuijzen P, Albertyn Z, . Genome sequence of the pathogenic intestinal spirochete brachyspira hyodysenteriae reveals adaptations to its lifestyle in the porcine large intestine. PLoS One 2009; 4:e4641;

Tables and figures

Figure 1a : Transmission electron micrograph of thin section of cell of Simkania negevensis. Pirellulosome (PI) and paryphoplasm (P) are separated by the intracytoplasmic membrane (ICM). The nucleoid (N) is contained within the pirellulosome. The bacterium lives in the cytoplasm of an amoeba and is surrounded by the membrane of phagocytic vacuole (VM). Bar marker, 0.2 µm.

Figure 1b and 1C : Transmission electron micrograph of thin section of cell of *Coraliomargarite akajimensis*. Pirellulosome (PI) and paryphoplasm (P) are separated by the intracytoplasmic membrane (ICM). The nucleoid (N) is contained within the pirellulosome. Bar marker, 0.2 µm.

Figure 2 : Phylogeny of PVC super phylum with indications of cell plan and lifestyles for modern species and their ancestors. (The points at the nodes indicates the cell plan : black for compartmentalization,

white for no compartmentalization and red for unknown cell plan. The color squares give the lifestyles of bacteria)

Figures 3 : Proportion of genes acquired by transfer recently and formerly in each genomes of the 65 bacteria studied. The black points at the nodes of phylogeny show the quantity of HGT occurred between the node and the following. The rates of recent and ancient HGT are presented in tables, a high rate of HGT is indicates by a red color, a low rate by yellow and very low by green.

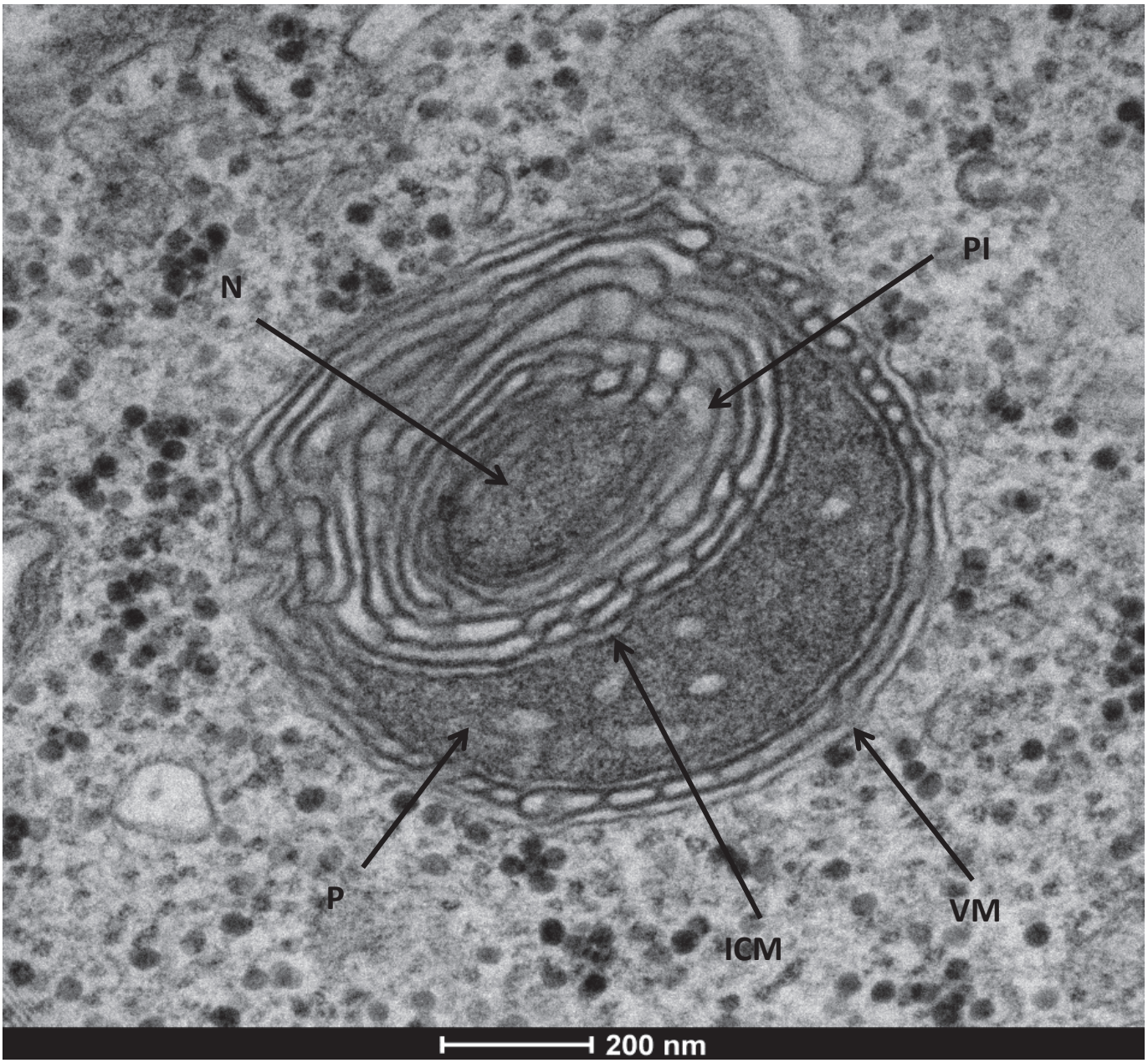


Figure 1.a

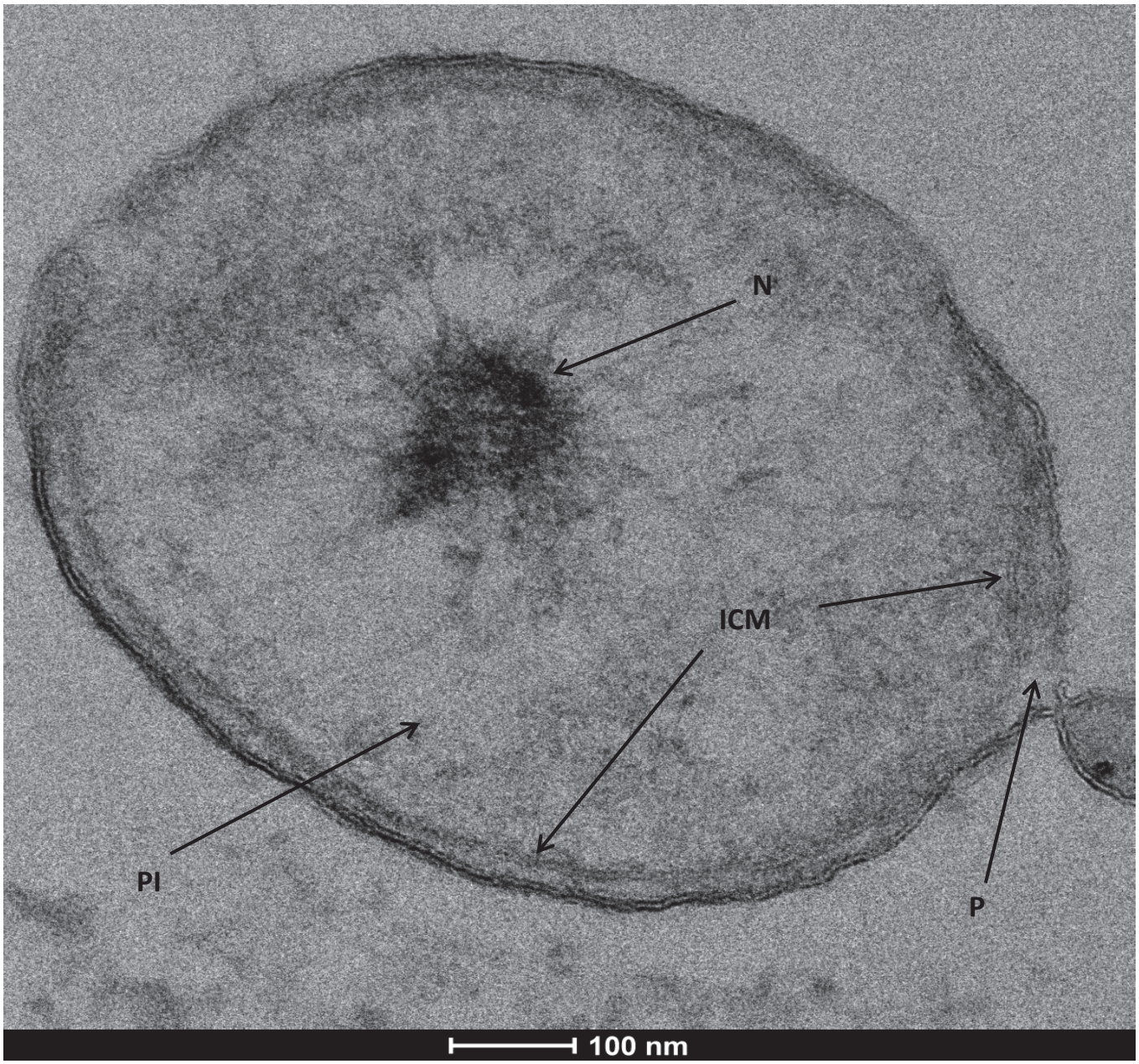


Figure 1.b

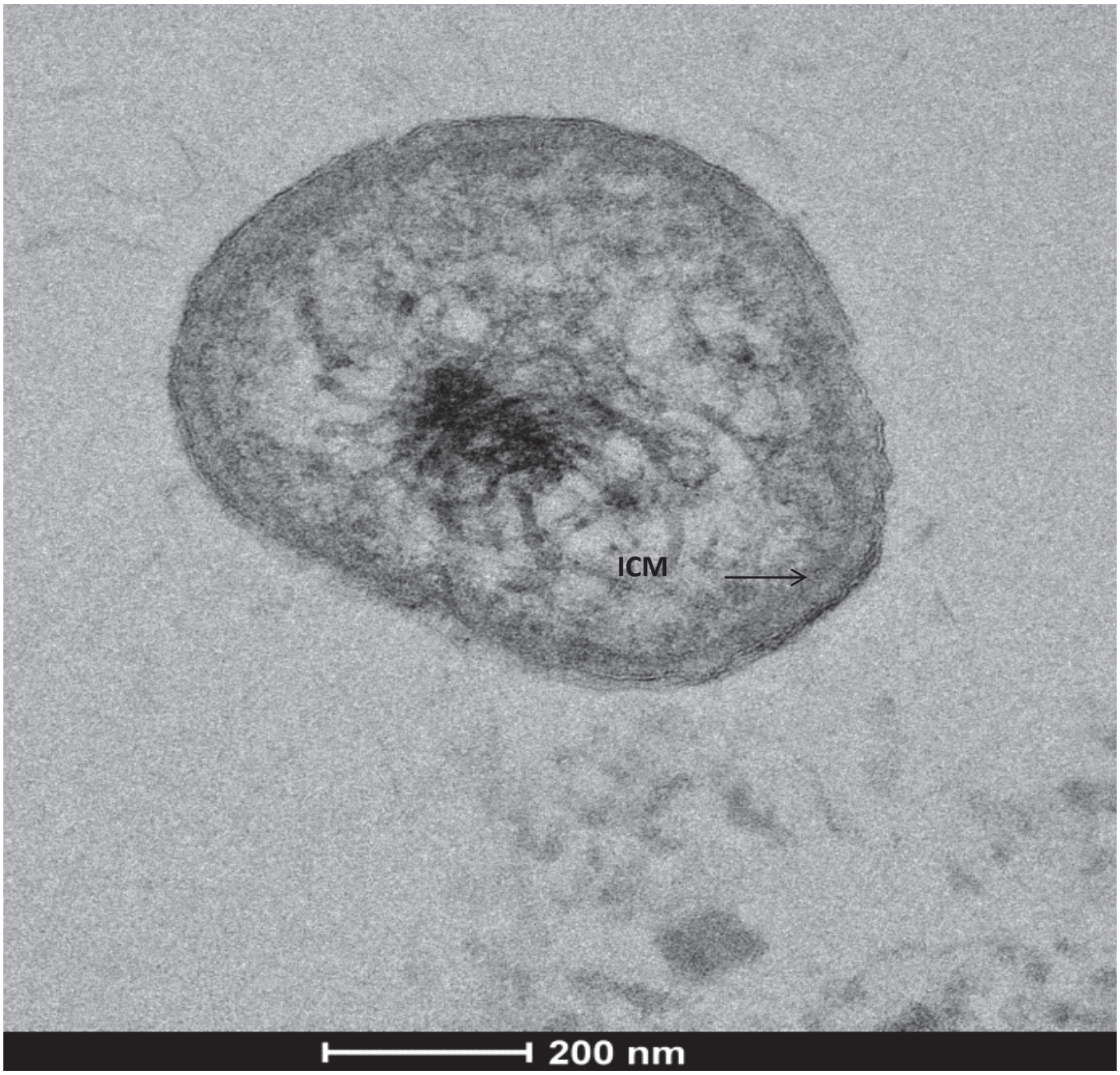


Figure 1.c

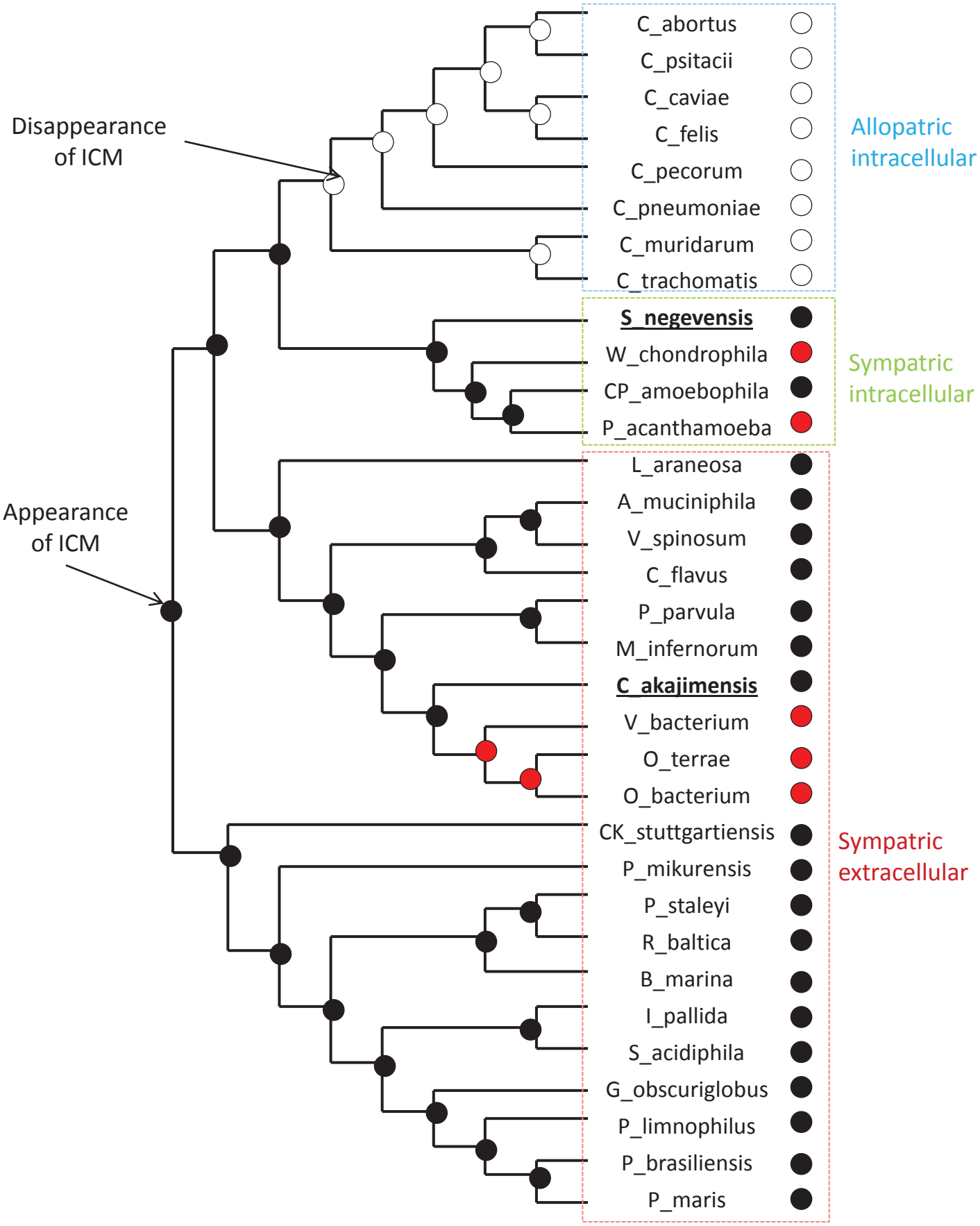
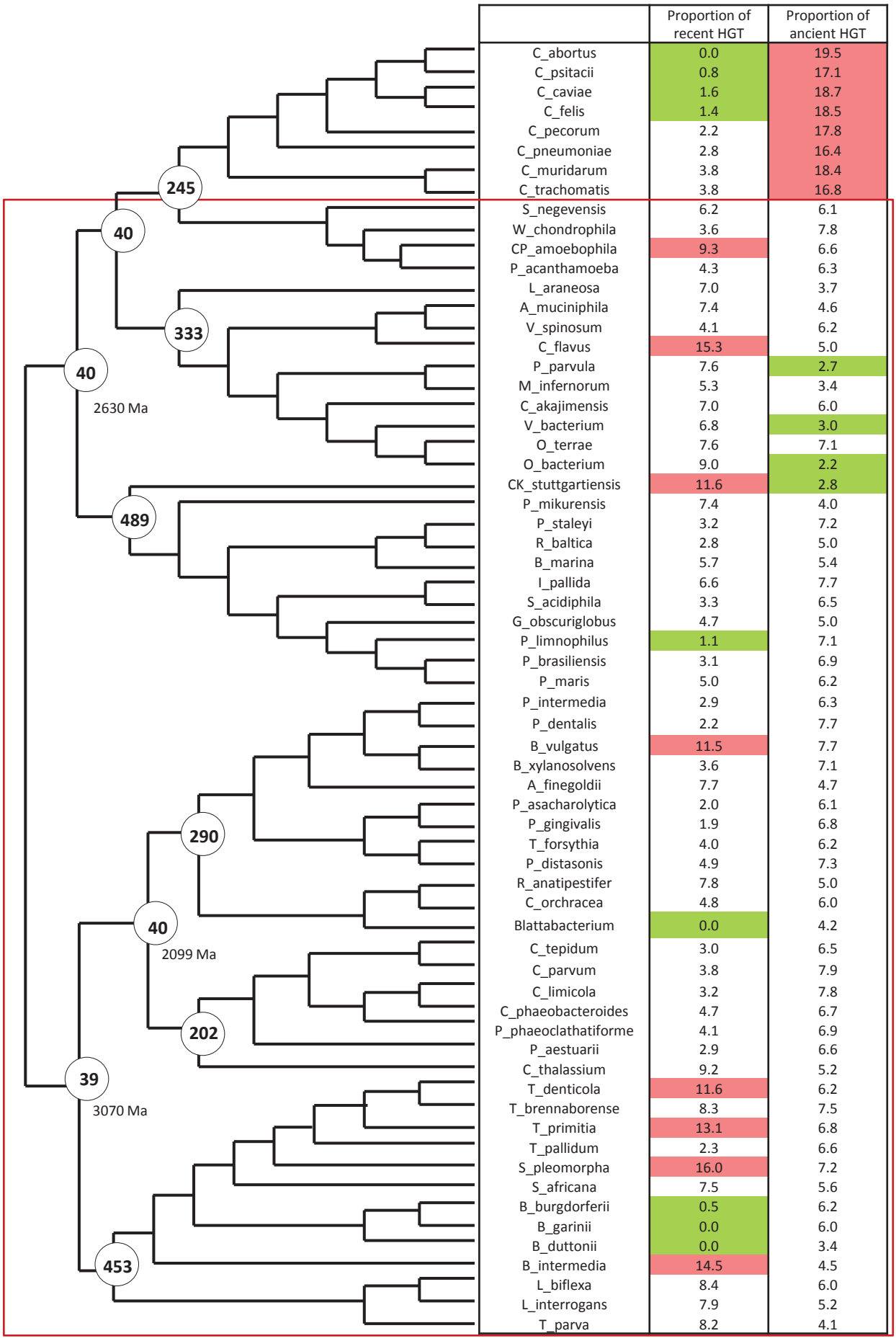

Disappearance of ICM
Appearance of ICM
C_abortus
C_psitacii
C_caviae
C_felis
C_pecorum
C_pneumoniae
C_muridarum
C_trachomatis
Allopatric intracellular
S_negevensis
W_chondrophila
CP_amoebophila
P_acanthamoeba
Sympatric intracellular
L_araneosa
A_muciniphila
V_spinosum
C_flavus
P_parvula
M_infernorum
C_akajimensis
V_bacterium
O_terrae
O_bacterium
CK_stuttgartiensis
P_mikurensis
P_staleyi
R_baltica
B_marina
I_pallida
S_acidiphila
G_obscuriglobus
P_limnophilus
P_brasiliensis
P_maris
Sympatric extracellular

Figure 2



	Proportion of recent HGT	Proportion of ancient HGT
C_abortus	0.0	19.5
C_psitacii	0.8	17.1
C_caviae	1.6	18.7
C_felis	1.4	18.5
C_pecorum	2.2	17.8
C_pneumoniae	2.8	16.4
C_muridarum	3.8	18.4
C_trachomatis	3.8	16.8
S_negevensis	6.2	6.1
W_chondrophila	3.6	7.8
CP_amoebophila	9.3	6.6
P_acanthamoeba	4.3	6.3
L_araneosa	7.0	3.7
A_muciniphila	7.4	4.6
V_spinosum	4.1	6.2
C_flavus	15.3	5.0
P_parvula	7.6	2.7
M_infernorum	5.3	3.4
C_akajimensis	7.0	6.0
V_bacterium	6.8	3.0
O_terrae	7.6	7.1
O_bacterium	9.0	2.2
CK_stuttgartiensis	11.6	2.8
P_mikurensis	7.4	4.0
P_staleyi	3.2	7.2
R_baltica	2.8	5.0
B_marina	5.7	5.4
I_pallida	6.6	7.7
S_acidiphila	3.3	6.5
G_obscuriglobus	4.7	5.0
P_limnophilus	1.1	7.1
P_brasiliensis	3.1	6.9
P_maris	5.0	6.2
P_intermedia	2.9	6.3
P_dentalis	2.2	7.7
B_vulgatus	11.5	7.7
B_xylanosolvens	3.6	7.1
A_finegoldii	7.7	4.7
P_asacharolytica	2.0	6.1
P_gingivalis	1.9	6.8
T_forsythia	4.0	6.2
P_distasonis	4.9	7.3
R_anatipestifer	7.8	5.0
C_orchracea	4.8	6.0
Blattabacterium	0.0	4.2
C_tepidum	3.0	6.5
C_parvum	3.8	7.9
C_limicola	3.2	7.8
C_phaeobacteroides	4.7	6.7
P_phaeoclathatiforme	4.1	6.9
P_aestuarii	2.9	6.6
C_thalassium	9.2	5.2
T_denticola	11.6	6.2
T_brennaborense	8.3	7.5
T_primitia	13.1	6.8
T_pallidum	2.3	6.6
S_pleomorpha	16.0	7.2
S_africana	7.5	5.6
B_burgdorferii	0.5	6.2
B_garinii	0.0	6.0
B_duttonii	0.0	3.4
B_intermedia	14.5	4.5
L_biflexa	8.4	6.0
L_interrogans	7.9	5.2
T_parva	8.2	4.1

Sympatric bacteria

Figure 3

Chapitre 2

Au vue des résultats obtenus dans notre étude précédente nous avons donc élargi notre domaine de recherche en travaillant sur l'influence des environnements et des modes de vie des bactéries sur le taux de transferts de gènes. Ces deux caractéristiques forment ce que nous avons appelé «l'habitat» des bactéries. Nous avons repris les résultats obtenus précédemment en nous concentrant uniquement sur les transferts de gènes récents (ayant donc eu lieu chez des bactéries pour lesquelles nous connaissons l'environnement). Les bactéries ont donc été triées en fonction de leur habitat, et les taux de transferts de gènes, ainsi que la fonction des gènes transférés et les partenaires de transferts, ont été comparés entre les différents habitats.

Les génomes des bactéries étudiés comprennent entre 0.2% et 1.5% de gènes acquis par transfert. Les *Chlamydia* vivant chez les métazoaires présentent un taux de gènes acquis par transfert extrêmement faible comparé aux autres bactéries. L'analyse des différents partenaires de transfert à permis de définir trois groupes distincts de bactéries : les bactéries environnementales (issues des

sols, de milieux aquatiques, des sols et des milieux aquatiques et les bactéries ubiquitaires), les bactéries vivant au sein des amibes et les bactéries issues des métazoaires (intracellulaires ou extracellulaires). Certaines fonctions de gènes transférés sont surreprésentées dans un ou plusieurs habitats, tels que le métabolisme, la transduction du signal et les mécanismes de défense. Ces fonctions pourraient donc jouer un rôle important dans l'adaptation des bactéries à leurs habitats respectifs.

Cette article est en cours de révision pour la journal Applied Environmental Biology.

Impact of habitat on bacterial genomes content

Sandrine PINOS[1,2] , Pierre PONTAROTTI[1], Didier RAOULT[2] and Vicky MERHEJ[2#]

[1] Institute of Mathematic of Marseille - UMR 7373 AMU-CNRS - Thème Evolution Biologique et Modélisation

[2] URMITE (unité de recherche sur les maladies infectieuses et tropicales émergentes) - UM63, AMU-CNRS

Running head : Impact of habitat on bacterial genomes content

[#] Corresponding author: Vicky Merhej, pharmD-PhD;

URMITE, Faculté de Médecine, 27 boulevard Jean Moulin, 13385 Marseille CEDEX 05, France.

Fax: +33 491 387 772

E-mail: vicky_merhej@hotmail.com

Keywords : Horizontal transfer, bacteria, environments, lifestyle, genomes, functions

Abstract

The genome analysis has dramatically opened wide horizons for evolutionary studies, identifying genetic events in relation with environmental conditions. To evaluate the impact of habitat of bacteria on genome evolution, we studied the genome content of 64 bacteria belonging to 6 different phyla and occupying different habitats. A comparative phylogenomic approach, phylogenetic profiling and an implicit phylogenetic analysis allowed us to determine the rate of the horizontal transfers for each studied species as well as the functional distribution of the horizontally transferred genes and the partners for transfers in the different habitats.

Our results showed that recent horizontal transfers represent 0.2 to 15% of the genome's content; bacteria from soil, from insects and ubiquitous bacteria have the largest proportions of horizontal transfers. The statistical analysis allowed classifying bacteria into groups according to their preferences for transfer partners. Some functions including metabolic functions, signal transduction and

defense mechanisms were found to be significantly overrepresented in some habitats. These functions may play an important role in the species adaptation to their habitats sources and interaction with their environment.

Background

The genome sequence of bacterial species provides insight into the interrelationship between the ecology and the evolution of these species. Environment is one of the important forces determining the evolution of bacterial genomes. Thus gene content of bacterial genomes might relate their adaptation to successive environments and/or lifestyles. Indeed, extremophile bacteria acquired genes or genetic structures implicated in essential metabolisms for survival in extreme habitats. The amino acid composition of the genome plays a major role in their adaptation to thermophilicity (1) and adaptation to acidophilic environments induced the appearance of specific metabolisms such as nitrogen fixation (2). Likewise, bacteria of ocean present different genes functions according to the nutriments available in their habitats (3) and *Phytoplasma* got complex genomic structures such as potentially mobile unities that are necessary for plants parasitism (4). In the same idea, intracellular bacteria and endosymbionts have very small genomes due to the deterioration of genes became useless in the new environment of the host cell (5-8).

More generally, some functions expression is known to characterize lifestyle of bacteria, hence transcription regulator expression differs significantly between intracellular pathogenic bacteria and free living species (9).

Among the different genetic changes allowing habitat adaptation, gain is the most interesting, due to its capacity to bring new functions to genomes. Horizontal transfer of DNA (HT) appears as a major driving force of innovation and evolution (10, 11) as it provides additional functions allowing adaptation to specific conditions and environmental changes. Many examples in the literature indicate that genes encoding for metabolic functions (12-14) represent the most transferred genes and some metabolic networks are modulated by incorporation of horizontally transferred sequences (15). Thereby, *Salmonella senftenberg* presents lots of HT implicated in sucrose production in order to adapt to its host (12). Other trials studying the mechanisms of pathogenesis showed that the acquisition of genes encoding for antibiotic resistance (16-19) and virulence (20-27) is a key step for host's colonization and

development of pathogenesis. Nevertheless transfers also depend on possibilities offered by the habitat. Several studies about specific “ecological niches” proved the influence of the environment on HT between bacteria (28). Thus, biofilms environments are known to facilitate the transfers between bacteria (29-32) and HT between bacteria seems to be more related to similarity of environment than to geographic or phylogenetic proximity (33). Environmental changing conditions are also well known factors for HT regulation, thereby, UV irradiation or starvation were shown to affect the mobility of transposons and Iss (34, 35). In naturally competent Gram-negative bacteria, the stress conditions may also enhance the quantity of HT (28, 36-39). Altogether, habitat influences HT in bacteria by two ways, it creates the necessity of functional adaptation and provides the necessary conditions for transfers.

The majority of genomic studies focused on one or few species, or on bacteria living in only one habitat (40-43) but no comparative study concerning the impact of the different existing habitats on the genomes evolution is available nowadays. In the present work, we

proposed to compare bacteria belonging to two sister super-phyla but from different habitats, in order to evaluate the real impact of habitat on bacterial genomes contents and their organization, with a special interest for HT. We wanted to verify if the genome sequence permits to reconstruct the habitat of species, as it is supported by reverse ecology theory (44). Therefore, we have included the genomes of 64 bacteria from various habitats. We considered that the environment i.e. water, soil, insects... and lifestyle i.e. isolated bacteria (such as intracellular bacteria) or bacteria living in contact with other microorganisms, are components of one bacterial characteristic, the habitat. We used a strategy based on comparative phylogenomics in order to determine genes specific to the different habitats and genes acquired by horizontal transfer. Statistical analyses of the partners and functions involved in the HT allowed us to evaluate the real impact of habitat on the genomes evolution.

Materials and methods

Bacterial set selection, definition of lifestyles and environments

The genomes of 64 bacteria have been retrieved from NCBI (http://www.ncbi.nlm.nih.gov/) and (http://mirrors.vbi.vt.edu/mirrors/ftp.ncbi.nih.gov/genomes/Bacteria/). These bacteria belong to six different phyla: The three phyla of PVC superphylum, *Planctomycetes*, *Verrucomicrobiae* (*Lentisphaerae*) and *Chlamydia* and the phylogenetically closest phyla, *Bacteroidetes, Chlorobi* and *Spirochaeta*. The phylogenetic relationships among bacteria were determined thanks to a reference tree built with 44 phylogenetic markers for 99 bacteria published in 2012 by Kamneva and al (45). We reconstructed the species tree of PVC bacteria and Bacteroidetes-Chlorobi-Spirochaeta thanks to 16S and 23S RNA using reference alignment and Mega5 (46) software. The studied bacteria have different lifestyles (intracellular or extracellular, allopatric or sympatric (47, 48)) and live in different environments (amoeba, mammals, soils, water, insects). Lifestyles and living environment were defined for each bacteria thanks to literature search (49-98)

and databases ((JGI (http://genome.jgi.doe.gov/), GOLD (https://gold.jgi-psf.org/index), LPSN (http://www.bacterio.net/index.html), NCBI (http://www.ncbi.nlm.nih.gov/taxonomy and http://www.ncbi.nlm.nih.gov/genome) (99-101)). We considered only the main habitat of species.

Genomes analysis: core gene, specific genes, ORFans

OrthoMCL (102) was used for construction of orthologous groups. Genes that do not belong to any orthologous groups are either acquired by recent HT or generated *de novo*. Blast against NR database allowed the identification of ORFans in the genome with no identifiable homologous genes (e-value<10-4 and coverage>50%). Among orthologous groups constructed by OrthoMCL, we identified those that are specific to one habitat i.e. groups containing only proteins of bacteria from the same habitat (present at least in 2 or 3 genomes for habitats containing more than 6 bacteria, as water, Extracellular and intracellular bacteria of vertebrates). We calculated the rate and determined the function of proteins that are specific to

habitat in each species using three different softwares, COGnitor, WGA and Interpro (103-106).

HT detection, functions and partners identification

The horizontal transfers were identified using a comparative phylogenomic approach, phylogenetic profiling of proteins and phylogenetic analysis of gene trees in comparison with species tree. We used Phylopattern (107) pipeline to study the genetic events in four steps: **1**, Comparison between species tree and trees constructed on the basis of each orthologous groups; **2**, Detection of missing species in orthologous groups; **3**, Reconstruction of ancestral state for each proteins; **4**, Identification of two types of genetic events: gains and losses. The gains could be a possible ancient HT, recent HT (occurred in one or several studied species), *de novo* genes, or artifacts. We focused on recent gene gains and Blast (108) each sequence against NR database. For each blast result, if the first ten hits belong to another phylum than that of the query and were validated as ortholog by a reciprocal best hit, the enquired gene could be considered as horizontally acquired. Sequences with e-value

> 10-5, coverage < 60% or identities < 30% were not considered. We calculated the number of HT, the percentage of proteins present in each species due to HT and the length of horizontally transferred sequences. We determined with which organism the transfer was realized thanks to the Blast results and we identified the function of transferred genes by using three different softwares, COGnitor, WGA and Interpro (103-106). These softwares allow to attribute one type of function to proteins, according the COG to which proteins belong. We studied the possible significant difference of HT distribution among the different studied groups of bacteria concerning the HT partners and functions. Therefore, the difference was estimated thanks to a comparison of variance (tests of normality followed by Kruskal Wallis (109) or ANOVA tests and Tukey (110) or Nemenyi (111) tests). We realized a Principal Components Analysis (PCA) (only with variables identified as differentially distributed by analysis of variance) followed by a hierarchical clustering (HCPC) to identify clustering of bacteria according to their transfer partners or their functions. We looked for a possible relation between clustering and

habitats. All analyzes are realized with R software. The statistical analyzes were completed by correlation tests (Pagel test with matrix transformation and Correlation test with Spearman coefficient on R software). For each function, we determined if the different proportions observed vary according to the bacterial habitat.

Results

Core gene

The genome size varied widely among the 64 studied bacteria ranging from 0.63 Mb for *Blattabacterium* to 9.76 Mb for *Singulisphaera acidiphila* with an average of 3.83 +/- 2.2. Bacteria of soils, insects and water-soils presented the largest genome sizes with an average of 7.07 Mb, 6.45 Mb and 5.69 Mb respectively, while the smallest genome sizes were found among intracellular and extracellular bacteria of vertebrates and bacteria from amoeba with an average of 1.14 Mb, 2.56 Mb and 2.91 Mb, respectively. Ubiquitous bacteria and bacteria from water with an average genomes size of 4.63 Mb, formed medium genomes. Likewise, the

proteomes contents were very different among the studied bacteria, ranging from 579 proteins for *Blattabacterium* sp to 7,969 for *Gemmata obscuriglobus*, with an average of 3,227 +/- 2,513 (Table S1). The 206,508 proteins of the 64 proteomes could be assigned to 16,918 orthologous groups (OG) using OrthoMCL. Among these, 1,224 OGs were common to, at least one bacterium from six of the eight different habitats and constituted the core gene. We could infer a putative function to 74% of the proteins from the core gene (19.5% for storage and processing information, 36.8% metabolism and 43.8% cellular processes and signaling) while 24% are poorly characterized. Energy production and conversion with 8.5%, cell wall/membrane/envelop biogenesis with 12.4%, transcription with 9.1% and signal transduction mechanisms with 11.6% were significantly over-represented functions among the proteins of the core gene (Chi2 test: p-value = 9.4*10-7) (Figure 1).

Specific genes

The core gene represented an average of 26.2% of the genome content and varied from 20.1% for the bacteria of insects to 49.5 %

for the intracellular vertebrates (Figure 1). This wide difference was due to the set of genes that are specific to each environment. Bacteria of insects showed the highest proportion of their genes shared with the other groups and the lowest quantity of specific genes (29 genes) (Figure 2). The functional distribution of the specific genes was significantly different from that of the core genes in all the habitats with significantly less genes implicated in cell process and signaling (20% to 35%) (t-test ; p-value=4.3*10-4). Thus, among the specific genes with clearly identified function, we found a significantly high proportion of genes dedicated to information storage (31.6%) in bacteria from amoeba (Kruskal-Wallis test : p-value = 2.5*10-4) ; the genes specific to bacteria from insects were significantly more dedicated to metabolism (64.8%) and less to cell process (17.8%) (Kruskal-Wallis test: p-value = 1.8*10-4 and 2.0*10-5) (Figure 2 B).

The analysis of the functions of specific genes showed that some functions were significantly overrepresented in some habitats compared to other habitats. Thus, genes encoding for proteins involved in the amino acid transport and metabolism (15.6%) and the

coenzyme transport and metabolism (13.3%) were overrepresented in the specific set of genes of bacteria from insects (Kruskal-Wallis test: p-value = 1.0*10-2 and ANOVA test : p-value=1.1*10-7), genes encoding for the signal transduction mechanisms (15.7%) and defense mechanisms (8.1%) were significantly overrepresented in the set of 185 specific genes found for intracellular bacteria of vertebrates (Kruskal-Wallis test : p-value = 2.2*10-5 and 1.8*10-6 ; Correlation test : p-value= 5.5*10-2 and 3.1*10-3), genes encoding for transcription (16.3%) were overrepresented in the set of 178 specific genes of amoeba (Kruskal-Wallis and Correlation test : p-value = 3.9*10-6 and 4.7*10-3) and genes encoding for coenzyme transport and metabolism (9.7%) were overrepresented in the set of 95 specific genes of bacteria of soils (ANOVA test: p-value = 1.1*10-7) and genes encoding for transcription were over represented (18.8%) in bacteria from soil-water compared to the other habitats (Kruskal-Wallis and Correlation test : p-value = 3.9*10-6 and 4.7*10-3) (Figure 2). The ubiquitous bacteria with 556 genes have the significantly highest proportion of specific genes in average (Chi2: p-value =

2.2*10-16). However, the functional distribution of these genes didn’t reveal a particular function with a significant overrepresentation compared to the other habitats. Likewise, bacteria from water and extracellular bacteria of vertebrates present also a high proportion of specific genes with 345 and 347 specific proteins, respectively but no category of functions was significantly over-represented among these habitats compared to the others (Figure 2). Specific genes could be due to two different mechanisms: acquisition of gene in an ancestor followed by gene loss in all descents, except in bacteria living in one habitat; or an Horizontal Transfer (HT) that occurred only in one environment. Among the specific genes, between 1.1% (extracellular bacteria of vertebrates) and 8.6% (bacteria of soils) are probable HT.

Finally, we observed a low and non-significant variation in ORFans proportion. Intracellular bacteria from vertebrates presented the smallest proportion of ORFans (3.4%); the highest proportions of ORFans were observed in bacteria of soils, amoeba and water-soils (7.7% - 7.6%) (Figure 1). It is noteworthy that *Rhodopirellula Baltica*

harbored the largest percentage of ORFans in genomes (11.0%) and *Borrelia burgdorferii* the smallest percentage (0.9%).

Horizontal transfers

Globally, we counted 10,922 events of recent HT detected in the studied bacteria thanks to the phylogenetic profile analysis, followed by an implicit phylogenetic analysis. This genes correspond to 11,866,799 bp (4,9% of the totality of sequences). Thus, 12,885 proteins (6,3 % of all the proteins) were subject to HT. The incidence to which the HT events occurred was as high as 202.6 +/- 138.3 yet highly variable among bacteria ranging from 0 transferred protein in the genome of *Blattabacterium* and *Borrelia* spp. to 803 proteins in the genome of *Chthoniobacter flavus* (Figure 3). Bacteria from soil, from insects and ubiquitous bacteria presented the highest average of proteins transferred (624.3, 418.3 and 311.3 proteins/species, respectively), compared to extracellular bacteria of vertebrates (152.9/species), bacteria from amoeba (133.5/species) and intracellular bacteria of vertebrates (20 transferred proteins /species). Species living in water present an intermediate quantity

with 214.5 proteins transferred by proteome. The transferred fragments presented a mean length of 1,124+/- 274 bp yet very heterogeneous in the studied bacteria, ranging from 859 bp for *C. tepidum* to 1,492 pb for *S. acidophila*. However, the size of transferred sequences was closely similar among the different habitats (1022 +/- 383 bp). The proportion of the transferred sequence in the genome was variable among the studied bacteria. The statistic comparison of the bacterial group from different habitats allowed to define three classes thanks to percentages of sequences due to transfer (Kruskal-Wallis test : p-value = 2.3*10-2): Bacteria of soils (8.4%) and insects (7.3%) in one class, bacteria from amoeba (4.8%), water (4.9%) and extracellular bacteria of vertebrates (4.6%) presented similar proportions and formed the second class. Intracellular bacteria of vertebrates were grouped in the third class, with 1.9% of sequence transferred.

These genetic exchanges were mostly realized with bacteria (92.5%) but we observed also some exchanges with *Archaea* (2.43%), *Eukaryota* (1.8%) and viruses (0.2%). The frequency of partners

varied significantly for bacteria from amoeba (86.6% of exchanges with bacteria, 9.8% with Eukaryotes and 3.1% with *Archaea*), for water-soils bacteria (95.5%, 0.9% and 3.4%), for ubiquitous bacteria (92.8%, 4.2% and 2.4%) and bacteria from water (95.5%, 0.9% and 3.4%) where the exchanges of sequences were significantly more frequent with Eukaryotes or *Archaea* than in bacteria from insects, soils, extracellular and intracellular bacteria of vertebrates, for which more than 90% of exchanges were with bacteria, less than 1% with *Eukaryota* and less than 2% with *Archaea*) (Kruskal test : p-value= 3.3*10-3 and 3.1*10-2). Finally, it is noteworthy that bacteria from amoeba showed a significantly high amount of HT realized with eukaryotes, plants in particular, compared to bacteria of other habitats.

Among the different bacterial partners, the most common were *Proteobacteria* (41.5%), *Firmicutes* (23.15%), *Cyanobacteria* (6.3%), *Actinobacteria* (5.2%), *Fusobacteria* (1.7%), *Chloroflexi* (1.5%) and *Acidobacteria* (1.3%) (Figure 4). This partners' distribution was respected for bacteria from amoeba, soils, water-soils and bacteria

from water. However, bacteria from insects and ubiquitous bacteria exchanged in majority with *Proteobacteria*, in second with *Firmicutes* and in third with *Actinobacteria* (6.6% and 10.9% with *Actinobacteria*, respectively). *Gammaproteobacteria* spp. were the most represented transfer partners for bacteria from insects (14.7%), water (17.2%) and water-soils (11.3%) (Figure 4). Bacteria from intracellular vertebrates exchanged in majority with *Firmicutes* (21.4%), in second *Proteobacteria* (17.6%) and third with *Cyanobacteria* (6%). Finally bacteria from extracellular vertebrates had a significantly higher proportion of exchanges with *Firmicutes* (41.1%) and lesser with *Proteobacteria* (30.9%) than the other habitats. Bacteria from intracellular and extracellular vertebrates have not the same distribution but are both characterized by their preference for the *Firmicutes* as transfer partners and their significantly low proportion of transfers with *Actinobacteria* (Kruskal test : p-value= 3.6*10-3 and ANOVA test : p-value=3.8*10-2). The Principal components analysis and hierarchical clustering showed a relationship between bacteria habitats and partners transfers (Correlation test : p-value=7.9*10-

29). Clustering determined three groups of bacteria according to partners: the environmental bacteria in a first group (soils, water, soils-water and Ubiquitous bacteria), bacteria from Amoeba in a second group and the intracellular and extracellular bacteria of vertebrates in the third group. These results suggested that HT partners can be used to infer bacterial habitats.

The general functional distribution in the HT detected for the different habitats was not similar to that of the core genes. Genes involved in cell process and signaling (33 to 50%) seems to be significantly more subject to HT whereas genes dedicated to information storage (12% to 17%) were less subject to HT (t-test : p-value = 4.1*10-2 and 1.7*10-3). Using a strategy based on comparative phylogenomic, we found that there were also significant differences among the habitats. Bacteria from soils-water presented significantly bigger proportion of genes dedicated to cell process and signaling (50.2%) than bacteria from other habitats (Kruskal-Wallis test : p-value = 2.5*10-2). Differences between habitats were even more visible in studying the sub-categories of functions. HT in

bacteria from amoeba were characterized by the over representation of the amino acid transport and metabolism function (16%) (Kruskal-Wallis test and correlation test: p-value = 7.5*10-2 and 2.1*10-4). HT in ubiquitous bacteria were characterized by the over representation of two functions, the lipid transport and metabolism (10.5%) and the signal transduction mechanism function (20.2%) (ANOVA test: p-value= 6.9*10-3 and 2.3*10-4); the later function was also overrepresented in bacteria from soils (17.9 %). HT from extracellular vertebrates were significantly more implicated in defense mechanism (4.7%) (Kruskal-Wallis: p-value=3.5*10-2). Altogether, the analysis of HT sequences allowed the identification of genes/functions that are specific to some habitats. The statistical significance of some HT events that are concomitant with a habitat demonstrated that the occurrence of HT was not due to hazard. Some of these genes were acquired via horizontal transfer according to a strategy of colonization and adaptation to the environment.

Discussion

Our comparative phylogenomic analysis of 64 bacteria revealed the influence of the living environment on bacterial genome composition. Genes that are specific to each environment have been identified on the basis of orthologous analysis; they seem to be selected according to their functions, with preferential functions different in each habitat. This first approximation to study evolution in relation with the environment has been completed by phylogenetic profiling and phylogenetic analysis that identified the horizontal transfer events, described the functional profile of transferred genes and demonstrated the role played by the habitat in the choice of partners for transfer. Our findings were in agreement with previous studies (45, 125, 126), which validate our method as a powerful approach for the identification of HT and potential transfer partners.

The proportion of HT in species depends on the capacity to exchange with other microorganisms. Our findings indicate that intracellular bacteria of vertebrates have low percentages (1.9%) of

horizontally transferred sequences compared to the other bacteria (between 4.6% and 8.4%). This is probably due to the fact that intracellular bacteria are physically isolated from other microorganisms, which prevents the opportunities of HT (112). This agrees with previous studies showing that the predominant evolutionary process in intracellular bacteria is genome reduction leading to small genome sizes (113 114). Intracellular bacteria of amoeba with 5.8% of HT are the exception (115, 116). Indeed, amoeba can phagocytes several bacteria at once giving a particular field for potential genetic exchange moreover it can be a training ground for the emergence of parasitism (115). Finally, the low proportion of specific genes that has been detected in bacteria from insects, soils and soils-water (0.6% to 1.6%), compared to other bacteria (7.5% to 18.4%) is rather due to the species richness and phylogenetic relationships (117, 118). Indeed, in the bacterial groups representing the habitats insects, soils and soils-water, species richness are lower than in other groups (except for bacteria from

Amoeba and Ubiquitous), and species are phylogenetically distant (conversely to bacteria from Amoeba and Ubiquitous).

Some functions were found to be significantly overrepresented among the transferred genes in some habitats. These overrepresented functions included two metabolic functions the amino acid and lipid metabolism and transport, signal transduction and defense mechanisms. These functions may play an important role in the species adaptation to their habitats sources and interaction with their environment. Moreover, these findings indicate that selection of the transferred genes is rather specific to the different habitats and seems to be biased by the encoded functions. Numerous studies indicate that the genes identified for HT encode for metabolic mechanisms and other functions that enhance pathogenicity like genes for virulence and antimicrobial activity (11, 15-27, 119). Informational genes (those involved in transcription, translation, and related processes) were considered to be less prone to lateral transfer. However, an automatic analysis of 116 bacterial genomes showed that horizontal transfer occurred in genes involved

in cell surface and DNA binding besides the pathogenicity-related functions (120). A recent genomic study revealed that genes essential for transcription and translation had also experienced multiple lateral transfers in the genome of a *Rickettsia* species (48). Our findings confirm the random character of lateral transfers that can affect any function and support that selection is rather specific of the habitat and acts in order to promote adaptation to the environment.

The analysis of the microbial partners allowed the identification of two distinct groups: the environmental bacteria that have exchanged especially with the *Proteobacteria* on one hand and bacteria from microbiomes that have exchanged especially with the *Firmicutes* on the other hand. It has been already assumed that HT occurred preferentially between bacteria from the same habitats. *Firmicutes* are one of the two major phyla present in gut micorbiome (5, 128) and this is the main partner of our bacteria from vertebrates. In the same way, *Acidobacteria* are mainly detected in soil (122) and they are overrepresented as HT partners of bacteria from soils compared to bacteria from other habitat. In a previous study, it was

demonstrated that the habitat plays an important role in determining the composition of microbial genomes. Thus genome analysis of bacteria from different habitats allowed the formation of two groups: the environmental bacteria of ocean, soil and stromatolites one hand and bacteria from microbiomes (human and mouse) on the other hand (121). Our study showed that habitat influences the flux of horizontal transfers and subsequently the microbial composition of the genomes. The tendency of bacteria from Amoeba to exchange more with Eukaryotes, especially plants, is probably due to their ancestral habitats. Indeed, ancestral *Chlamydiae* are known to have lived in and acquired genes from plants (123, 124). Thus, we can support that one part of HT detected are ancient HT acquired by the interaction between the ancestors of the *Chlamydiae* and the plants, followed by the loss in the majority of bacteria. It is worthy to note that like previous studies for HT detections, it is difficult to distinguish between ancient and recent HT; yet HT partners are the witnesses of modern and ancestral habitats of the studied bacteria.

Thanks to this study, we know that the habitat influences the bacterial genome contents, especially the HT selection. However, in our set of bacteria, we have only few representatives of soil bacteria and three bacteria from insects or ubiquitous. This is low compared to the diversity of bacteria and environments existing. The ever increasing number of sequenced genomes will help to improve the estimation of HT. Indeed, the introduction of additional genomes of different bacterial species to have several representatives of each environment might be very interesting to establish a more reliable evaluation of the impact of the environment on the genome evolution.

Funding Information

This work was supported by the Assistance Publique - Hopitaux de Marseille (Marseille Public University Hospital System). The funders had no role in study design, data collection and interpretation, or the decision to submit the work.

Acknowledgements

We thank Olivier Chabrol for assistance in computer programming during the elaboration of HT detection strategy and Manuela Royer Carenzi for the assistance in statistical analyzes. We also thank the Xegen company for the assistance in HT detection by using of Phylopattern software.

The authors declare that they have no competing interests

References

1.Kreil, D.P. and Ouzounis, C.A. 2001. Identification of thermophilic species by the amino acid compositions deduced from their genomes. Nucleic Acids Res 29:1608– 1615
2.Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., ... & Banfield, J. F. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428(6978):37-43.
3.Lauro, F. M., McDougald, D., Thomas, T., Williams, T. J., Egan, S., Rice, S., ... & Cavicchioli, R. 2009. The genomic basis of trophic strategy in marine bacteria. Proceedings of the National Academy of Sciences 106(37):15527-15533.
4.Bai, X., Zhang, J., Ewing, A., Miller, S. A., Radek, A. J., Shevchenko, D. V., ... & Hogenhout, S. A. 2006. Living with genome instability: the adaptation of phytoplasmas to diverse environments of their insect and plant hosts. Journal of Bacteriology 188(10):3682-3696.
5.Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., ... & Weissenbach, J. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464(7285):59-65.
6.Ochman, H. and Moran, N.A. 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science 292:1096– 1099
7.Andersson, J.O. and Andersson, S.G. 2001. Pseudogenes, junk DNA, and the dynamics of Rickettsia genomes. Mol. Biol. Evol. 18:829 – 839

8.Tamas, I., Klasson, L. M., Sandström, J. P., & Andersson, S. G. 2001. Mutualists and parasites: how to paint yourself into a (metabolic) corner. FEBS letters 498(2):135-139.
9.Ildefonso Cases, Victor de Lorenzo and Christos A. Ouzounis. 2003. Transcription regulation and environmental adaptation in bacteria. TRENDS in Microbiology 11:248-253.
10.Le, P. T., Ramulu, H. G., Guijarro, L., Paganini, J., Gouret, P., Chabrol, O., ... & Pontarotti, P. 2012. An automated approach for the identification of horizontal gene transfers from complete genomes reveals the rhizome of Rickettsiales. BMC evolutionary biology 12(1):243.
11.Ochman, H., Lawrence, J. G., & Groisman, E. A. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405(6784):299-304.
12.Jörg Hacker & Elisabeth Carniel. 2001. Ecological fitness, genomic islands and bacterial pathogenicity : A Darwinian view of the evolution of microbes. EMBO reports 21:376–381
13.Viswanathan, V. K. 2012. Spirochaeta and their twisted ways. Gut microbes 3(5):399-400.
14.Bellgard MI, Wanchanthuek P, La T, Ryan K, Moolhuijzen P, Albertyn Z. 2009. Genome sequence of the pathogenic intestinal spirochete brachyspira hyodysenteriae reveals adaptations to its lifestyle in the porcine large intestine. PLoS One 4:e4641;
15.Pál, C., Papp, B., & Lercher, M. J. 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nature genetics 37(12):1372-1375.
16.Kleckner, N. in Mobile DNA (eds Berg, D. E. & Howe, M. M.) 227±268 (American Society for Microbiology, Washington DC, 1989).
17.Baquero, F., Alvarez-Ortega, C., & Martinez, J. L. 2009. Ecology and evolution of antibiotic resistance. Environmental Microbiology Reports 1(6):469-476.
18.Mazel, D., Dychinco, B.,Webb, V. A. & Davies, J. 1998. A distinctive class of integron in the Vibrio cholera genome. Science 280:605±608
19.Barlow M. 2009. What antimicrobial resistance has taught us about horizontal gene transfer. Methods in Molecular Biology 532: 397–411.
20.Portnoy, D. A., Moseley, S. L. & Falkow, S. 1981. Characterization of plasmids and plasmid-associated determinants of Yersinia enterocolitica pathogenesis. Infect. Immun 31:775±782
21.Maurelli, A. T., Baudry, B., d'Hauteville, H., Hale, T. L. & Sansonetti, P. J. 1985. Cloning of plasmid DNA sequences involved in invasion of HeLa cells by Shigella ¯exneri. Infect. Immun 49:164±171
22.Sasakawa, C., Kamata, K., Sakai, T., Makino, S., Yamada, M., Okada, N., & Yoshikawa, M. 1988. Virulence-associated genetic regions comprising 31 kilobases of the 230-kilobase plasmid in Shigella flexneri 2a. Journal of bacteriology 170(6):2480-2484.
23.Gemski, P., Lazere, J. R., Casey, T. & Wohlhieter, J. A. 1980. Presence of virulence-associated plasmid in Yersinia pseudotuberculosis. Infect. Immun 28:1044±1047
24.Groisman, E. A. & Ochman, H. 1996. Pathogenicity islands: bacterial evolution in quantum leaps. Cell 87:791±794.

25.Hacker, J., Blum-Oehler, G., Muhldorfer, I. & Tschape, H. 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol. Microbiol 23:1089±1097.
26.Sun, J., Inouye, M. & Inouye, S. 1991. Association of a retroelement with a P4-like cryptic prophage (retronphage Theta;R73) integrated into the selenocystyl-tRNA gene of Escherichia coli. J. Bacteriol 173:171±181 (1991).
27.Cheetham, B. F. & Katz, M. E. 1995. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. Mol Microbiol 18:201±208.
28.Aminov, R. I. 2011. Horizontal gene exchange in environmental microbiota. Frontiers in microbiology 2.
29.Ghigo, J. M. 2001. Natural conjugative plasmids induce bacterial biofilm development. Nature 412:442–445.
30.Molin, S., and Tolker-Nielsen, T. 2003. Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. Curr. Opin. Biotechnol 14:255–261.
31.Reisner, A., Höller, B. M., Molin, S.,and Zechner, E. L. 2006. Synergistic effects in mixed Escherichia coli biofilms: conjugative plasmid transfer drives biofilm expansion. J. Bacteriol 188:3582–3588.
32.Burmølle, M., Bahl, M. I., Jensen, L. B., Sørensen, S. J., and Hansen, L. H. 2008. Type 3 fimbriae,encoded by the conjugative plasmid pOLA52, enhance biofilm formation and transfer frequencies in Enterobacteriaceae strains. Microbiology 154:187–195.
33.Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., & Alm, E. J. 2011. Ecology drives a global network of gene exchange connecting the human microbiome. Nature 480(7376):241-244.
34.Levy, M. S., Balbinder, E., and Nagel, R. 1993. Effect of mutations in SOS genes on UV-induced precise excision of Tn10 in Escherichia coli. Mutat. Res 293:241–247.
35.Ilves, H., Horak, R., and Kivisaar, M. 2001. Involvement of sigma(S) in starvation-induced transposition of Pseudomonas putida transposon Tn4652. J. Bacteriol 183:5445–5448.
36.Meibom, K. L., Blokesch, M., Dolganov, N. A., Wu, C. Y., and Schoolnik, G. K. 2005. Chitin induces natural competence in Vibrio cholerae. Science 310:1824–1827.
37.Whittle, G., Shoemaker, N. B., and Salyers, A. A. 2002. Identification of two genes involved in the modulation of conjugal transfer of the Bacteroides conjugative transposon CTnDOT. J. Bacteriol 184:3839–3847.
38.Dorer, M. S., Fero, J., and Salama, N. R. 2010. DNA damage triggers genetic exchange in Helicobacter pylori. PLoS Pathog 6:e1001026
39.Stanton, T. B., Humphrey, S. B., Sharma, V. K., and Zuerner, R. L. 2014. Collateral effects of antibiotics: carbadox and metronidazole induce VSH-1 and facilitate gene transfer among Brachyspira hyodysenteriae strains. Appl. Environ. Microbiol 74:2950–2956.
40.Maccario, L., Vogel, T. M., & Larose, C. 2014. Potential drivers of microbial community structure and function in Arctic spring snow. Frontiers in microbiology 5.

41.Haq, I. U., Zhang, M., Yang, P., & van Elsas, J. D. 2014. The interactions of bacteria with fungi in soil: emerging concepts. Advances in applied microbiology 89:185.
42.Mendes, L. W., Tsai, S. M., Navarrete, A. A., de Hollander, M., van Veen, J. A., & Kuramae, E. E. 2015. Soil-borne microbiome: linking diversity to function. Microbial ecology 1-11.
43.Besemer, K., Singer, G., Quince, C., Bertuzzo, E., Sloan, W., & Battin, T. J. 2013. Headwaters are critical reservoirs of microbial diversity for fluvial networks. Proceedings of the Royal Society of London B: Biological Sciences 280(1771):20131760.
44.Lassalle, F., Muller, D., & Nesme, X. 2015. Ecological speciation in bacteria: reverse ecology approaches reveal the adaptive part of bacterial cladogenesis. Research in Microbiology
45.Kamneva, O. K., Knight, S. J., Liberles, D. A., & Ward, N. L. 2012. Analysis of genome content evolution in PVC bacterial super-phylum: assessment of candidate genes associated with cellular organization and lifestyle. Genome biology and evolution 4(12):1375-1390.
46.Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution 28:2731-2739.
47.Georgiades, K. 2013. Genomics of epidemic pathogens. Clinical Microbiology and Infection 18(3):213-217.
48.Merhej, V., Notredame, C., Royer-Carenzi, M., Pontarotti, P., & Raoult, D. 2011. The rhizome of life: the sympatric Rickettsia felis paradigm demonstrates the random transfer of DNA sequences. Molecular biology and evolution 28(11):3213-3223.
49.Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R., ... & Venter, J. C. 1997. Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi. Nature 390(6660):580-586.
50.Hovind-Hougen, K., Birch-Andersen, A., Henrik-Nielsen, R., Orholm, M., Pedersen, J. O., Teglbjaerg, P. S., & Thaysen, E. H. 1982. Intestinal spirochetosis: morphological characterization and cultivation of the spirochete Brachyspira aalborgi gen. nov., sp. nov. Journal of Clinical Microbiology 16(6):1127-1136.
51.Ren, S. X., Fu, G., Jiang, X. G., Zeng, R., Miao, Y. G., Xu, H., ... & Zhao, G. P. 2003. Unique physiological and pathogenic features of Leptospira interrogans revealed by whole-genome sequencing. Nature 422(6934):888-893.
52.Picardeau, M., Bulach, D. M., Bouchier, C., Zuerner, R. L., Zidane, N., Wilson, P. J., ... & Adler, B. 2008. Genome sequence of the saprophyte Leptospira biflexa provides insights into the evolution of Leptospira and the pathogenesis of leptospirosis. PLoS One 3(2):e1607.
53.Nascimento, A. L. T. O., Ko, A. I., Martins, E. A. L., Monteiro-Vitorello, C. B., Ho, P. L., Haake, D. A., ... & Van Sluys, M. A. 2004. Comparative genomics of two Leptospira interrogans serovars reveals novel insights into physiology and pathogenesis. Journal of bacteriology 186(7):2164-2172.
54.Ritalahti, K. M., Justicia-Leon, S. D., Cusick, K. D., Ramos-Hernandez, N., Rubin, M., Dornbush, J., & Löffler, F. E. 2012. Sphaerochaeta globosa gen. nov., sp. nov. and

Sphaerochaeta pleomorpha sp. nov., free-living, spherical Spirochaeta. International journal of systematic and evolutionary microbiology 62(1):210-216.
55.Schrank, K., Choi, B. K., Grund, S., Moter, A., Heuner, K., Nattermann, H., & Göbel, U. B. 1999. Treponema brennaborense sp. nov., a novel spirochaete isolated from a dairy cow suffering from digital dermatitis. International journal of systematic bacteriology 49(1):43-50.
56.Fraser, C. M., Norris, S. J., Weinstock, G. M., White, O., Sutton, G. G., Dodson, R., ... & Venter, J. C. 1998. Complete genome sequence of Treponema pallidum, the syphilis spirochete. Science 281(5375):375-388.
57.Seshadri, R., Myers, G. S., Tettelin, H., Eisen, J. A., Heidelberg, J. F., Dodson, R. J., ... & Paulsen, I. T. 2004. Comparison of the genome of the oral pathogen Treponema denticola with other spirochete genomes. Proceedings of the National Academy of Sciences of the United States of America 101(15):5646-5651.
58.Graber, J. R., Leadbetter, J. R., & Breznak, J. A. 2004. Description of Treponema azotonutricium sp. nov. and Treponema primitia sp. nov., the first spirochetes isolated from termite guts. Applied and environmental microbiology 70(3):1315-1320.
59.Levett, P. N., Morey, R. E., Galloway, R., Steigerwalt, A. G., & Ellis, W. A. 2005. Reclassification of Leptospira parva Hovind-Hougen et al. 1982 as Turneriella parva gen. nov., comb. nov. International journal of systematic and evolutionary microbiology 55(4):1497-1499.
60.McGroty, S. E., Pattaniyil, D. T., Patin, D., Blanot, D., Ravichandran, A. C., Suzuki, H., ... & Hudson, A. O. 2013. Biochemical characterization of UDP-N-acetylmuramoyl-L-alanyl-D-glutamate: meso-2, 6-diaminopimelate ligase (MurE) from Verrucomicrobium spinosum DSM 4136T. PloS one 8(6):e66458.
61.Sait, M., Kamneva, O. K., Fay, D. S., Kirienko, N. V., Polek, J., Shirasu-Hiza, M. M., & Ward, N. L. 2011. Genomic and experimental evidence suggests that Verrucomicrobium spinosum interacts with eukaryotes. Frontiers in microbiology 2.
62.Kant, R., Van Passel, M. W., Sangwan, P., Palva, A., Lucas, S., Copeland, A., ... & Smidt, H. 2011. Genome sequence of Pedosphaera parvula Ellin514, an aerobic verrucomicrobial isolate from pasture soil. Journal of bacteriology.
63.Mavromatis, K., Abt, B., Brambilla, E., Lapidus, A., Copeland, A., Deshpande, S., ... & Kyrpides, N. C. 2010. Complete genome sequence of Coraliomargarita akajimensis type strain (04OKA010-24T). Standards in genomic sciences 2(3):290.
64.Yoon, J., Yasumoto-Hirose, M., Katsuta, A., Sekiguchi, H., Matsuda, S., Kasai, H., & Yokota, A. 2007. Coraliomargarita akajimensis gen. nov., sp. nov., a novel member of the phylum 'Verrucomicrobia'isolated from seawater in Japan. International journal of systematic and evolutionary microbiology 57(5):959-963.
65.Kant, R., Van Passel, M. W., Palva, A., Lucas, S., Lapidus, A., del Rio, T. G., ... & Smidt, H. 2011. Genome sequence of Chthoniobacter flavus Ellin428, an aerobic heterotrophic soil bacterium. Journal of bacteriology.

66.Thrash, J. C., Cho, J. C., Vergin, K. L., Morris, R. M., & Giovannoni, S. J. 2010. Genome sequence of Lentisphaera araneosa HTCC2155T, the type species of the order Lentisphaerales in the phylum Lentisphaerae. Journal of bacteriology 192(11):2938-2939.
67.Cho, J. C., Vergin, K. L., Morris, R. M., & Giovannoni, S. J. 2004. Lentisphaera araneosa gen. nov., sp. nov, a transparent exopolymer producing marine bacterium, and the description of a novel bacterial phylum, Lentisphaerae. Environmental Microbiology 6(6):611-621.
68.Sharp, C. E., Smirnova, A. V., Graham, J. M., Stott, M. B., Khadka, R., Moore, T. R., ... & Dunfield, P. F. 2014. Distribution and diversity of Verrucomicrobia methanotrophs in geothermal and acidic environments. Environmental microbiology 16(6):1867-1878.
69.Hou, S., Makarova, K. S., Saw, J. H., Senin, P., Ly, B. V., Zhou, Z., ... & Alam, M. 2008. Complete genome sequence of the extremely acidophilic methanotroph isolate V4, Methylacidiphilum infernorum, a representative of the bacterial phylum Verrucomicrobia. Biol Direct 3(11).
70.Isanapong, J., Goodwin, L., Bruce, D., Chen, A., Detter, C., Han, J., ... & Rodrigues, J. L. 2012. High-quality draft genome sequence of the Opitutaceae bacterium strain TAV1, a symbiont of the wood-feeding termite Reticulitermes flavipes. Journal of bacteriology 194(10):2744-2745.
71.Van Passel, M. W., Kant, R., Palva, A., Copeland, A., Lucas, S., Lapidus, A., ... & Smidt, H. 2011. Genome sequence of the Verrucomicrobium Opitutus terrae PB90-1, an abundant inhabitant of rice paddy soil ecosystems. Journal of bacteriology.
72.Derrien, M., Collado, M. C., Ben-Amor, K., Salminen, S., & de Vos, W. M. 2008. The mucin degrader Akkermansia muciniphila is an abundant resident of the human intestinal tract. Applied and environmental microbiology 74(5), 1646-1648.
73.Göker, M., Cleland, D., Saunders, E., Lapidus, A., Nolan, M., Lucas, S., ... & Klenk, H. P. 2011. Complete genome sequence of Isosphaera pallida type strain (IS1BT). Standards in genomic sciences 4(1):63.
74.Kulichevskaya, I. S., Ivanova, A. O., Baulina, O. I., Bodelier, P. L., Damsté, J. S. S., & Dedysh, S. N. 2008. Singulisphaera acidiphila gen. nov., sp. nov., a non-filamentous, Isosphaera-like planctomycete from acidic northern wetlands. International journal of systematic and evolutionary microbiology 58(5):1186-1193.
75.Van Niftrik, L., Van Helden, M., Kirchen, S., Van Donselaar, E. G., Harhangi, H. R., Webb, R. I., ... & Strous, M. 2010. Intracellular localization of membrane-bound ATPases in the compartmentalized anammox bacterium 'Candidatus Kuenenia stuttgartiensis'. Molecular microbiology 77(3):701-715.
76.Schlesner, H. 1989. Planctomyces brasiliensis sp. nov., a halotolerant bacterium from a salt pit. Systematic and applied microbiology 12(2): 159-161.
77.Hirsch, P., & Müller, M. 1985. Planctomyces limnophilus sp. nov., a stalked and budding bacterium from freshwater. Systematic and applied microbiology 6(3):276-280.

78.Bauld, J., & STALEY, J. T. 1976. Planctomyces maris sp. nov.: a marine isolate of the Planctomyces-Blastocaulis group of budding bacteria. Journal of General Microbiology 97(1): 45-55.
79.Fukunaga, Y., Kurahashi, M., Sakiyama, Y., Ohuchi, M., Yokota, A., & Harayama, S. 2009. Phycisphaera mikurensis gen. nov., sp. nov., isolated from a marine alga, and proposal of Phycisphaeraceae fam. nov., Phycisphaerales ord. nov. and Phycisphaerae classis nov. in the phylum Planctomycetes. The Journal of general and applied microbiology 55(4):267-275.
80.Clum, A., Tindall, B. J., Sikorski, J., Ivanova, N., Mavrommatis, K., Lucas, S., ... & Lapidus, A. 2009. Complete genome sequence of Pirellula staleyi type strain (ATCC 27377T). Standards in genomic sciences 1(3):308.
81.Butler, M. K., Wang, J., Webb, R. I., & Fuerst, J. A. 2002. Molecular and ultrastructural confirmation of classification of ATCC 35122 as a strain of Pirellula staleyi. International journal of systematic and evolutionary microbiology 52(5):1663-1667.
82.Schlesner, H., Rensmann, C., Tindall, B. J., Gade, D., Rabus, R., Pfeiffer, S., & Hirsch, P. 2004. Taxonomic heterogeneity within the Planctomycetales as derived by DNA–DNA hybridization, description of Rhodopirellula baltica gen. nov., sp. nov., transfer of Pirellula marina to the genus Blastopirellula gen. nov. as Blastopirellula marina comb. nov. and emended description of the genus Pirellula. International journal of systematic and evolutionary microbiology 54(5):1567-1580.
83.Glöckner, F. O., Kube, M., Bauer, M., Teeling, H., Lombardot, T., Ludwig, W., ... & Reinhardt, R. 2003. Complete genome sequence of the marine planctomycete Pirellula sp. strain 1. Proceedings of the National Academy of Sciences 100(14):8298-8303.
84.Fenner, L., Roux, V., Ananian, P., & Raoult, D. 2007. Alistipes finegoldii in blood cultures from colon cancer patients. Emerging infectious diseases 13(8):1260.
85.Takeda, K., Kenzaka, T., Morita, Y., Kuroki, S., & Kajii, E. 2013. A rare case of Lemierres syndrome caused by Porphyromonas asaccharolytica. Infection 41(4):889-892.
86.Mavrommatis, K., Gronow, S., Saunders, E., Land, M., Lapidus, A., Copeland, A., ... & Hugenholtz, P. 2009. Complete genome sequence of Capnocytophaga ochracea type strain (VPI 2845T). Standards in genomic sciences 1(2):101.
87.Huang, C. Y., Sabree, Z. L., & Moran, N. A. 2012. Genome sequence of Blattabacterium sp. strain BGIGA, endosymbiont of the Blaberus giganteus cockroach. Journal of bacteriology 194(16):4450-4451.
88.Ulsemer, P., Toutounian, K., Schmidt, J., Karsten, U., & Goletz, S. 2012. Preliminary safety evaluation of a new Bacteroides xylanisolvens isolate. Applied and environmental microbiology 78(2):528-535.
89.Chassard, C., Delmas, E., Lawson, P. A., & Bernalier-Donadille, A. 2008. Bacteroides xylanisolvens sp. nov., a xylan-degrading bacterium isolated from human faeces. International journal of systematic and evolutionary microbiology 58(4):1008-1013.
90.Bamba, T., Matsuda, H., Endo, M., & Fujiyama, Y. 1995. The pathogenic role of Bacteroides vulgatus in patients with ulcerative colitis. Journal of gastroenterology 30:45-47.

91.Nagaoka, K., Yanagihara, K., Morinaga, Y., Nakamura, S., Harada, T., Hasegawa, H., ... & Kohno, S. 2014. Prevotella intermedia Induces Severe Bacteremic Pneumococcal Pneumonia in Mice with Upregulated Platelet-Activating Factor Receptor Expression. Infection and immunity 82(2):587-593.
92.Sharma, A. 2010. Virulence mechanisms of Tannerella forsythia. Periodontology 54(1):106-116.
93.Wang, X., Zhu, D., Wang, M., Cheng, A., Jia, R., Zhou, Y., ... & Chen, X. Y. 2012. Complete genome sequence of Riemerella anatipestifer reference strain. Journal of bacteriology 194(12):3270-3271.
94.Nelson, K. E., Fleischmann, R. D., DeBoy, R. T., Paulsen, I. T., Fouts, D. E., Eisen, J. A., ... & Fraser, C. 2003. M. Complete genome sequence of the oral pathogenic bacterium Porphyromonas gingivalis strain W83. Journal of bacteriology 185(18):5591-5601.
95.Lamont, R. J., Chan, A., Belton, C. M., Izutsu, K. T., Vasel, D., & Weinberg, A. 1995. Porphyromonas gingivalis invasion of gingival epithelial cells. Infection and immunity 63(10):3878-3885.
96.Sakamoto, M., & Benno, Y. 2006. Reclassification of Bacteroides distasonis, Bacteroides goldsteinii and Bacteroides merdae as Parabacteroides distasonis gen. nov., comb. nov., Parabacteroides goldsteinii comb. nov. and Parabacteroides merdae comb. nov. International journal of systematic and evolutionary microbiology 56(7):1599-1605.
97.Xu, J., Mahowald, M. A., Ley, R. E., Lozupone, C. A., Hamady, M., Martens, E. C., ... & Gordon, J. I. 2007. Evolution of symbiotic bacteria in the distal human intestine. PLoS biology 5(7):e156.
98.Franzmann, P. D., & Skerman, V. B. D. 1984. Gemmata obscuriglobus, a new genus and species of the budding bacteria. Antonie van Leeuwenhoek 50(3):261-268.
99.Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., ... & Dubchak, I. 2014. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. Nucleic acids research 42(D1):D26-D31.
100.Grigoriev, I. V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., ... & Dubchak, I. 2011. The genome portal of the department of energy joint genome institute. Nucleic acids research gkr947.
101.Reddy, T. B. K., Thomas, A. D., Stamatis, D., Bertsch, J., Isbandi, M., Jansson, J., ... & Kyrpides, N. C. 2014. The Genomes OnLine Database (GOLD) v. 5: a metadata management system based on a four level (meta) genome project classification. Nucleic acids research gku950.
102.Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., ... & Stoeckert, C. J. 2011. Using OrthoMCL to Assign Proteins to OrthoMCL-DB Groups or to Cluster Proteomes Into New Ortholog Groups. Current protocols in bioinformatics 6-12.
103.Wu, S., Zhu, Z., Fu, L., Niu, B., & Li, W. 2011. WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC genomics 12(1):444.

104.Galperin, M. Y., Makarova, K. S., Wolf, Y. I., & Koonin, E. V. 2014. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic acids research gku1223.
105.Tatusov, R. L., Koonin, E. V., & Lipman, D. J. 1997. A genomic perspective on protein families. Science 278(5338):631-637.
106.Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., ... & Finn, R. D. 2014. The InterPro protein families database: the classification resource after 15 years. Nucleic acids research gku1243.
107.Gouret, P., Thompson, J. D., & Pontarotti, P. 2009. PhyloPattern: regular expressions to identify complex patterns in phylogenetic trees. BMC bioinformatics 10(1):298.
108.Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. 1990. Basic local alignment search tool. Journal of molecular biology 215(3):403-410.
109.Kruskal, W. H., & Wallis, W. A. 1952. Use of ranks in one-criterion variance analysis. Journal of the American statistical Association 47(260):583-621.
110.Tukey, J. W. 1949. Comparing individual means in the analysis of variance. Biometrics 99-114.
111.Douglas, C. E., & Michael, F. A. 1991. On distribution-free multiple comparisons in the one-way analysis of variance. Communications in statistics-theory and methods 20(1):127-139.
112.Hirt, R. P., Alsmark, C., & Embley, T. M. 2015. Lateral gene transfers and the origins of the eukaryote proteome: a view from microbial parasites. Current opinion in microbiology 23:155-162.
113.Merhej, V., Royer-Carenzi, M., Pontarotti, P., & Raoult, D. 2009. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biology direct 4(1):13.
114.Wolf, Y. I., & Koonin, E. V. 2013. Genome reduction as the dominant mode of evolution. Bioessays 35(9):829-837.
115.Molmeret, M., Horn, M., Wagner, M., Santic, M., & Kwaik, Y. A. 2005. Amoebae as training grounds for intracellular bacterial pathogens. Applied and environmental microbiology 71(1), 20-28.
116.Moliner, C., Fournier, P. E., & Raoult, D. 2010. Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. FEMS microbiology reviews 34(3):281-294
117.Daubin, V., Gouy, M., & Perriere, G. 2002. A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. Genome research 12(7):1080-1090.
118.Charlebois, R. L., Clarke, G. P., Beiko, R. G., & Jean, A. S. 2003. Characterization of species-specific genes using a flexible, web-based querying system. FEMS microbiology letters 225(2):213-220.
119.Lawrence, J. G. & Roth, J. R. 1996. Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics 143:1843±1860.
120.Nakamura, Y., Itoh, T., Matsuda, H., & Gojobori, T. 2004. Biased biological functions of horizontally transferred genes in prokaryotic genomes. Nature genetics 36(7):760-766.

121.Ruth E. Ley, Daniel A. Peterson, and Jeffrey I. Gordon. 2006. Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine. Cell 124:837–848
122.Quaiser, A., Ochsenreiter, T., Lanz, C., Schuster, S. C., Treusch, A. H., Eck, J., & Schleper, C. 2006. Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. Molecular microbiology 50(2):563-575.
123.Subtil, A., Collingro, A., & Horn, M. 2014. Tracing the primordial Chlamydia: extinct parasites of plants?. Trends in plant science 19(1):36-43.
124.Moustafa, A., Reyes-Prieto, A., & Bhattacharya, D. 2008. Chlamydia has contributed at least 55 genes to Plantae with predominantly plastid functions. PLoS One 3(5):e2205.
125.Viswanathan, V. K. 2012. Spirochaetes and their twisted ways. Gut microbes 3(5):399-400.
126.Bellgard MI, Wanchanthuek P, La T, Ryan K, Moolhuijzen P, Albertyn Z. 2009. Genome sequence of the pathogenic intestinal spirochete brachyspira hyodysenteriae reveals adaptations to its lifestyle in the porcine large intestine. PLoS One 4:e4641.

Figures legends

Figure 1. Comparison of the genome contents in the different habitats. The quantity of genes in core genome and their functions are indicated in the center of flower. Genes specific to each habitats and genes shared by different habitats are represented by the different petals. The clouds represents the ORFans percentages in genomes. The width of the petals and the clouds is proportional to the quantity of specific genes and percentage of ORFans, respectively.

Figure 2. Functional distribution of the specific genes. A) Proportions of each COG categories in the bacteria from different habitats. B) The repartition of the four functional categories in each studied habitat.

Figure 3. Functional description of the detected horizontally transferred genes in each studied bacteria. Only the four functional categories of COGs are presented. The habitats of species and their

ancestors are indicated by the color of the branches (red : ubiquitous, cyan : water, green : insects, yellow: amoeba, pink: intracellular vertebrates, purple: extracellular vertebrates, blue: water and soils, brown: soils), the black dotted branches indicate an unknown habitat

Figure 4. Preferences for horizontal transfer partners among habitats. The colored point corresponds to the bacteria from the different habitats (red : ubiquitous, cyan : water, green : insects, yellow: amoeba, pink: intracellular vertebrates, purple: extracellular vertebrates, blue: water and soils, brown: soils). The traits are colored according to the habitats of studied bacteria and their thickness is proportional to the amount of genes exchanged.

Supplementary figure

Table S1. Genomic features and habitats of bacteria studied.

For each bacteria studied, we present the genomes identifiers, genomic features (proteins quantity, GC percent) and the habitat (environment and lifestyle).

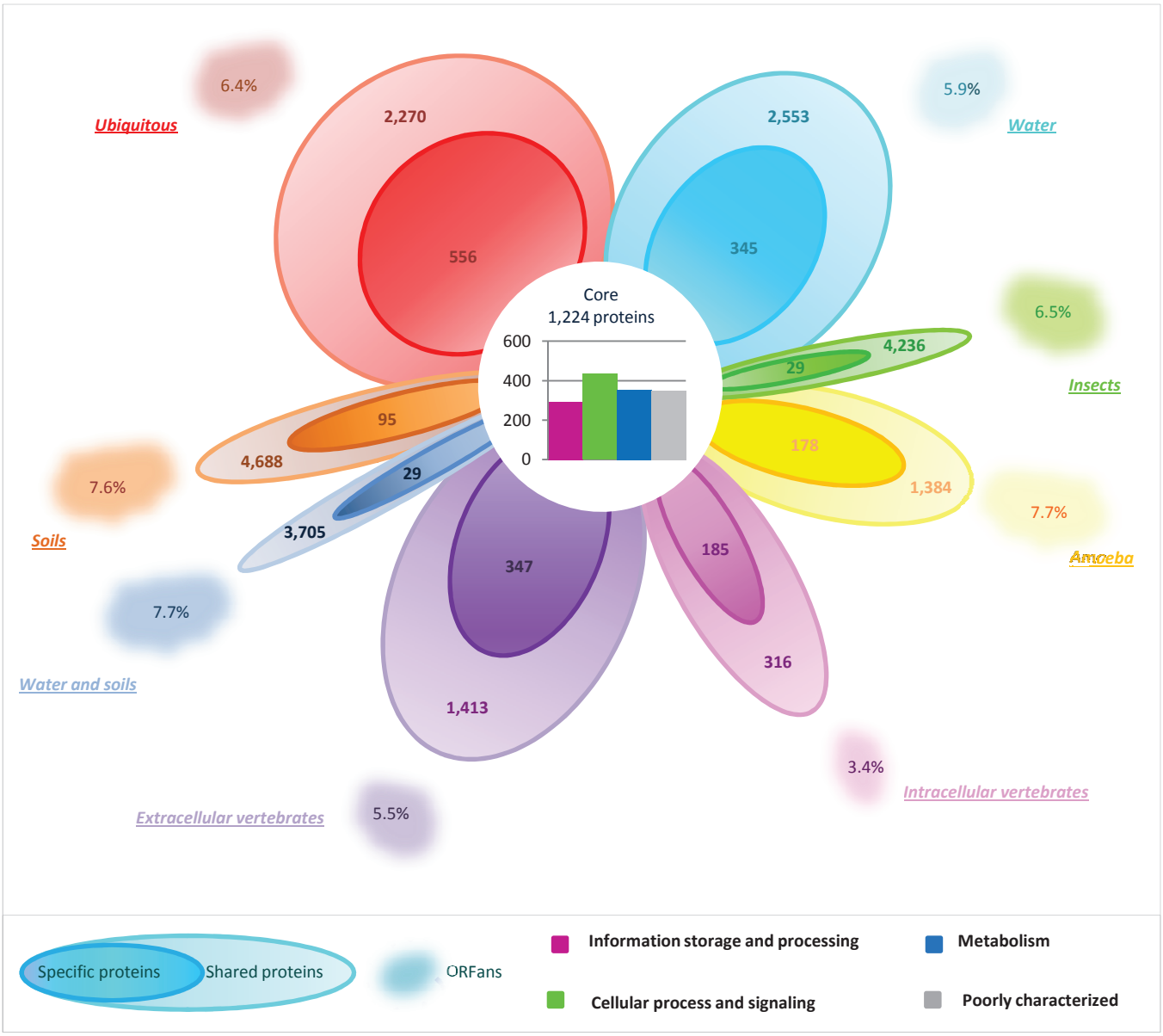


Figure 1. Comparison of the genome contents in the different habitats. The quantity of genes in core genome and their functions are indicated in the center of flower. Genes specific to each habitats and genes shared by different habitats are represented by the different petals. The clouds represents the ORFans percentages in genomes. The width of the petals and the clouds is proportional to the quantity of specific genes and percentage of ORFans, respectively.

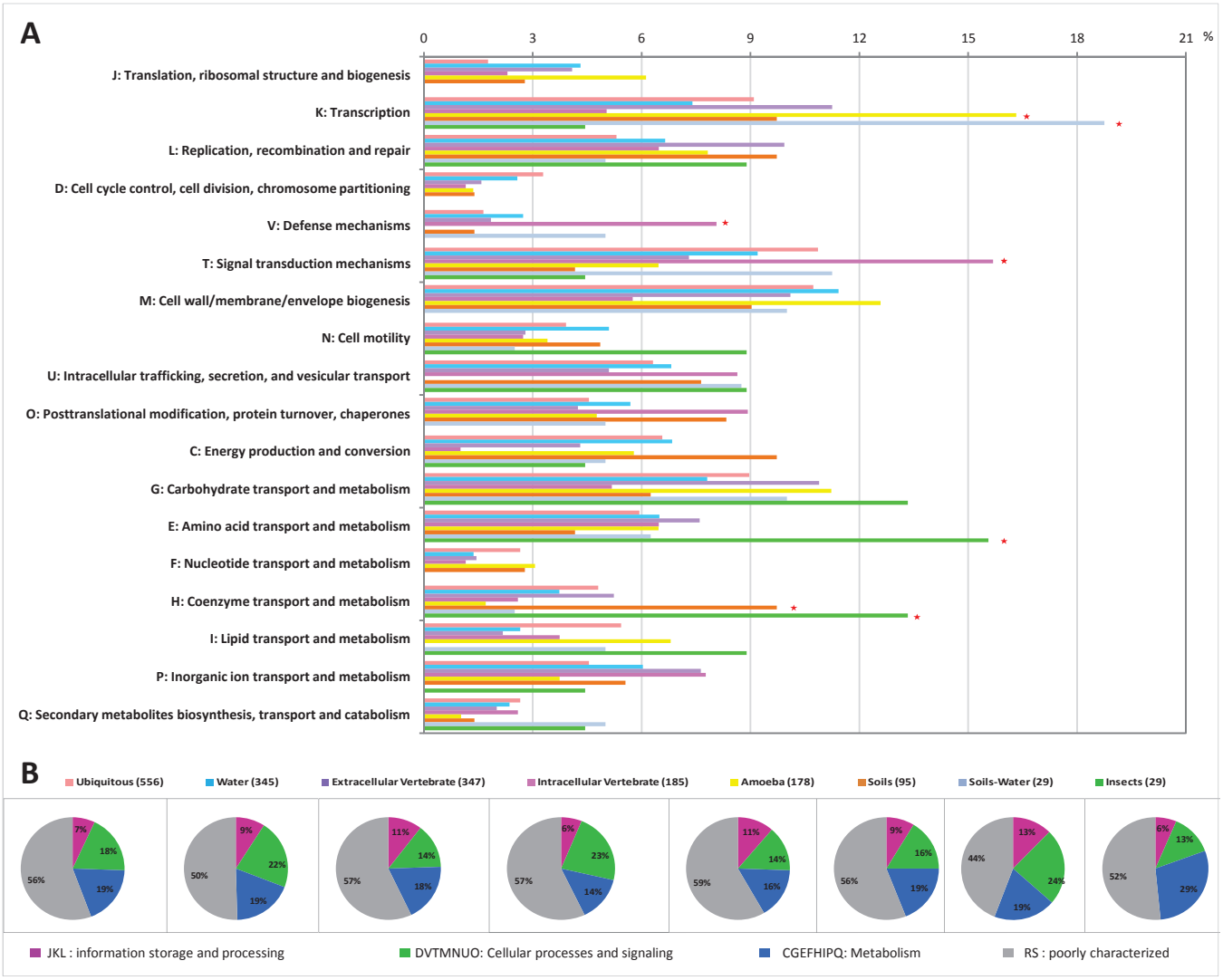


Figure 2: Functional distribution of the specific genes. A) Proportions of each COG categories in the bacteria from different habitats. B) The repartition of the four functional categories in each studied habitat.

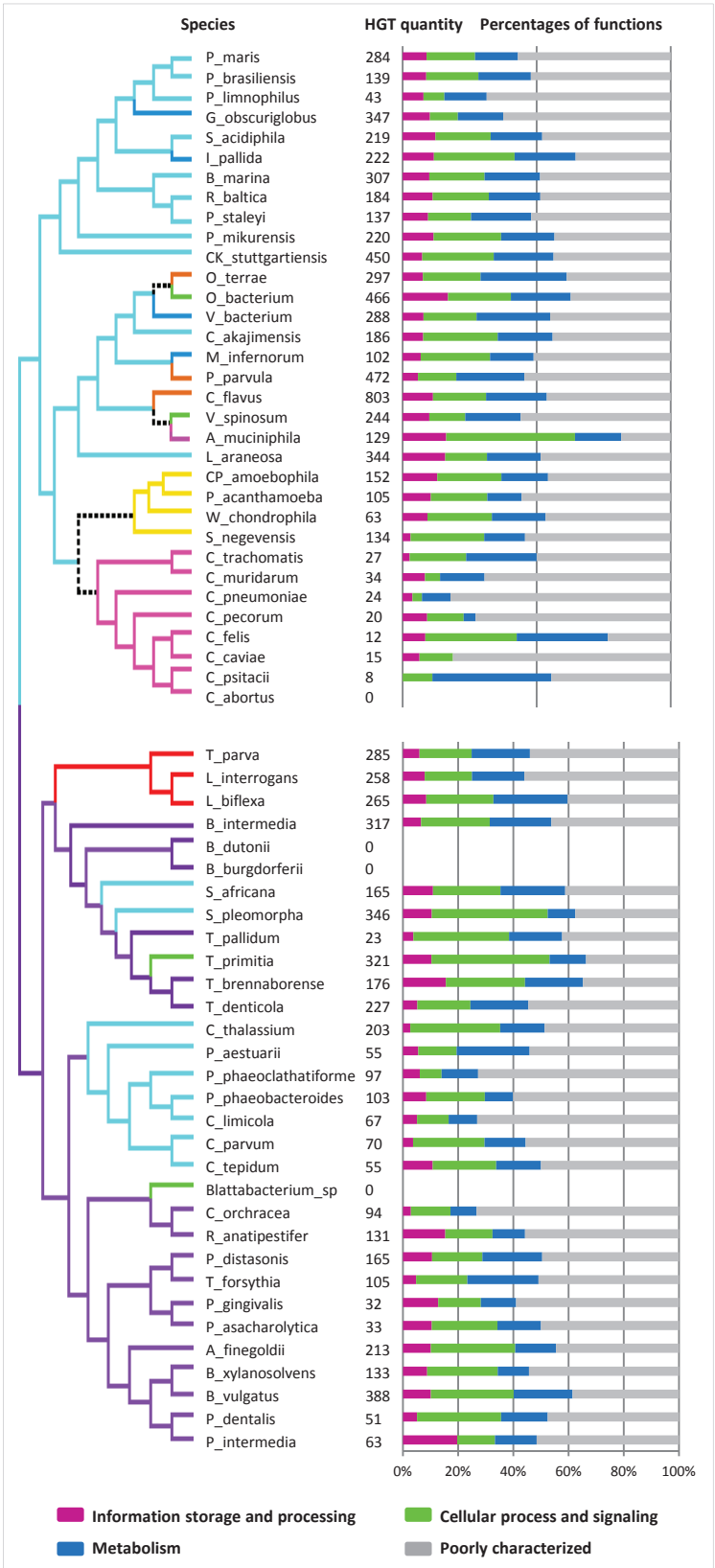


Figure 3: Functional description of the detected horizontally transferred genes in each studied bacteria. Only the four functional categories of COGs are presented. The habitats of species and their ancestors are indicated by the color of the branches (red : ubiquitous, cyan : water, green : insects, yellow: Amoeba, pink: intracellular vertebrates, purple: extracellular vertebrates, blue: water and soils, brown: soils), the black dotted branches indicate an unknown habitat

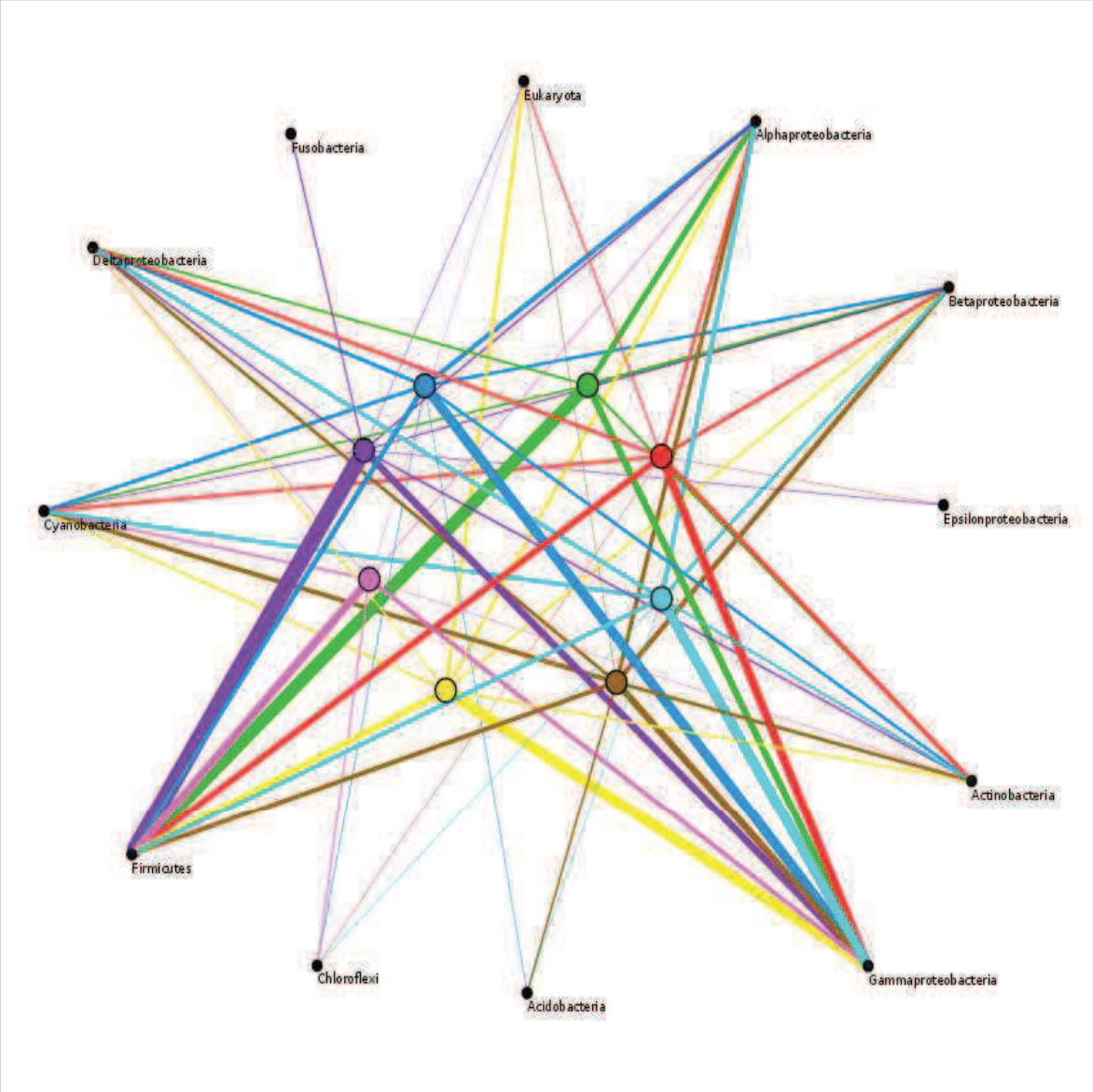


Figure 4: Preferences for horizontal transfer partners among habitats. The colored point corresponds to the bacteria from the different habitats (red : ubiquitous, cyan : water, green : insects, yellow: Amoeba, pink: intracellular vertebrates, purple: extracellular vertebrates, blue: water and soils, brown: soils). The traits are colored according to the habitats of studied bacteria and their thickness is proportional to the amount of genes exchanged.

Species	Genome Size	GC percent	Proteins	Genome identifier	Environment	Lifestyle
Chlamydia muridarum Nigg	1,09	40,30%	911	NC_002620	Vertebrates	Intracellular
Chlamydophila abortus S26/3	1,14	39,90%	932	NC_004552	Vertebrates	Intracellular
Chlamydia trachomatis 434/Bu	1,04	41,30%	885	NC_010287	Vertebrates	Intracellular
Chlamydophila pecorum E58	1,11	41,10%	988	NC_015408	Vertebrates	Intracellular
Chlamydophila caviae GPIC	1,18	39,20%	1005	NC_003361	Vertebrates	Intracellular
Chlamydophila felis Fe/C-56	1,17	39,40%	1013	NC_007899	Vertebrates	Intracellular
Chlamydia psittaci NJ1	1,17	39,00%	1052	NC_018626	Vertebrates	Intracellular
Chlamydophila pneumoniae J138	1,23	40,60%	1069	NC_002491	Vertebrates	Intracellular
Candidatus Protochlamydia amoebophila	2,41	34,70%	2031	NC_005861	Amoeba	Intracellular
Parachlamydia acanthamoebae UV-7	3,07	39,00%	2782	NC_015702	Amoeba	Intracellular
Simkania negevensis Z	2,63	41,60%	2512	NC_015713	Amoeba	Intracellular
Waddlia chondrophila WSU 86-1044	2,13	43,80%	1956	NC_014225	Amoeba	Intracellular
Blattabacterium sp (Blaberus giganteus)	0,63	25,70%	576	NC_017924	Insects	Intracellular
Verrucomicrobium spinosum DSM 4136	8,22	58,60%	6477	NZ_ABIZ00000000	Insects	Extracellular
Treponema primitia ZAS-2	4,06	50,80%	3523	NC_015578	Insects	Extracellular
Opitutaceae bacterium TAV1	7,07	62,10%	5830	NZ_AHKS00000000	Insects	Extracellular
Brachyspira intermedia PWS/A	3,31	27,20%	2872	NC_017243	Vertebrates	Extracellular
Treponema brennaborense DSM 12168	3,06	51,50%	2531	NC_015500	Vertebrates	Extracellular
Treponema denticola ATCC 35405	2,84	37,90%	2767	NC_002967	Vertebrates	Extracellular
Treponema pallidum subsp. pallidum DAL-1	1,14	52,80%	1056	NC_016844	Vertebrates	Extracellular
Akkermansia muciniphila ATCC BAA-835	2,66	55,80%	2138	NC_010655	Vertebrates	Extracellular
Alistipes finegoldii DSM 17242	3,73	56,60%	3110	NC_018011	Vertebrates	Extracellular
Bacteroides vulgatus ATCC 8482	5,16	42,20%	4066	NC_009614	Vertebrates	Extracellular
Bacteroides xylanisolvens XB1A	5,98	42,10%	4405	NC_021017	Vertebrates	Extracellular
Capnocytophaga ochracea DSM 7271	2,61	39,60%	2171	NC_013162	Vertebrates	Extracellular
Porphyromonas asaccharolytica DSM 20707	2,19	52,50%	1699	NC_015501	Vertebrates	Extracellular
Prevotella dentalis DSM 3688	3,35	55,90%	2544	NC_019960 NC_019968	Vertebrates	Extracellular
Parabacteroides distasonis ATCC 8503	4,81	45,10%	3849	NC_009615	Vertebrates	Extracellular
Porphyromonas gingivalis W83	2,34	48,30%	1909	NC_002950	Vertebrates	Extracellular
Prevotella intermedia 17	2,7	43,50%	2266	NC_017860 NC_017861	Vertebrates	Extracellular
Riemerella anatipestifer ATCC 11845 = DSM 15868	2,16	35,00%	1972	NC_014738	Vertebrates	Extracellular
Tannerella forsythia ATCC 43037	3,41	47,00%	3001	NC_016610	Vertebrates	Extracellular
Borrelia burgdorferi JD1	1,53	28,00%	1458	NC_017403	Vertebrates	Extracellular
Borrelia garinii BgVir	0,99	28,20%	950	NC_017717	Vertebrates	Extracellular
Chthoniobacter flavus Ellin428	7,85	61,10%	6716	NZ_ABVL00000000	Soils	Extracellular
Opitutus terrae PB90-1	5,96	65,30%	4612	NC_010571	Soils	Extracellular
Pedosphaera parvula Ellin514	7,41	52,60%	6510	NZ_ABOX00000000	Soils	Extracellular
Methylacidiphilum infernorum V4	2,29	45,50%	2472	NC_010794	Soils, water	Extracellular
Verrucomicrobiae bacterium DG1235	5,78	54,30%	4909	NZ_ABSI00000000	Soils, water	Extracellular
Gemmata obscuriglobus UQM 2246	9,16	67,20%	7969	NZ_ABGO00000000	Soils, water	Extracellular
Isosphaera pallida ATCC 43644	5,53	62,40%	3722	NC_014962	Soils, water	Extracellular
Leptospira biflexa serovar Patoc	3,96	38,90%	3600	NC_010842 NC_010845	Ubiquitous	Extracellular
Leptospira interrogans serovar Lai str. 56601	4,7	35,00%	3702	NC_004342 NC_004343	Ubiquitous	Extracellular
Turneriella parva DSM 21527	4,41	53,60%	4139	NC_018020.1	Ubiquitous	Extracellular
Spirochaeta africana DSM 8902	3,29	57,80%	2782	NC_017098	Water	Extracellular
Sphaerochaeta pleomorpha str. Grapes	3,59	46,20%	3159	NC_016633	Water	Extracellular
Coraliomargarita akajimensis DSM 45221	3,75	53,60%	3120	NC_014008	Water	Extracellular
Lentisphaera araneosa HTCC2155	6,02	40,90%	5104	NZ_ABCK00000000	Water	Extracellular
Planctomyces brasiliensis DSM 5305	6,01	56,40%	4750	NC_015174	Water	Extracellular
Planctomyces limnophilus DSM 3776	5,46	53,70%	4258	NC_014148	Water	Extracellular
Planctomyces maris DSM 8797	7,78	50,50%	6480	NZ_ABCE00000000	Water	Extracellular
Phycisphaera mikurensis NBRC 102666	3,88	73,20%	3282	NC_017080	Water	Extracellular
Pirellula staleyi DSM 6068	6,2	57,50%	4717	NC_013720	Water	Extracellular
Rhodopirellula baltica SH 1	7,15	55,40%	7325	NC_005027	Water	Extracellular
Singulisphaera acidiphila DSM 18658	9,76	62,30%	7251	NC_019892	Water	Extracellular
Blastopirellula marina DSM 3645	6,66	57,00%	6025	NZ_AANZ00000000	Water	Extracellular
Candidatus Kuenenia stuttgartiensis	4,2	41,00%	4663	_	Water	Extracellular
Chlorobium limicola DSM 245	2,76	51,30%	2434	NC_010803	Water	Extracellular
Chlorobaculum parvum NCIB 8327	2,29	55,80%	2043	NC_011027	Water	Extracellular
Chlorobium phaeobacteroides BS1	2,74	48,90%	2469	NC_010831	Water	Extracellular
Chlorobium tepidum TLS	2,15	56,50%	2245	NC_002932	Water	Extracellular
Chloroherpeton thalassium ATCC 35110	3,29	45,00%	2710	NC_011026	Water	Extracellular
Prosthecochloris aestuarii DSM 271	2,58	50,10%	2327	NC_011059	Water	Extracellular
Pelodictyon phaeoclathratiforme BU-1	3,02	48,10%	2707	NC_011060	Water	Extracellular

Table S1 Genomic features and habitats of bacteria studied.

For each bacteria studied, we present the genomes identifiers, genomic features (proteins quantity, GC percent) and the habitat (environment and lifestyle).

Additional tables not submitted

Table S2: Results for general analyzes of genomes (Detail for core gene, orthologous groups, specific genes and HGT)

Tables S3 : Functions for specific and transferred genes (quantity and percent) ,

Table S4 : Partners for genes transfers (quantity and percent)

Species		Habitats		Genomes				Protein
		Environment	Lifestyle	Identfifiers	Size	Proteins	Orthologous group (OG)	Quantity
CP_amoebophila	Candidatus Protochlamydia amoebophila	Amoeba	Intracellular	NC_005861	2.4	2031	1223	1301
P_acanthamoeba	Parachlamydia acanthamoebae UV-7	Amoeba	Intracellular	NC_015702	3.1	2782	1573	1711
S_negevensis	Simkania negevensis Z	Amoeba	Intracellular	NC_015713	2.6	2512	1229	1328
W_chondrophila	Waddlia chondrophila WSU 86-1044	Amoeba	Intracellular	NC_014225	2.1	1956	1309	1420
O_bacterium	Opitutaceae bacterium TAV1	Insects	Extracellular	NZ_AHKS00000000	7.1	5830	2822	3240
T_primitia	Treponema primitia ZAS-2	Insects	Extracellular	NC_015578	4.1	3523	2052	2377
V_spinosum	Verrucomicrobium spinosum DSM 4136	Insects	Extracellular	NZ_ABIZ00000000	8.2	6477	3516	4073
Blattabacterium_sp	Blattabacterium sp (Blaberus giganteus)	Insects	Intracellular	NC_017924	0.6	576	521	526
C_flavus	Chthoniobacter flavus Ellin428	Soils	Extracellular	NZ_ABVL00000000	7.9	6716	3620	4213
O_terrae	Opitutus terrae PB90-1	Soils	Extracellular	NC_010571	6.0	4612	2979	3435
P_parvula	Pedosphaera parvula Ellin514	Soils	Extracellular	NZ_ABOX00000000	7.4	6510	3195	3761
G_obscuriglobus	Gemmata obscuriglobus UQM 2246	Soils, water	Extracellular	NZ_ABGO00000000	9.2	7969	3436	4245
I_pallida	Isosphaera pallida ATCC 43644	Soils, water	Extracellular	NC_014962	5.5	3722	2737	2905
M_infernorum	Methylacidiphilum infernorum V4	Soils, water	Extracellular	NC_010794	2.3	2472	1348	1433
V_bacterium	Verrucomicrobiae bacterium DG1235	Soils, water	Extracellular	NZ_ABSI00000000	5.8	4909	2908	3288
L_biflexa	Leptospira biflexa serovar Patoc	Ubiquitous	Extracellular	NC_010842 NC_010845	4.0	3600	2579	2802
L_interrogans	Leptospira interrogans serovar Lai str. 56601	Ubiquitous	Extracellular	NC_004342 NC_004343	4.7	3702	2383	2561
T_parva	Turneriella parva DSM 21527	Ubiquitous	Extracellular	NC_018020.1	4.4	4139	2228	2518
A_finegoldii	Alistipes finegoldii DSM 17242	Vertebrates	Extracellular	NC_018011	3.7	3110	1853	2097
A_muciniphila	Akkermansia muciniphila ATCC BAA-835	Vertebrates	Extracellular	NC_010655	2.7	2138	1461	1579
B_burgdorferii	Borrelia burgdorferi JD1	Vertebrates	Extracellular	NC_017403	1.5	1458	746	762
B_dutonii	Borrelia dutonii	Vertebrates	Extracellular	NC_017717	1.0	950	740	755
B_intermedia	Brachyspira intermedia PWS/A	Vertebrates	Extracellular	NC_017243	3.3	2872	1458	1637
B_vulgatus	Bacteroides vulgatus ATCC 8482	Vertebrates	Extracellular	NC_009614	5.2	4066	2681	3196
B_xylanosolvens	Bacteroides xylanisolvens XB1A	Vertebrates	Extracellular	NC_021017	6.0	4405	2709	3200
C_orchracea	Capnocytophaga ochracea DSM 7271	Vertebrates	Extracellular	NC_013162	2.6	2171	1518	1602
P_asacharolytica	Porphyromonas asaccharolytica DSM 20707	Vertebrates	Extracellular	NC_015501	2.2	1699	1292	1381
P_dentalis	Prevotella dentalis DSM 3688	Vertebrates	Extracellular	NC_019960 NC_019968	3.4	2544	1839	2016
P_distasonis	Parabacteroides distasonis ATCC 8503	Vertebrates	Extracellular	NC_009615	4.8	3849	2632	3087
P_gingivalis	Porphyromonas gingivalis W83	Vertebrates	Extracellular	NC_002950	2.3	1909	1424	1524
P_intermedia	Prevotella intermedia 17	Vertebrates	Extracellular	NC_017860 NC_017861	2.7	2266	1549	1649
R_anatipestifer	Riemerella anatipestifer ATCC 11845 = DSM 15868	Vertebrates	Extracellular	NC_014738	2.2	1972	1397	1466
T_brennaborense	Treponema brennaborense DSM 12168	Vertebrates	Extracellular	NC_015500	3.1	2531	1876	2037
T_denticola	Treponema denticola ATCC 35405	Vertebrates	Extracellular	NC_002967	2.8	2767	1621	1778
T_forsythia	Tannerella forsythia ATCC 43037	Vertebrates	Extracellular	NC_016610	3.4	3001	1904	2059
T_pallidum	Treponema pallidum subsp. pallidum DAL-1	Vertebrates	Extracellular	NC_016844	1.1	1056	817	838
C_abortus	Chlamydophila abortus S26/3	Vertebrates	Intracellular	NC_004552	1.1	932	858	874
C_caviae	Chlamydophila caviae GPIC	Vertebrates	Intracellular	NC_003361	1.2	1005	876	896
C_felis	Chlamydophila felis Fe/C-56	Vertebrates	Intracellular	NC_007899	1.2	1013	883	907
C_muridarum	Chlamydia muridarum Nigg	Vertebrates	Intracellular	NC_002620	1.1	911	796	809
C_pecorum	Chlamydophila pecorum E58	Vertebrates	Intracellular	NC_015408	1.1	988	824	845
C_pneumoniae	Chlamydophila pneumoniae J138	Vertebrates	Intracellular	NC_002491	1.2	1069	834	863
C_psitacii	Chlamydia psittaci NJ1	Vertebrates	Intracellular	NC_018626	1.2	1052	874	905
C_trachomatis	Chlamydia trachomatis 434/Bu	Vertebrates	Intracellular	NC_010287	1.0	885	802	817
B_marina	Blastopirellula marina DSM 3645	Water	Extracellular	NZ_AANZ00000000	6.7	6025	3621	4004
C_akajimensis	Coraliomargarita akajimensis DSM 45221	Water	Extracellular	NC_014008	3.8	3120	2121	2309
C_limicola	Chlorobium limicola DSM 245	Water	Extracellular	NC_010803	2.8	2434	1898	2018
C_parvum	Chlorobaculum parvum NCIB 8327	Water	Extracellular	NC_011027	2.3	2043	1725	1796
C_phaeobacteroides	Chlorobium phaeobacteroides BS1	Water	Extracellular	NC_010831	2.7	2469	1850	1978
C_tepidum	Chlorobium tepidum TLS	Water	Extracellular	NC_002932	2.2	2245	1610	1681
C_thalassium	Chloroherpeton thalassium ATCC 35110	Water	Extracellular	NC_011026	3.3	2710	1927	2073
CK_stuttgartiensis	Candidatus Kuenenia stuttgartiensis	Water	Extracellular	_	4.2	4663	1856	2191
L_araneosa	Lentisphaera araneosa HTCC2155	Water	Extracellular	NZ_ABCK00000000	6.0	5104	2360	2853
P_aestuarii	Prosthecochloris aestuarii DSM 271	Water	Extracellular	NC_011059	2.6	2327	1799	1915
P_brasiliensis	Planctomyces brasiliensis DSM 5305	Water	Extracellular	NC_015174	6.0	4750	3295	3664
P_limnophilus	Planctomyces limnophilus DSM 3776	Water	Extracellular	NC_014148	5.5	4258	2978	3187
P_maris	Planctomyces maris DSM 8797	Water	Extracellular	NZ_ABCE00000000	7.8	6480	3843	4383
P_mikurensis	Phycisphaera mikurensis NBRC 102666	Water	Extracellular	NC_017080	3.9	3282	1932	2083
P_phaeoclathatiforme	Pelodictyon phaeoclathratiforme BU-1	Water	Extracellular	NC_011060	3.0	2707	1903	2073
P_staleyi	Pirellula staleyi DSM 6068	Water	Extracellular	NC_013720	6.2	4717	3328	3629
R_baltica	Rhodopirellula baltica SH 1	Water	Extracellular	NC_005027	7.2	7325	3435	3881
S_acidiphila	Singulisphaera acidiphila DSM 18658	Water	Extracellular	NC_019892	9.8	7251	4117	4917
S_africana	Spirochaeta africana DSM 8902	Water	Extracellular	NC_017098	3.3	2782	1849	2043
S_pleomorpha	Sphaerochaeta pleomorpha str. Grapes	Water	Extracellular	NC_016633	3.6	3159	1931	2299

Table S2: Results for general analyzes of genomes (Detail for core gene, orthologous groups, specific genes and HGT)

s in OG		ORFans		Specific gens		Core genes		Proteins transferred		Sequences transferred				
Percent	Proteins/OG	Quantity	Percent	Quantity	Percent	Quantity	Percent	Quantity	Percent of proteomes	Quantity	Genes by sequences	Average size	Total size	Percent of genomes
64.1	1.1	148	7.3	170	8.4	734	36.1	188	9.3	152	1.2	1213.9	184517.0	7.7
61.5	1.1	189	6.8	232	8.3	871	31.3	120	4.3	105	1.1	989.5	103893.0	3.4
52.9	1.1	340	14.3	113	4.5	736	29.3	156	6.2	134	1.2	1052.9	141086.0	5.4
72.6	1.1	132	6.8	184	9.4	765	39.1	70	3.6	63	1.1	998.6	62909.0	3.0
55.6	1.2	430	3.7	37	0.6	1085	18.6	525	9.0	466	1.1	1289.2	600787.0	8.5
67.5	1.2	233	6.6	15	0.4	874	24.8	462	13.1	321	1.4	1465.3	471818.0	11.6
62.9	1.2	923	14.9	34	0.5	1220	18.8	268	4.1	244	1.1	1032.8	253043.0	3.1
91.3	1.0	1	0.2	0	0.0	399	69.3	82	14.2	0	0.0	0.0	0.0	0.0
62.7	1.2	676	10.0	110	1.6	1220	18.2	1026	15.3	803	1.3	1134.3	910602.0	11.5
74.5	1.2	238	5.2	50	1.1	1198	26.0	351	7.6	297	1.2	1473.9	437758.0	7.3
57.8	1.2	715	11.0	123	1.9	1248	19.2	496	7.6	472	1.1	966.1	455983.0	6.2
53.3	1.2	863	11.1	56	0.7	1375	17.3	377	4.7	347	1.1	859.4	299077.0	3.3
78.1	1.1	318	8.6	42	1.1	1019	27.4	245	6.6	222	1.1	1295.8	287664.0	5.2
58.0	1.1	285	11.5	6	0.2	804	32.5	130	5.3	102	1.3	1186.8	121051.0	5.3
67.0	1.1	544	11.1	12	0.2	1126	22.9	333	6.8	288	1.2	1247.2	359186.0	6.2
77.8	1.1	152	4.6	709	19.7	1032	28.7	304	8.4	265	1.1	1103.7	292471.0	7.4
69.2	1.1	372	10.1	664	17.9	968	26.1	292	7.9	258	1.1	1004.3	259107.0	5.5
60.8	1.1	426	10.4	291	7.0	1048	25.3	338	8.2	285	1.2	1075.4	306492.0	6.9
67.4	1.1	311	10.0	431	13.9	761	24.5	240	7.7	213	1.1	1161.1	247318.0	6.6
73.9	1.1	214	10.0	47	2.2	822	38.4	159	7.4	129	1.2	1362.2	175717.0	6.6
52.3	1.0	6	0.7	164	11.2	358	24.6	0	0.0	0	0.0	0.0	0.0	0.0
79.5	1.0	6	0.8	152	16.0	369	38.8	0	0.0	0	0.0	0.0	0.0	0.0
57.0	1.1	133	4.6	52	1.8	780	27.2	416	14.5	317	1.3	1213.0	384535.0	11.6
78.6	1.2	112	2.7	780	19.2	913	22.5	466	11.5	388	1.2	1421.4	551496.0	10.7
72.6	1.2	366	8.3	801	18.2	876	19.9	157	3.6	133	1.2	1359.2	180773.0	3.0
73.8	1.1	93	4.3	251	11.6	660	30.4	104	4.8	94	1.1	878.3	82562.0	3.2
81.3	1.1	82	4.8	233	13.7	601	35.4	34	2.0	33	1.0	983.4	32453.0	1.5
79.3	1.1	33	2.3	494	19.4	671	26.4	55	2.2	51	1.1	1287.3	65650.0	2.0
80.2	1.2	153	4.0	661	17.2	928	24.1	188	4.9	165	1.1	1188.9	196167.0	4.1
79.8	1.1	112	5.9	300	15.7	616	32.3	37	1.9	32	1.2	1217.8	38971.0	1.7
72.8	1.1	71	3.1	400	17.7	591	26.1	66	2.9	63	1.0	1061.7	66888.0	2.5
74.3	1.1	63	3.2	188	9.5	673	34.1	153	7.8	131	1.2	972.3	127371.0	5.9
80.5	1.1	93	3.7	105	4.1	711	28.1	211	8.3	176	1.2	1213.8	213631.0	7.0
64.3	1.1	139	5.0	136	4.9	640	23.1	319	11.5	227	1.4	1330.5	302027.0	10.6
68.6	1.1	186	6.2	446	14.9	665	22.2	119	4.0	105	1.1	1362.0	144368.0	4.2
79.4	1.0	107	10.1	64	6.1	400	37.9	24	2.3	23	1.0	1124.0	25853.0	2.3
93.8	1.0	6	0.7	192	20.6	486	52.1	0	0.0	0	0.0	0.0	0.0	0.0
89.2	1.0	11	1.1	207	20.6	489	48.7	16	1.6	15	1.1	1519.1	22787.0	1.9
89.5	1.0	18	1.8	202	19.9	498	49.2	14	1.4	12	1.2	1021.5	12258.0	1.0
88.8	1.0	20	2.2	160	17.6	469	51.5	35	3.8	34	1.0	1190.2	40467.0	3.7
85.5	1.0	43	4.3	155	15.7	492	49.8	22	2.2	20	1.1	1213.4	24268.0	2.2
80.7	1.0	73	6.6	164	15.3	489	45.7	30	2.8	24	1.3	1246.8	29924.0	2.4
86.0	1.0	31	2.9	206	19.6	493	46.9	8	0.8	8	1.0	601.6	4813.0	0.4
92.3	1.0	58	5.7	159	18.0	473	53.4	34	3.8	27	1.3	1478.7	39925.0	3.8
66.5	1.1	731	11.9	669	11.1	1071	17.8	343	5.7	307	1.1	1088.2	334066.0	5.0
74.0	1.1	293	9.4	73	2.3	923	29.6	217	7.0	186	1.2	1164.6	216618.0	5.8
82.9	1.1	96	3.9	377	15.5	808	33.2	77	3.2	67	1.1	1290.5	86461.0	3.1
87.9	1.0	32	1.6	322	15.8	762	37.3	78	3.8	70	1.1	1049.6	73471.0	3.2
80.1	1.1	145	5.9	350	14.2	765	31.0	115	4.7	103	1.1	892.6	91934.0	3.4
74.9	1.0	151	6.7	284	12.7	750	33.4	68	3.0	55	1.2	1152.5	63388.0	2.9
76.5	1.1	141	5.2	217	8.0	859	31.7	248	9.2	203	1.2	1218.3	247323.0	7.5
47.0	1.2	691	14.7	104	2.2	969	20.8	542	11.6	450	1.2	1172.5	527640.0	12.6
55.9	1.2	786	15.3	163	3.2	935	18.3	358	7.0	344	1.0	1054.8	362842.0	6.0
82.3	1.1	91	4.0	352	15.1	775	33.3	67	2.9	55	1.2	1167.9	64234.0	2.5
77.1	1.1	475	10.0	602	12.7	1021	21.5	146	3.1	139	1.1	913.3	126950.0	2.1
74.9	1.1	477	11.3	488	11.5	995	23.4	48	1.1	43	1.1	1007.5	43322.0	0.8
67.6	1.1	807	12.5	723	11.2	1067	16.5	322	5.0	284	1.1	1061.9	301569.0	3.9
63.5	1.1	467	14.6	112	3.4	902	27.5	244	7.4	220	1.1	1173.1	258088.0	6.7
76.6	1.1	51	1.9	356	13.2	793	29.3	111	4.1	97	1.1	1000.3	97026.0	3.2
76.9	1.1	399	8.5	561	11.9	1028	21.8	152	3.2	137	1.1	1041.2	142643.0	2.3
53.0	1.1	775	10.6	512	7.0	1036	14.1	205	2.8	184	1.1	1054.5	194026.0	2.7
67.8	1.2	802	11.2	276	3.8	1298	17.9	238	3.3	219	1.1	1149.0	251632.0	2.6
73.4	1.1	176	6.3	63	2.3	844	30.3	208	7.5	165	1.3	1315.7	217094.0	6.6
72.8	1.2	148	4.7	77	2.4	823	26.1	504	16.0	345	1.5	1492.2	514824.0	14.3

Species		Quantity															
	Complete name	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
CP_amoebophila	Candidatus Protochlamydia amoebophila	1.0	0.0	3.0	1.0	4.0	3.0	9.0	0.0	4.0	5.0	10.0	5.0	8.0	2.0	3.0	3.0
P_acanthamoeba	Parachlamydia acanthamoebae UV-7	1.0	0.0	6.0	2.0	5.0	2.0	10.0	2.0	7.0	5.0	14.0	7.0	10.0	3.0	5.0	3.0
S_negevensis	Simkania negevensis Z	1.0	0.0	3.0	0.0	3.0	1.0	6.0	1.0	2.0	3.0	9.0	5.0	7.0	1.0	2.0	3.0
W_chondrophila	Waddlia chondrophila WSU 86-1044	1.0	0.0	5.0	1.0	6.0	3.0	10.0	2.0	7.0	3.0	13.0	4.0	11.0	4.0	4.0	2.0
O_bacterium	Opitutaceae bacterium TAV1	0.0	0.0	1.0	0.0	3.0	0.0	3.0	1.0	1.0	0.0	1.0	2.0	1.0	0.0	0.0	0.0
T_primitia	Treponema primitia ZAS-2	0.0	0.0	1.0	0.0	2.0	0.0	0.0	2.0	2.0	0.0	0.0	0.0	0.0	2.0	0.0	1.0
V_spinosum	Verrucomicrobium spinosum DSM 4136	0.0	0.0	0.0	0.0	2.0	0.0	3.0	3.0	0.0	0.0	1.0	2.0	1.0	2.0	0.0	1.0
Blattabacterium_sp	Blattabacterium sp (Blaberus giganteus)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
C_flavus	Chthoniobacter flavus Ellin428	0.0	0.0	6.0	1.0	2.0	2.0	3.0	4.0	0.0	2.0	7.0	3.0	5.0	2.0	5.0	3.0
O_terrae	Opitutus terrae PB90-1	0.0	0.0	2.0	1.0	1.0	0.0	2.0	4.0	0.0	2.0	0.0	1.0	1.0	3.0	3.0	2.0
P_parvula	Pedosphaera parvula Ellin514	0.0	0.0	6.0	0.0	3.0	2.0	4.0	6.0	0.0	0.0	7.0	4.0	4.0	2.0	3.0	3.0
G_obscuriglobus	Gemmata obscuriglobus UQM 2246	0.0	0.0	2.0	0.0	2.0	0.0	3.0	1.0	2.0	0.0	6.0	2.0	3.0	1.0	1.0	0.0
I_pallida	Isosphaera pallida ATCC 43644	0.0	0.0	2.0	0.0	2.0	0.0	3.0	1.0	0.0	0.0	5.0	0.0	2.0	1.0	2.0	0.0
M_infernorum	Methylacidiphilum infernorum V4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0
V_bacterium	Verrucomicrobiae bacterium DG1235	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	2.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0
L_biflexa	Leptospira biflexa serovar Patoc	2.0	1.0	22.0	11.0	18.0	9.0	31.0	16.0	17.0	5.0	29.0	14.0	37.0	13.0	16.0	15.0
L_interrogans	Leptospira interrogans serovar Lai str. 56601	2.0	1.0	22.0	9.0	19.0	9.0	28.0	15.0	17.0	5.0	27.0	16.0	32.0	10.0	13.0	14.0
T_parva	Turneriella parva DSM 21527	1.0	1.0	8.0	6.0	10.0	3.0	12.0	7.0	9.0	4.0	13.0	10.0	16.0	8.0	7.0	7.0
A_finegoldii	Alistipes finegoldii DSM 17242	1.0	0.0	7.0	4.0	16.0	2.0	24.0	10.0	6.0	4.0	26.0	26.0	20.0	5.0	6.0	14.0
A_muciniphila	Akkermansia muciniphila ATCC BAA-835	0.0	0.0	2.0	0.0	1.0	1.0	5.0	0.0	0.0	1.0	4.0	2.0	1.0	3.0	2.0	2.0
B_burgdorferii	Borrelia burgdorferi JD1	0.0	0.0	4.0	3.0	5.0	1.0	6.0	1.0	0.0	2.0	0.0	2.0	2.0	3.0	3.0	4.0
B_dutonii	Borrelia dutonii	0.0	0.0	3.0	3.0	5.0	0.0	6.0	1.0	1.0	3.0	0.0	2.0	2.0	3.0	3.0	4.0
B_intermedia	Brachyspira intermedia PWS/A	0.0	0.0	0.0	0.0	4.0	0.0	4.0	2.0	0.0	2.0	4.0	2.0	1.0	3.0	2.0	0.0
B_vulgatus	Bacteroides vulgatus ATCC 8482	3.0	0.0	17.0	7.0	31.0	6.0	37.0	20.0	6.0	13.0	41.0	37.0	36.0	11.0	13.0	32.0
B_xylanosolvens	Bacteroides xylanisolvens XB1A	3.0	0.0	17.0	7.0	27.0	7.0	50.0	22.0	8.0	14.0	32.0	37.0	38.0	6.0	13.0	32.0
C_orchracea	Capnocytophaga ochracea DSM 7271	0.0	0.0	6.0	0.0	14.0	2.0	12.0	4.0	3.0	5.0	12.0	10.0	11.0	1.0	6.0	7.0
P_asacharolytica	Porphyromonas asaccharolytica DSM 20707	0.0	0.0	3.0	2.0	8.0	2.0	10.0	8.0	5.0	2.0	8.0	7.0	11.0	4.0	6.0	11.0
P_dentalis	Prevotella dentalis DSM 3688	1.0	0.0	6.0	3.0	15.0	4.0	23.0	10.0	3.0	8.0	26.0	22.0	29.0	2.0	6.0	19.0
P_distasonis	Parabacteroides distasonis ATCC 8503	2.0	0.0	16.0	3.0	20.0	6.0	39.0	20.0	7.0	14.0	40.0	28.0	35.0	10.0	11.0	26.0
P_gingivalis	Porphyromonas gingivalis W83	0.0	0.0	5.0	2.0	13.0	2.0	13.0	10.0	4.0	4.0	14.0	16.0	10.0	3.0	5.0	12.0
P_intermedia	Prevotella intermedia 17	2.0	0.0	4.0	3.0	9.0	4.0	22.0	8.0	6.0	5.0	20.0	21.0	16.0	3.0	5.0	14.0
R_anatipestifer	Riemerella anatipestifer ATCC 11845 = DSM 15868	0.0	0.0	3.0	0.0	8.0	0.0	11.0	3.0	4.0	2.0	8.0	8.0	11.0	0.0	6.0	3.0
T_brennaborense	Treponema brennaborense DSM 12168	0.0	0.0	2.0	1.0	3.0	0.0	5.0	3.0	2.0	2.0	7.0	4.0	2.0	5.0	5.0	4.0
T_denticola	Treponema denticola ATCC 35405	0.0	0.0	2.0	1.0	3.0	0.0	3.0	3.0	1.0	3.0	8.0	8.0	5.0	7.0	1.0	2.0
T_forsythia	Tannerella forsythia ATCC 43037	0.0	0.0	6.0	3.0	18.0	4.0	20.0	17.0	4.0	11.0	19.0	23.0	20.0	6.0	7.0	17.0
T_pallidum	Treponema pallidum subsp. pallidum DAL-1	0.0	0.0	2.0	1.0	1.0	0.0	1.0	3.0	1.0	1.0	5.0	3.0	1.0	3.0	2.0	2.0
C_abortus	Chlamydophila abortus S26/3	2.0	0.0	1.0	1.0	6.0	1.0	6.0	3.0	2.0	3.0	5.0	7.0	5.0	2.0	7.0	7.0
C_caviae	Chlamydophila caviae GPIC	2.0	0.0	1.0	1.0	7.0	1.0	5.0	3.0	5.0	3.0	4.0	7.0	5.0	3.0	7.0	7.0
C_felis	Chlamydophila felis Fe/C-56	0.0	0.0	1.0	1.0	6.0	1.0	5.0	3.0	3.0	2.0	4.0	5.0	5.0	3.0	5.0	7.0
C_muridarum	Chlamydia muridarum Nigg	2.0	0.0	1.0	1.0	5.0	1.0	4.0	1.0	3.0	0.0	3.0	5.0	5.0	1.0	9.0	8.0
C_pecorum	Chlamydophila pecorum E58	1.0	0.0	1.0	1.0	4.0	1.0	3.0	2.0	3.0	1.0	5.0	5.0	5.0	3.0	6.0	5.0
C_pneumoniae	Chlamydophila pneumoniae J138	1.0	0.0	1.0	1.0	4.0	1.0	3.0	2.0	3.0	2.0	5.0	5.0	5.0	3.0	6.0	5.0
C_psitacii	Chlamydia psittaci NJ1	2.0	0.0	1.0	1.0	7.0	1.0	6.0	3.0	4.0	3.0	5.0	6.0	5.0	3.0	7.0	7.0
C_trachomatis	Chlamydia trachomatis 434/Bu	2.0	0.0	0.0	1.0	5.0	1.0	4.0	1.0	2.0	0.0	3.0	5.0	5.0	1.0	9.0	8.0
B_marina	Blastopirellula marina DSM 3645	1.0	0.0	26.0	5.0	22.0	5.0	27.0	14.0	9.0	31.0	24.0	18.0	30.0	22.0	18.0	16.0
C_akajimensis	Coraliomargarita akajimensis DSM 45221	0.0	0.0	1.0	3.0	3.0	0.0	7.0	2.0	0.0	2.0	2.0	2.0	7.0	0.0	2.0	0.0
C_limicola	Chlorobium limicola DSM 245	0.0	0.0	15.0	5.0	12.0	2.0	14.0	3.0	7.0	1.0	12.0	19.0	34.0	8.0	14.0	16.0
C_parvum	Chlorobaculum parvum NCIB 8327	1.0	0.0	19.0	5.0	11.0	1.0	9.0	5.0	5.0	4.0	12.0	14.0	22.0	6.0	10.0	18.0
C_phaeobacteroides	Chlorobium phaeobacteroides BS1	1.0	0.0	17.0	5.0	13.0	4.0	6.0	5.0	5.0	4.0	14.0	18.0	31.0	11.0	12.0	17.0
C_tepidum	Chlorobium tepidum TLS	0.0	0.0	15.0	5.0	10.0	1.0	8.0	5.0	6.0	2.0	11.0	11.0	18.0	7.0	10.0	18.0
C_thalassium	Chloroherpeton thalassium ATCC 35110	1.0	0.0	13.0	5.0	9.0	2.0	10.0	7.0	4.0	4.0	9.0	15.0	12.0	5.0	7.0	7.0
CK_stuttgartiensis	Candidatus Kuenenia stuttgartiensis	0.0	0.0	4.0	2.0	4.0	3.0	4.0	1.0	1.0	0.0	5.0	7.0	4.0	1.0	2.0	4.0
L_araneosa	Lentisphaera araneosa HTCC2155	0.0	0.0	1.0	6.0	8.0	1.0	11.0	7.0	0.0	9.0	8.0	8.0	7.0	6.0	3.0	3.0
P_aestuarii	Prosthecochloris aestuarii DSM 271	0.0	0.0	15.0	5.0	11.0	4.0	9.0	5.0	4.0	4.0	12.0	18.0	34.0	10.0	12.0	14.0
P_brasiliensis	Planctomyces brasiliensis DSM 5305	3.0	0.0	19.0	7.0	25.0	4.0	27.0	16.0	6.0	10.0	24.0	11.0	38.0	23.0	21.0	16.0
P_limnophilus	Planctomyces limnophilus DSM 3776	2.0	0.0	9.0	5.0	14.0	3.0	20.0	7.0	7.0	8.0	21.0	12.0	33.0	19.0	12.0	13.0
P_maris	Planctomyces maris DSM 8797	2.0	0.0	22.0	12.0	24.0	5.0	33.0	17.0	9.0	21.0	36.0	20.0	40.0	24.0	27.0	15.0
P_mikurensis	Phycisphaera mikurensis NBRC 102666	0.0	0.0	1.0	1.0	5.0	2.0	12.0	5.0	2.0	4.0	7.0	5.0	6.0	3.0	4.0	3.0
P_phaeoclathatiforme	Pelodictyon phaeoclathratiforme BU-1	1.0	0.0	21.0	4.0	11.0	2.0	10.0	5.0	6.0	2.0	11.0	20.0	26.0	9.0	11.0	18.0
P_staleyi	Pirellula staleyi DSM 6068	2.0	0.0	18.0	7.0	16.0	3.0	22.0	10.0	6.0	19.0	14.0	10.0	24.0	16.0	13.0	10.0
R_baltica	Rhodopirellula baltica SH 1	1.0	0.0	16.0	7.0	18.0	4.0	32.0	14.0	8.0	19.0	25.0	11.0	23.0	12.0	16.0	15.0
S_acidiphila	Singulisphaera acidiphila DSM 18658	0.0	0.0	15.0	2.0	6.0	3.0	14.0	7.0	6.0	6.0	11.0	8.0	16.0	3.0	7.0	11.0
S_africana	Spirochaeta africana DSM 8902	0.0	0.0	1.0	0.0	1.0	0.0	5.0	1.0	2.0	4.0	2.0	0.0	7.0	2.0	3.0	4.0
S_pleomorpha	Sphaerochaeta pleomorpha str. Grapes	0.0	0.0	1.0	0.0	2.0	0.0	7.0	1.0	2.0	6.0	3.0	2.0	6.0	0.0	1.0	2.0

Tables S3 : Functions for specific and transferred genes (quantity and percent) ,

Specific genes								percentages																					
Q	R	S	T	U	V	W	Z	A	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	Z
1.0	50.0	58.0	7.0	1.0	2.0	0.0	0.0	0.6	1.7	0.6	2.2	1.7	5.0	0.0	2.2	2.8	5.6	2.8	4.4	1.1	1.7	1.7	0.6	27.8	32.2	3.9	0.6	1.1	0.0	0.0
1.0	65.0	80.0	9.0	3.0	4.0	0.0	0.0	0.4	2.5	0.8	2.0	0.8	4.1	0.8	2.9	2.0	5.7	2.9	4.1	1.2	2.0	1.2	0.4	26.6	32.8	3.7	1.2	1.6	0.0	0.0
0.0	27.0	39.0	3.0	0.0	1.0	0.0	0.0	0.9	2.6	0.0	2.6	0.9	5.1	0.9	1.7	2.6	7.7	4.3	6.0	0.9	1.7	2.6	0.0	23.1	33.3	2.6	0.0	0.9	0.0	0.0
1.0	54.0	55.0	5.0	3.0	2.0	0.0	0.0	0.5	2.6	0.5	3.1	1.5	5.1	1.0	3.6	1.5	6.6	2.0	5.6	2.0	2.0	1.0	0.5	27.6	28.1	2.6	1.5	1.0	0.0	0.0
0.0	14.0	10.0	1.0	0.0	0.0	0.0	0.0	0.0	2.6	0.0	7.9	0.0	7.9	2.6	2.6	0.0	2.6	5.3	2.6	0.0	0.0	0.0	0.0	36.8	26.3	2.6	0.0	0.0	0.0	0.0
1.0	1.0	4.0	0.0	2.0	0.0	0.0	0.0	0.0	5.6	0.0	11.1	0.0	0.0	11.1	11.1	0.0	0.0	0.0	0.0	11.1	0.0	5.6	5.6	5.6	22.2	0.0	11.1	0.0	0.0	0.0
0.0	13.0	6.0	1.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	5.4	0.0	8.1	8.1	0.0	0.0	2.7	5.4	2.7	5.4	0.0	2.7	0.0	35.1	16.2	2.7	5.4	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	38.0	24.0	3.0	4.0	1.0	0.0	1.0	0.0	5.2	0.9	1.7	1.7	2.6	3.4	0.0	1.7	6.0	2.6	4.3	1.7	4.3	2.6	0.0	32.8	20.7	2.6	3.4	0.9	0.0	0.0
1.0	14.0	15.0	1.0	4.0	0.0	0.0	1.0	0.0	3.4	1.7	1.7	0.0	3.4	6.9	0.0	3.4	0.0	1.7	1.7	5.2	5.2	3.4	1.7	24.1	25.9	1.7	6.9	0.0	0.0	0.0
1.0	42.0	35.0	2.0	3.0	1.0	0.0	0.0	0.0	4.7	0.0	2.3	1.6	3.1	4.7	0.0	0.0	5.5	3.1	3.1	1.6	2.3	2.3	0.8	32.8	27.3	1.6	2.3	0.8	0.0	0.0
2.0	20.0	8.0	3.0	4.0	1.0	0.0	0.0	0.0	3.3	0.0	3.3	0.0	4.9	1.6	3.3	0.0	9.8	3.3	4.9	1.6	1.6	0.0	3.3	32.8	13.1	4.9	6.6	1.6	0.0	0.0
1.0	11.0	10.0	4.0	2.0	1.0	0.0	0.0	0.0	4.3	0.0	4.3	0.0	6.4	2.1	0.0	0.0	10.6	0.0	4.3	2.1	4.3	0.0	2.1	23.4	21.3	8.5	4.3	2.1	0.0	0.0
0.0	1.0	2.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	14.3	0.0	0.0	0.0	14.3	0.0	0.0	14.3	28.6	14.3	0.0	14.3	0.0	0.0
1.0	2.0	3.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.1	0.0	14.3	0.0	7.1	7.1	7.1	0.0	0.0	0.0	7.1	14.3	21.4	7.1	0.0	7.1	0.0	0.0
9.0	251.0	176.0	34.0	21.0	5.0	0.0	0.0	0.3	2.9	1.5	2.4	1.2	4.1	2.1	2.3	0.7	3.9	1.9	4.9	1.7	2.1	2.0	1.2	33.4	23.4	4.5	2.8	0.7	0.0	0.0
9.0	235.0	168.0	33.0	17.0	3.0	0.0	0.0	0.3	3.1	1.3	2.7	1.3	4.0	2.1	2.4	0.7	3.8	2.3	4.5	1.4	1.8	2.0	1.3	33.4	23.9	4.7	2.4	0.4	0.0	0.0
3.0	82.0	72.0	18.0	12.0	5.0	0.0	0.0	0.3	2.5	1.9	3.2	1.0	3.8	2.2	2.9	1.3	4.1	3.2	5.1	2.5	2.2	2.2	1.0	26.1	22.9	5.7	3.8	1.6	0.0	0.0
7.0	131.0	126.0	8.0	7.0	3.0	0.0	1.0	0.2	1.5	0.9	3.5	0.4	5.3	2.2	1.3	0.9	5.7	5.7	4.4	1.1	1.3	3.1	1.5	28.9	27.8	1.8	1.5	0.7	0.0	0.0
1.0	11.0	11.0	1.0	3.0	0.0	0.0	0.0	0.0	3.9	0.0	2.0	2.0	9.8	0.0	0.0	2.0	7.8	3.9	2.0	5.9	3.9	3.9	2.0	21.6	21.6	2.0	5.9	0.0	0.0	0.0
0.0	57.0	65.0	5.0	4.0	1.0	0.0	0.0	0.0	2.4	1.8	3.0	0.6	3.6	0.6	0.0	1.2	0.0	1.2	1.2	1.8	1.8	2.4	0.0	33.9	38.7	3.0	2.4	0.6	0.0	0.0
1.0	50.0	60.0	5.0	4.0	1.0	0.0	0.0	0.0	1.9	1.9	3.2	0.0	3.8	0.6	0.6	1.9	0.0	1.3	1.3	1.9	1.9	2.5	0.6	31.8	38.2	3.2	2.5	0.6	0.0	0.0
2.0	18.0	10.0	3.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	6.9	0.0	6.9	3.4	0.0	3.4	6.9	3.4	1.7	5.2	3.4	0.0	3.4	31.0	17.2	5.2	1.7	0.0	0.0	0.0
7.0	240.0	216.0	26.0	19.0	5.0	0.0	1.0	0.4	2.1	0.8	3.8	0.7	4.5	2.4	0.7	1.6	5.0	4.5	4.4	1.3	1.6	3.9	0.8	29.1	26.2	3.2	2.3	0.6	0.0	0.0
9.0	257.0	218.0	23.0	15.0	7.0	0.0	0.0	0.4	2.0	0.8	3.2	0.8	5.9	2.6	1.0	1.7	3.8	4.4	4.5	0.7	1.5	3.8	1.1	30.5	25.9	2.7	1.8	0.8	0.0	0.0
1.0	79.0	72.0	9.0	1.0	4.0	0.0	0.0	0.0	2.3	0.0	5.4	0.8	4.6	1.5	1.2	1.9	4.6	3.9	4.2	0.4	2.3	2.7	0.4	30.5	27.8	3.5	0.4	1.5	0.0	0.0
3.0	76.0	60.0	6.0	9.0	1.0	0.0	0.0	0.0	1.2	0.8	3.3	0.8	4.1	3.3	2.1	0.8	3.3	2.9	4.5	1.7	2.5	4.5	1.2	31.4	24.8	2.5	3.7	0.4	0.0	0.0
3.0	160.0	156.0	19.0	5.0	3.0	0.0	0.0	0.2	1.1	0.6	2.9	0.8	4.4	1.9	0.6	1.5	5.0	4.2	5.5	0.4	1.1	3.6	0.6	30.6	29.8	3.6	1.0	0.6	0.0	0.0
8.0	197.0	175.0	21.0	18.0	7.0	0.0	1.0	0.3	2.3	0.4	2.8	0.9	5.5	2.8	1.0	2.0	5.7	4.0	5.0	1.4	1.6	3.7	1.1	28.0	24.9	3.0	2.6	1.0	0.0	0.0
2.0	92.0	88.0	11.0	6.0	4.0	0.0	0.0	0.0	1.6	0.6	4.1	0.6	4.1	3.2	1.3	1.3	4.4	5.1	3.2	0.9	1.6	3.8	0.6	29.1	27.8	3.5	1.9	1.3	0.0	0.0
3.0	126.0	132.0	12.0	9.0	2.0	0.0	0.0	0.5	0.9	0.7	2.1	0.9	5.2	1.9	1.4	1.2	4.7	4.9	3.8	0.7	1.2	3.3	0.7	29.6	31.0	2.8	2.1	0.5	0.0	0.0
2.0	57.0	54.0	6.0	4.0	2.0	0.0	0.0	0.0	1.6	0.0	4.2	0.0	5.7	1.6	2.1	1.0	4.2	4.2	5.7	0.0	3.1	1.6	1.0	29.7	28.1	3.1	2.1	1.0	0.0	0.0
2.0	30.0	29.0	7.0	3.0	0.0	0.0	0.0	0.0	1.7	0.9	2.6	0.0	4.3	2.6	1.7	1.7	6.0	3.4	1.7	4.3	4.3	3.4	1.7	25.9	25.0	6.0	2.6	0.0	0.0	0.0
0.0	45.0	40.0	11.0	4.0	0.0	0.0	0.0	0.0	1.4	0.7	2.0	0.0	2.0	2.0	0.7	2.0	5.4	5.4	3.4	4.8	0.7	1.4	0.0	30.6	27.2	7.5	2.7	0.0	0.0	0.0
4.0	142.0	125.0	15.0	13.0	7.0	0.0	0.0	0.0	1.2	0.6	3.7	0.8	4.2	3.5	0.8	2.3	4.0	4.8	4.2	1.2	1.5	3.5	0.8	29.5	26.0	3.1	2.7	1.5	0.0	0.0
1.0	21.0	12.0	8.0	2.0	0.0	0.0	0.0	0.0	2.9	1.4	1.4	0.0	1.4	4.3	1.4	1.4	7.1	4.3	1.4	4.3	2.9	2.9	1.4	30.0	17.1	11.4	2.9	0.0	0.0	0.0
2.0	53.0	62.0	14.0	8.0	8.0	1.0	1.0	1.0	0.5	0.5	2.9	0.5	2.9	1.4	1.0	1.4	2.4	3.4	2.4	1.0	3.4	3.4	1.0	25.6	30.0	6.8	3.9	3.9	0.0	0.0
2.0	59.0	68.0	14.0	8.0	8.0	1.0	1.0	0.9	0.5	0.5	3.2	0.5	2.3	1.4	2.3	1.4	1.8	3.2	2.3	1.4	3.2	3.2	0.9	26.6	30.6	6.3	3.6	3.6	0.0	0.0
2.0	43.0	49.0	11.0	2.0	7.0	0.0	0.0	0.0	0.6	0.6	3.6	0.6	3.0	1.8	1.8	1.2	2.4	3.0	3.0	1.8	3.0	4.2	1.2	26.1	29.7	6.7	1.2	4.2	0.0	0.0
3.0	43.0	53.0	13.0	4.0	6.0	0.0	1.0	1.2	0.6	0.6	2.9	0.6	2.3	0.6	1.7	0.0	1.7	2.9	2.9	0.6	5.2	4.7	1.7	25.0	30.8	7.6	2.3	3.5	0.0	0.0
2.0	45.0	49.0	14.0	6.0	6.0	0.0	1.0	0.6	0.6	0.6	2.4	0.6	1.8	1.2	1.8	0.6	3.0	3.0	3.0	1.8	3.6	3.0	1.2	26.6	29.0	8.3	3.6	3.6	0.0	0.0
2.0	49.0	53.0	14.0	7.0	5.0	0.0	1.0	0.6	0.6	0.6	2.2	0.6	1.7	1.1	1.7	1.1	2.8	2.8	2.8	1.7	3.4	2.8	1.1	27.5	29.8	7.9	3.9	2.8	0.0	0.0
2.0	60.0	66.0	14.0	8.0	8.0	1.0	1.0	0.9	0.5	0.5	3.2	0.5	2.7	1.4	1.8	1.4	2.3	2.7	2.3	1.4	3.2	3.2	0.9	27.1	29.9	6.3	3.6	3.6	0.0	0.0
3.0	43.0	53.0	13.0	5.0	6.0	0.0	1.0	1.2	0.0	0.6	2.9	0.6	2.3	0.6	1.2	0.0	1.8	2.9	2.9	0.6	5.3	4.7	1.8	25.1	31.0	7.6	2.9	3.5	0.0	0.0
7.0	176.0	206.0	25.0	31.0	6.0	0.0	2.0	0.1	3.6	0.7	3.1	0.7	3.7	1.9	1.2	4.3	3.3	2.5	4.2	3.1	2.5	2.2	1.0	24.4	28.6	3.5	4.3	0.8	0.0	0.0
1.0	24.0	15.0	4.0	1.0	1.0	0.0	0.0	0.0	1.3	3.9	3.9	0.0	9.1	2.6	0.0	2.6	2.6	2.6	9.1	0.0	2.6	0.0	1.3	31.2	19.5	5.2	1.3	1.3	0.0	0.0
5.0	112.0	99.0	17.0	6.0	7.0	0.0	0.0	0.0	3.7	1.2	2.9	0.5	3.4	0.7	1.7	0.2	2.9	4.7	8.3	2.0	3.4	3.9	1.2	27.5	24.3	4.2	1.5	1.7	0.0	0.0
3.0	87.0	81.0	16.0	6.0	7.0	0.0	0.0	0.3	5.6	1.5	3.2	0.3	2.6	1.5	1.5	1.2	3.5	4.1	6.4	1.8	2.9	5.3	0.9	25.4	23.7	4.7	1.8	2.0	0.0	0.0
6.0	92.0	88.0	18.0	8.0	7.0	0.0	0.0	0.3	4.5	1.3	3.4	1.0	1.6	1.3	1.3	1.0	3.7	4.7	8.1	2.9	3.1	4.5	1.6	24.1	23.0	4.7	2.1	1.8	0.0	0.0
3.0	74.0	75.0	19.0	7.0	4.0	0.0	0.0	0.0	4.9	1.6	3.2	0.3	2.6	1.6	1.9	0.6	3.6	3.6	5.8	2.3	3.2	5.8	1.0	23.9	24.3	6.1	2.3	1.3	0.0	0.0
2.0	56.0	42.0	13.0	7.0	3.0	0.0	0.0	0.4	5.6	2.1	3.9	0.9	4.3	3.0	1.7	1.7	3.9	6.4	5.2	2.1	3.0	3.0	0.9	24.0	18.0	5.6	3.0	1.3	0.0	0.0
1.0	21.0	32.0	8.0	3.0	3.0	0.0	0.0	0.0	3.6	1.8	3.6	2.7	3.6	0.9	0.9	0.0	4.5	6.4	3.6	0.9	1.8	3.6	0.9	19.1	29.1	7.3	2.7	2.7	0.0	0.0
2.0	43.0	37.0	9.0	7.0	1.0	0.0	0.0	0.0	0.6	3.4	4.5	0.6	6.2	4.0	0.0	5.1	4.5	4.5	4.0	3.4	1.7	1.7	1.1	24.3	20.9	5.1	4.0	0.6	0.0	0.0
4.0	105.0	86.0	17.0	8.0	5.0	0.0	0.0	0.0	3.9	1.3	2.9	1.0	2.4	1.3	1.0	1.0	3.1	4.7	8.9	2.6	3.1	3.7	1.0	27.5	22.5	4.5	2.1	1.3	0.0	0.0
8.0	158.0	167.0	28.0	33.0	9.0	0.0	2.0	0.5	2.9	1.1	3.8	0.6	4.1	2.4	0.9	1.5	3.7	1.7	5.8	3.5	3.2	2.4	1.2	24.1	25.5	4.3	5.0	1.4	0.0	0.0
4.0	133.0	156.0	23.0	27.0	3.0	0.0	2.0	0.4	1.7	0.9	2.6	0.6	3.8	1.3	1.3	1.5	3.9	2.3	6.2	3.6	2.3	2.4	0.8	25.0	29.3	4.3	5.1	0.6	0.0	0.0
11.0	174.0	218.0	31.0	33.0	9.0	0.0	3.0	0.3	2.8	1.5	3.1	0.6	4.2	2.2	1.1	2.7	4.6	2.5	5.1	3.1	3.4	1.9	1.4	22.1	27.7	3.9	4.2	1.1	0.0	0.0
2.0	28.0	20.0	6.0	3.0	2.0	0.0	0.0	0.0	0.8	0.8	4.1	1.7	9.9	4.1	1.7	3.3	5.8	4.1	5.0	2.5	3.3	2.5	1.7	23.1	16.5	5.0	2.5	1.7	0.0	0.0
3.0	93.0	96.0	18.0	8.0	7.0	0.0	0.0	0.3	5.5	1.0	2.9	0.5	2.6	1.3	1.6	0.5	2.9	5.2	6.8	2.4	2.9	4.7	0.8	24.3	25.1	4.7	2.1	1.8	0.0	0.0
8.0	164.0	181.0	25.0	24.0	5.0	0.0	5.0	0.3	3.0	1.2	2.7	0.5	3.7	1.7	1.0	3.2	2.3	1.7	4.0	2.7	2.2	1.7	1.3	27.2	30.1	4.2	4.0	0.8	0.0	0.0
10.0	135.0	134.0	31.0	18.0	8.0	0.0	2.0	0.2	2.9	1.3	3.2	0.7	5.7	2.5	1.4	3.4	4.5	2.0	4.1	2.1	2.9	2.7	1.8	24.2	24.0	5.5	3.2	1.4	0.0	0.0
3.0	79.0	67.0	12.0	6.0	6.0	0.0	0.0	0.0	5.2	0.7	2.1	1.0	4.9	2.4	2.1	2.1	3.8	2.8	5.6	1.0	2.4	3.8	1.0	27.4	23.3	4.2	2.1	2.1	0.0	0.0
0.0	13.0	17.0	2.0	4.0	2.0	0.0	0.0	0.0	1.4	0.0	1.4	0.0	7.1	1.4	2.9	5.7	2.9	0.0	10.0	2.9	4.3	5.7	0.0	18.6	24.3	2.9	5.7	2.9	0.0	0.0
1.0	17.0	23.0	4.0	1.0	2.0	0.0	0.0	0.0	1.2	0.0	2.5	0.0	8.6	1.2	2.5	7.4	3.7	2.5	7.4	0.0	1.2	2.5	1.2	21.0	28.4	4.9	1.2	2.5	0.0	0.0

	Quantity																			
Complete name	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Candidatus Protochlamydia amoebophila	0.0	0.0	10.0	3.0	17.0	4.0	3.0	1.0	3.0	9.0	5.0	12.0	10.0	3.0	7.0	8.0	2.0	51.0	41.0	1.0
Parachlamydia acanthamoebae UV-7	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	4.0	0.0	2.0	2.0	5.0	0.0	2.0	1.0	1.0	23.0	16.0	3.0
Simkania negevensis Z	0.0	0.0	4.0	0.0	6.0	2.0	43.0	0.0	2.0	3.0	30.0	2.0	7.0	8.0	6.0	6.0	1.0	29.0	49.0	20.0
Waddlia chondrophila WSU 86-1044	0.0	0.0	0.0	0.0	2.0	2.0	6.0	1.0	2.0	0.0	1.0	2.0	4.0	0.0	1.0	3.0	2.0	16.0	12.0	2.0
Opitutaceae bacterium TAV1	0.0	0.0	15.0	9.0	18.0	4.0	18.0	6.0	4.0	6.0	17.0	6.0	30.0	14.0	12.0	13.0	6.0	77.0	74.0	40.0
Treponema primitia ZAS-2	0.0	0.0	0.0	5.0	9.0	0.0	9.0	2.0	7.0	3.0	17.0	10.0	14.0	7.0	2.0	8.0	5.0	97.0	71.0	19.0
Verrucomicrobium spinosum DSM 4136	0.0	0.0	10.0	2.0	15.0	4.0	13.0	5.0	16.0	3.0	16.0	7.0	15.0	11.0	12.0	7.0	5.0	70.0	54.0	39.0
Blattabacterium sp (Blaberus giganteus)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Chthoniobacter flavus Ellin428	1.0	1.0	24.0	1.0	46.0	3.0	47.0	29.0	28.0	8.0	77.0	41.0	62.0	18.0	33.0	26.0	21.0	251.0	271.0	104.0
Opitutus terrae PB90-1	0.0	0.0	2.0	0.0	11.0	2.0	7.0	0.0	2.0	3.0	4.0	7.0	5.0	1.0	2.0	2.0	2.0	41.0	33.0	5.0
Pedosphaera parvula Ellin514	0.0	0.0	1.0	3.0	2.0	1.0	1.0	0.0	1.0	0.0	5.0	2.0	3.0	1.0	1.0	3.0	0.0	42.0	41.0	5.0
Gemmata obscuriglobus UQM 2246	2.0	0.0	12.0	0.0	6.0	2.0	6.0	7.0	5.0	3.0	16.0	20.0	11.0	11.0	6.0	4.0	0.0	138.0	113.0	27.0
Isosphaera pallida ATCC 43644	1.0	0.0	12.0	1.0	10.0	5.0	14.0	18.0	3.0	3.0	16.0	13.0	19.0	2.0	13.0	13.0	10.0	59.0	41.0	22.0
Methylacidiphilum infernorum V4	0.0	0.0	15.0	1.0	34.0	2.0	44.0	15.0	4.0	2.0	95.0	4.0	17.0	44.0	3.0	21.0	4.0	101.0	121.0	14.0
Verrucomicrobiae bacterium DG1235	0.0	0.0	3.0	1.0	5.0	3.0	1.0	3.0	2.0	2.0	3.0	2.0	7.0	1.0	3.0	0.0	1.0	16.0	19.0	2.0
Leptospira biflexa serovar Patoc	0.0	0.0	7.0	2.0	6.0	1.0	9.0	7.0	11.0	4.0	13.0	8.0	10.0	10.0	6.0	6.0	7.0	92.0	84.0	22.0
Leptospira interrogans serovar Lai str. 56601	0.0	0.0	10.0	0.0	7.0	1.0	7.0	3.0	0.0	3.0	2.0	4.0	11.0	0.0	3.0	6.0	1.0	39.0	30.0	6.0
Turneriella parva DSM 21527	2.0	0.0	11.0	3.0	15.0	8.0	12.0	5.0	4.0	4.0	18.0	4.0	16.0	9.0	6.0	11.0	5.0	79.0	82.0	44.0
Alistipes finegoldii DSM 17242	0.0	0.0	22.0	0.0	8.0	7.0	21.0	7.0	1.0	5.0	7.0	14.0	15.0	0.0	7.0	11.0	2.0	62.0	53.0	6.0
Akkermansia muciniphila ATCC BAA-835	0.0	0.0	14.0	1.0	33.0	10.0	7.0	11.0	5.0	13.0	11.0	6.0	8.0	0.0	10.0	8.0	1.0	24.0	10.0	4.0
Borrelia burgdorferi JD1	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	0.0
Borrelia dutonii	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Brachyspira intermedia PWS/A	1.0	0.0	18.0	2.0	24.0	5.0	27.0	7.0	7.0	10.0	15.0	4.0	13.0	26.0	11.0	19.0	7.0	125.0	86.0	32.0
Bacteroides vulgatus ATCC 8482	0.0	1.0	18.0	4.0	23.0	9.0	44.0	14.0	17.0	10.0	20.0	19.0	46.0	2.0	13.0	17.0	7.0	122.0	70.0	13.0
Bacteroides xylanisolvens XB1A	0.0	0.0	3.0	1.0	4.0	1.0	20.0	3.0	4.0	0.0	7.0	7.0	5.0	2.0	4.0	5.0	1.0	51.0	36.0	1.0
Capnocytophaga ochracea DSM 7271	0.0	0.0	2.0	0.0	2.0	2.0	5.0	0.0	0.0	0.0	2.0	1.0	2.0	0.0	2.0	4.0	0.0	37.0	40.0	4.0
Porphyromonas asaccharolytica DSM 20707	0.0	0.0	4.0	0.0	5.0	1.0	4.0	7.0	1.0	0.0	9.0	5.0	2.0	4.0	4.0	5.0	4.0	42.0	41.0	11.0
Prevotella dentalis DSM 3688	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	3.0	1.0	1.0	3.0	3.0	0.0	1.0	0.0	1.0	11.0	12.0	1.0
Parabacteroides distasonis ATCC 8503	0.0	0.0	0.0	2.0	1.0	2.0	3.0	1.0	1.0	0.0	2.0	11.0	2.0	0.0	2.0	1.0	0.0	23.0	11.0	1.0
Porphyromonas gingivalis W83	0.0	0.0	0.0	0.0	1.0	0.0	2.0	0.0	1.0	1.0	0.0	3.0	1.0	2.0	2.0	0.0	0.0	22.0	13.0	1.0
Prevotella intermedia 17	0.0	0.0	13.0	0.0	12.0	2.0	9.0	8.0	7.0	1.0	20.0	10.0	15.0	2.0	10.0	7.0	5.0	91.0	108.0	17.0
Riemerella anatipestifer ATCC 11845 = DSM 15868	0.0	0.0	12.0	1.0	7.0	1.0	10.0	10.0	6.0	4.0	22.0	7.0	21.0	3.0	5.0	5.0	5.0	78.0	52.0	16.0
Treponema brennaborense DSM 12168	1.0	0.0	0.0	0.0	4.0	2.0	2.0	5.0	1.0	0.0	2.0	3.0	14.0	4.0	4.0	8.0	1.0	28.0	35.0	3.0
Treponema denticola ATCC 35405	0.0	0.0	4.0	0.0	2.0	0.0	2.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	2.0	1.0	0.0	7.0	4.0	1.0
Tannerella forsythia ATCC 43037	0.0	0.0	9.0	2.0	8.0	1.0	6.0	5.0	24.0	3.0	10.0	8.0	10.0	5.0	10.0	6.0	7.0	109.0	80.0	38.0
Treponema pallidum subsp. pallidum DAL-1	0.0	0.0	28.0	2.0	82.0	3.0	34.0	21.0	3.0	5.0	35.0	13.0	10.0	3.0	6.0	46.0	5.0	102.0	72.0	40.0
Chlamydophila abortus S26/3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Chlamydophila caviae GPIC	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	8.0	0.0
Chlamydophila felis Fe/C-56	0.0	0.0	0.0	0.0	6.0	1.0	3.0	2.0	4.0	0.0	2.0	2.0	0.0	8.0	4.0	0.0	0.0	8.0	3.0	3.0
Chlamydia muridarum Nigg	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	2.0	2.0	2.0	1.0	0.0	0.0	15.0	10.0	0.0
Chlamydophila pecorum E58	0.0	0.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	9.0	7.0	0.0
Chlamydophila pneumoniae J138	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	13.0	10.0	1.0
Chlamydia psittaci NJ1	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	3.0	1.0	0.0
Chlamydia trachomatis 434/Bu	0.0	0.0	1.0	0.0	3.0	0.0	0.0	1.0	2.0	0.0	0.0	1.0	0.0	3.0	2.0	1.0	0.0	12.0	7.0	0.0
Blastopirellula marina DSM 3645	0.0	0.0	12.0	0.0	22.0	1.0	8.0	12.0	8.0	4.0	23.0	12.0	14.0	10.0	10.0	10.0	7.0	111.0	79.0	29.0
Coraliomargarita akajimensis DSM 45221	0.0	0.0	4.0	1.0	10.0	3.0	15.0	9.0	4.0	2.0	12.0	4.0	16.0	6.0	7.0	16.0	4.0	50.0	53.0	9.0
Chlorobium limicola DSM 245	0.0	0.0	2.0	0.0	1.0	0.0	2.0	0.0	0.0	0.0	1.0	3.0	4.0	1.0	0.0	3.0	1.0	35.0	22.0	0.0
Chlorobaculum parvum NCIB 8327	0.0	0.0	6.0	0.0	1.0	2.0	4.0	3.0	0.0	2.0	0.0	1.0	3.0	0.0	3.0	5.0	0.0	21.0	24.0	3.0
Chlorobium phaeobacteroides BS1	0.0	0.0	4.0	2.0	9.0	1.0	2.0	4.0	4.0	1.0	9.0	6.0	6.0	6.0	5.0	2.0	1.0	49.0	39.0	10.0
Chlorobium tepidum TLS	1.0	0.0	4.0	1.0	1.0	0.0	2.0	5.0	2.0	0.0	2.0	5.0	3.0	2.0	1.0	3.0	0.0	15.0	22.0	1.0
Chloroherpeton thalassium ATCC 35110	0.0	0.0	24.0	1.0	11.0	2.0	9.0	16.0	6.0	2.0	4.0	1.0	8.0	3.0	8.0	15.0	2.0	70.0	57.0	17.0
Candidatus Kuenenia stuttgartiensis	0.0	0.0	59.0	7.0	15.0	3.0	11.0	31.0	6.0	5.0	15.0	22.0	32.0	11.0	21.0	18.0	10.0	131.0	119.0	35.0
Lentisphaera araneosa HTCC2155	1.0	0.0	9.0	0.0	11.0	2.0	29.0	5.0	7.0	12.0	36.0	21.0	14.0	19.0	7.0	3.0	3.0	121.0	93.0	28.0
Prosthecochloris aestuarii DSM 271	0.0	0.0	2.0	0.0	1.0	0.0	1.0	2.0	2.0	0.0	2.0	2.0	1.0	1.0	2.0	1.0	0.0	12.0	7.0	1.0
Planctomyces brasiliensis DSM 5305	0.0	0.0	3.0	0.0	4.0	1.0	16.0	3.0	2.0	2.0	12.0	7.0	19.0	2.0	4.0	6.0	1.0	63.0	35.0	7.0
Planctomyces limnophilus DSM 3776	1.0	0.0	7.0	2.0	8.0	4.0	15.0	9.0	7.0	2.0	19.0	9.0	16.0	11.0	0.0	9.0	9.0	60.0	57.0	12.0
Planctomyces maris DSM 8797	0.0	0.0	23.0	1.0	10.0	3.0	10.0	8.0	10.0	2.0	23.0	7.0	24.0	20.0	9.0	5.0	11.0	161.0	145.0	50.0
Phycisphaera mikurensis NBRC 102666	0.0	0.0	8.0	0.0	2.0	0.0	3.0	4.0	3.0	1.0	3.0	6.0	6.0	1.0	2.0	5.0	0.0	39.0	32.0	2.0
Pelodictyon phaeoclathratiforme BU-1	0.0	0.0	1.0	1.0	3.0	0.0	2.0	8.0	7.0	4.0	13.0	8.0	6.0	0.0	6.0	4.0	3.0	48.0	43.0	4.0
Pirellula staleyi DSM 6068	0.0	0.0	8.0	1.0	11.0	3.0	5.0	3.0	4.0	2.0	15.0	8.0	10.0	3.0	4.0	9.0	4.0	61.0	48.0	14.0
Rhodopirellula baltica SH 1	0.0	0.0	2.0	0.0	10.0	2.0	24.0	2.0	7.0	2.0	17.0	5.0	7.0	5.0	5.0	6.0	1.0	42.0	49.0	31.0
Singulisphaera acidiphila DSM 18658	0.0	0.0	4.0	0.0	13.0	3.0	18.0	3.0	1.0	0.0	4.0	1.0	16.0	0.0	3.0	3.0	0.0	42.0	47.0	1.0
Spirochaeta africana DSM 8902	0.0	0.0	20.0	2.0	29.0	6.0	139.0	2.0	3.0	0.0	49.0	6.0	8.0	0.0	4.0	17.0	7.0	118.0	81.0	29.0
Sphaerochaeta pleomorpha str. Grapes	0.0	0.0	6.0	0.0	16.0	2.0	6.0	5.0	4.0	2.0	12.0	3.0	9.0	3.0	5.0	22.0	2.0	92.0	86.0	14.0

HGT				Percentages																						
U	V	W	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	Z
10.0	1.0	0.0	0.0	0.0	0.0	5.0	1.5	8.5	2.0	1.5	0.5	1.5	4.5	2.5	6.0	5.0	1.5	3.5	4.0	1.0	25.4	20.4	0.5	5.0	0.5	0.0	0.0
8.0	1.0	0.0	0.0	0.0	0.0	1.5	0.0	8.3	1.5	5.3	0.0	1.5	2.3	3.0	5.3	3.8	0.8	1.5	1.5	1.5	30.8	24.8	3.8	0.8	2.3	0.0	0.0
3.0	3.0	0.0	0.0	0.0	0.0	2.4	0.0	7.9	1.8	11.0	1.8	0.6	0.0	2.4	0.6	9.8	0.0	1.8	1.8	0.0	25.6	28.7	0.6	1.2	1.8	0.0	0.0
0.0	3.0	0.0	0.0	0.0	0.0	4.0	1.3	6.7	4.0	1.3	4.0	2.7	2.7	4.0	2.7	9.3	1.3	4.0	0.0	1.3	21.3	25.3	2.7	1.3	0.0	0.0	0.0
9.0	10.0	0.0	1.0	0.0	0.0	2.5	0.2	5.7	0.3	7.4	2.5	0.7	0.3	16.0	0.7	2.9	7.4	0.5	3.5	0.7	17.0	20.3	2.4	8.2	0.8	0.0	0.0
9.0	3.0	0.0	3.0	0.0	0.0	5.4	0.4	15.9	0.6	6.6	4.1	0.6	1.0	6.8	2.5	1.9	0.6	1.2	8.9	1.0	19.7	13.9	7.7	0.6	0.8	0.0	0.0
2.0	2.0	0.0	0.0	0.0	0.0	0.0	1.7	3.0	0.0	3.0	0.7	2.3	1.0	5.7	3.3	4.7	2.3	0.7	2.7	1.7	32.3	23.7	6.3	3.0	1.0	0.0	1.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
20.0	18.0	0.0	0.0	0.1	0.1	2.1	0.1	4.1	0.3	4.2	2.6	2.5	0.7	6.8	3.6	5.5	1.6	2.9	2.3	1.9	22.2	24.0	9.2	1.8	1.6	0.0	0.0
1.0	3.0	0.0	0.0	0.0	0.0	3.9	2.3	4.6	1.0	4.6	1.5	1.0	1.5	4.4	1.5	7.7	3.6	3.1	3.3	1.5	19.8	19.0	10.3	2.3	2.6	0.0	0.3
1.0	1.0	0.0	0.0	0.0	0.0	4.1	0.2	1.8	0.5	1.8	1.4	1.8	0.4	4.1	1.3	4.3	3.6	1.6	0.9	2.0	28.8	25.9	8.9	5.9	0.7	0.0	0.2
7.0	5.0	0.0	1.0	0.5	0.0	3.0	0.0	1.5	0.5	1.5	1.7	1.2	0.7	4.0	5.0	2.7	2.7	1.5	1.0	0.0	34.3	28.1	6.7	1.7	1.2	0.0	0.2
5.0	2.0	0.0	0.0	0.4	0.0	4.3	0.4	3.5	1.8	5.0	6.4	1.1	1.1	5.7	4.6	6.7	0.7	4.6	4.6	3.5	20.9	14.5	7.8	1.8	0.7	0.0	0.0
49.0	5.0	0.0	0.0	0.0	0.0	7.4	0.0	5.2	0.7	5.2	2.2	0.0	2.2	1.5	3.0	8.1	0.0	2.2	4.4	0.7	28.9	22.2	4.4	0.7	0.7	0.0	0.0
1.0	0.0	0.0	0.0	0.6	0.0	3.1	0.8	4.2	2.2	3.4	1.4	1.1	1.1	5.0	1.1	4.5	2.5	1.7	3.1	1.4	22.1	22.9	12.3	3.4	2.2	0.0	0.0
6.0	4.0	0.0	0.0	0.0	0.0	3.2	0.6	4.9	1.3	4.2	1.6	5.2	1.0	5.2	2.3	4.9	3.6	3.9	2.3	1.6	22.7	17.5	12.7	0.6	0.6	0.0	0.0
1.0	1.0	0.0	0.0	0.0	0.0	2.2	0.6	1.9	0.3	2.9	2.2	3.5	1.3	4.1	2.5	3.2	3.2	1.9	1.9	2.2	29.2	26.7	7.0	1.9	1.3	0.0	0.0
12.0	8.0	0.0	0.0	0.0	0.0	2.6	0.6	2.3	0.3	1.7	1.4	6.9	0.9	2.9	2.3	2.9	1.4	2.9	1.7	2.0	31.1	22.9	10.9	0.6	2.0	0.0	0.0
3.0	7.0	0.0	0.0	0.0	0.0	8.5	0.0	3.1	2.7	8.1	2.7	0.4	1.9	2.7	5.4	5.8	0.0	2.7	4.3	0.8	24.0	20.5	2.3	1.2	2.7	0.0	0.0
2.0	7.0	0.0	0.0	0.0	0.0	7.6	0.5	17.8	5.4	3.8	5.9	2.7	7.0	5.9	3.2	4.3	0.0	5.4	4.3	0.5	13.0	5.4	2.2	1.1	3.8	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6.0	12.0	0.0	0.0	0.2	0.0	3.9	0.4	5.3	1.1	5.9	1.5	1.5	2.2	3.3	0.9	2.8	5.7	2.4	4.2	1.5	27.4	18.8	7.0	1.3	2.6	0.0	0.0
12.0	16.0	0.0	0.0	0.0	0.2	3.6	0.8	4.6	1.8	8.9	2.8	3.4	2.0	4.0	3.8	9.3	0.4	2.6	3.4	1.4	24.5	14.1	2.6	2.4	3.2	0.0	0.0
3.0	2.0	0.0	0.0	0.0	0.0	1.9	0.6	2.5	0.6	12.5	1.9	2.5	0.0	4.4	4.4	3.1	1.3	2.5	3.1	0.6	31.9	22.5	0.6	1.9	1.3	0.0	0.0
0.0	2.0	0.0	0.0	0.0	0.0	1.9	0.0	1.9	1.9	4.8	0.0	0.0	0.0	1.9	1.0	1.9	0.0	1.9	3.8	0.0	35.2	38.1	3.8	0.0	1.9	0.0	0.0
6.0	4.0	0.0	0.0	0.0	0.0	5.3	0.0	2.6	0.0	2.6	5.3	5.3	0.0	5.3	5.3	2.6	2.6	5.3	2.6	0.0	31.6	18.4	2.6	2.6	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.4	3.4	10.2	1.7	3.4	0.0	1.7	3.4	6.8	0.0	1.7	5.1	3.4	27.1	20.3	3.4	0.0	5.1	0.0	0.0
1.0	2.0	0.0	0.0	0.0	0.0	1.5	0.0	2.0	0.5	8.1	1.5	1.0	1.0	6.1	3.5	9.6	1.0	2.0	3.0	0.5	31.8	17.7	3.5	4.5	1.0	0.0	0.0
2.0	0.0	0.0	0.0	0.0	0.0	2.6	0.0	2.6	0.0	0.0	0.0	7.7	2.6	2.6	7.7	7.7	0.0	2.6	0.0	2.6	28.2	30.8	2.6	0.0	0.0	0.0	0.0
8.0	4.0	0.0	0.0	0.0	0.0	0.0	3.0	1.5	3.0	4.5	1.5	1.5	0.0	3.0	16.7	3.0	0.0	3.0	1.5	0.0	34.8	16.7	1.5	1.5	3.0	0.0	0.0
6.0	0.0	0.0	0.0	0.0	0.0	0.6	0.6	1.8	0.0	1.2	4.9	4.3	2.5	8.0	4.9	3.7	0.0	3.7	2.5	1.8	29.4	26.4	2.5	0.6	0.6	0.0	0.0
2.0	5.0	0.0	0.0	0.0	0.0	1.8	0.0	2.7	0.9	19.2	0.0	0.9	1.3	13.4	0.9	3.1	3.6	2.7	2.7	0.4	12.9	21.9	8.9	1.3	1.3	0.0	0.0
1.0	0.0	0.0	0.0	0.0	0.0	1.8	0.0	4.9	0.6	1.8	1.5	1.2	0.6	3.7	0.9	2.8	0.9	1.5	6.7	0.6	28.2	26.4	4.3	0.0	11.3	0.0	0.0
2.0	7.0	0.0	0.0	0.8	0.0	0.0	0.0	3.2	1.6	1.6	4.0	0.8	0.0	1.6	2.4	11.3	3.2	3.2	6.5	0.8	22.6	28.2	2.4	1.6	4.0	0.0	0.0
3.0	4.0	0.0	0.0	0.0	0.0	15.4	0.0	7.7	0.0	7.7	0.0	0.0	0.0	3.8	0.0	3.8	0.0	7.7	3.8	0.0	26.9	15.4	3.8	3.8	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.3	0.0	0.0	6.3	0.0	0.0	6.3	0.0	0.0	0.0	0.0	0.0	0.0	31.3	50.0	0.0	0.0	0.0	0.0	0.0
0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	12.8	2.1	6.4	4.3	8.5	0.0	4.3	4.3	0.0	17.0	8.5	0.0	0.0	17.0	6.4	6.4	0.0	2.1	0.0	0.0
1.0	0.0	0.0	0.0	2.8	0.0	0.0	0.0	0.0	2.8	0.0	0.0	2.8	0.0	0.0	5.6	5.6	5.6	2.8	0.0	0.0	41.7	27.8	0.0	2.8	0.0	0.0	0.0
1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.5	4.5	4.5	0.0	0.0	4.5	4.5	0.0	0.0	0.0	0.0	0.0	0.0	40.9	31.8	0.0	4.5	0.0	0.0	0.0
1.0	0.0	0.0	0.0	0.0	0.0	0.0	3.6	0.0	0.0	0.0	0.0	3.6	3.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	46.4	35.7	3.6	3.6	0.0	0.0	0.0
1.0	0.0	0.0	0.0	0.0	0.0	0.0	11.1	0.0	11.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.1	11.1	0.0	0.0	33.3	11.1	0.0	11.1	0.0	0.0	0.0
5.0	0.0	0.0	0.0	0.0	0.0	2.6	0.0	7.9	0.0	0.0	2.6	5.3	0.0	0.0	2.6	0.0	7.9	5.3	2.6	0.0	31.6	18.4	0.0	13.2	0.0	0.0	0.0
16.0	2.0	0.0	0.0	0.0	0.0	3.1	0.0	5.6	0.3	2.1	3.1	2.1	1.0	5.9	3.1	3.6	2.6	2.6	2.6	1.8	28.5	20.3	7.4	4.1	0.5	0.0	0.0
7.0	2.0	0.0	0.0	0.0	0.0	1.7	0.4	4.3	1.3	6.4	3.8	1.7	0.9	5.1	1.7	6.8	2.6	3.0	6.8	1.7	21.4	22.6	3.8	3.0	0.9	0.0	0.0
3.0	0.0	0.0	0.0	0.0	0.0	2.6	0.0	1.3	0.0	2.6	0.0	0.0	0.0	1.3	3.8	5.1	1.3	0.0	3.8	1.3	44.9	28.2	0.0	3.8	0.0	0.0	0.0
3.0	0.0	0.0	0.0	0.0	0.0	7.4	0.0	1.2	2.5	4.9	3.7	0.0	2.5	0.0	1.2	3.7	0.0	3.7	6.2	0.0	25.9	29.6	3.7	3.7	0.0	0.0	0.0
7.0	0.0	0.0	2.0	1.4	0.0	5.4	1.4	1.4	0.0	2.7	6.8	2.7	0.0	2.7	6.8	4.1	2.7	1.4	4.1	0.0	20.3	29.7	1.4	2.7	2.7	0.0	0.0
2.0	2.0	0.0	0.0	0.0	0.0	9.2	0.4	4.2	0.8	3.4	6.1	2.3	0.8	1.5	0.4	3.1	1.1	3.1	5.7	0.8	26.8	21.8	6.5	1.5	0.4	0.0	0.0
4.0	1.0	0.0	0.0	0.0	0.0	10.3	1.2	2.6	0.5	1.9	5.4	1.0	0.9	2.6	3.8	5.6	1.9	3.7	3.1	1.7	22.9	20.8	6.1	2.4	0.7	0.0	0.5
14.0	4.0	0.0	3.0	0.2	0.0	2.0	0.0	2.5	0.5	6.6	1.1	1.6	2.7	8.2	4.8	3.2	4.3	1.6	0.7	0.7	27.4	21.1	6.3	4.5	0.0	0.0	0.0
20.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.4	1.4	1.4	1.4	5.6	0.0	2.8	2.8	6.9	0.0	2.8	1.4	1.4	31.9	22.2	4.2	11.1	1.4	0.0	0.0
1.0	0.0	0.0	0.0	0.0	0.0	2.5	0.0	3.1	0.6	2.5	4.4	0.6	0.0	5.7	3.1	1.3	2.5	2.5	3.1	2.5	26.4	25.8	6.9	3.8	2.5	0.0	0.0
9.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	3.9	0.0	2.0	2.0	0.0	5.9	2.0	3.9	3.9	0.0	0.0	43.1	25.5	2.0	3.9	0.0	0.0	0.0
13.0	0.0	0.0	0.0	0.0	0.0	3.7	0.0	3.4	0.6	2.6	2.3	2.0	0.3	5.7	2.9	4.3	0.6	2.9	2.0	1.4	26.1	30.9	4.9	2.3	1.1	0.0	0.0
33.0	4.0	0.0	1.0	0.4	0.0	2.6	0.7	3.0	1.5	5.6	3.3	2.6	0.7	7.0	3.3	5.9	4.1	0.0	3.3	3.3	22.2	21.1	4.4	4.8	0.0	0.0	0.0
1.0	0.0	0.0	0.0	0.0	0.0	6.8	0.0	1.7	0.0	2.5	3.4	2.5	0.8	2.5	5.1	5.1	0.8	1.7	4.2	0.0	33.1	27.1	1.7	0.8	0.0	0.0	0.0
1.0	1.0	0.0	0.0	0.0	0.0	0.9	2.6	1.8	0.9	0.9	0.0	0.9	0.0	4.4	1.8	2.6	0.9	0.9	2.6	0.0	36.8	36.0	4.4	0.9	0.9	0.0	0.0
8.0	3.0	0.0	0.0	0.0	0.0	2.4	1.2	5.3	0.6	1.2	2.4	2.4	0.6	5.3	3.6	3.6	3.6	3.0	1.2	0.6	29.0	23.1	5.9	4.1	0.0	0.0	1.2
0.0	4.0	0.0	0.0	0.0	0.0	3.6	0.4	4.9	1.3	2.2	1.3	1.8	0.9	6.7	3.6	4.5	1.3	1.8	4.0	1.8	27.2	21.4	6.3	3.6	1.3	0.0	0.0
2.0	3.0	0.0	0.0	0.0	0.0	4.4	0.4	2.6	0.4	3.7	3.7	2.2	1.5	8.1	2.6	7.7	1.1	1.8	1.8	1.8	28.8	19.2	5.9	2.2	0.0	0.0	0.0
2.0	8.0	0.0	0.0	0.0	0.0	0.9	0.0	4.5	0.9	10.9	0.9	3.2	0.9	7.7	2.3	3.2	2.3	2.3	2.7	0.5	19.0	22.2	14.0	0.0	1.8	0.0	0.0
0.0	37.0	0.0	0.0	0.0	0.0	3.8	0.4	5.5	1.1	26.2	0.4	0.6	0.0	9.2	1.1	1.5	0.0	0.8	3.2	1.3	22.3	15.3	5.5	0.4	1.5	0.0	0.0

Species		Habitats		Bacterial partners							
		Environment	Lifestyle	Bacteria	Acidobacteria	Actinobacteria	Aquificae	Chloroflexi	Chrysiogenetes	Cyanobacteria	Deinococcus
CP_amoebophila	Candidatus Protochlamydia amoebophila	Amoeba	Intracellular	1371	5	52	0	17	0	150	23
P_acanthamoeba	Parachlamydia acanthamoebae UV-7	Amoeba	Intracellular	1081	8	33	3	3	0	72	7
S_negevensis	Simkania negevensis Z	Amoeba	Intracellular	1526	4	142	7	14	0	84	12
W_chondrophila	Waddlia chondrophila WSU 86-1044	Amoeba	Intracellular	531	2	54	1	1	0	22	17
O_bacterium	Opitutaceae bacterium TAV1	Insects	Extracellular	5141	50	561	12	64	0	171	24
T_primitia	Treponema primitia ZAS-2	Insects	Extracellular	5007	12	217	4	12	0	144	11
V_spinosum	Verrucomicrobium spinosum DSM 4136	Insects	Extracellular	2338	37	117	6	23	0	208	17
Blattabacterium_sp	Blattabacterium sp (Blaberus giganteus)	Insects	Intracellular	0	0	0	0	0	0	0	0
C_flavus	Chthoniobacter flavus Ellin428	Soils	Extracellular	25	0	0	0	0	0	0	0
O_terrae	Opitutus terrae PB90-1	Soils	Extracellular	3710	358	235	4	39	0	285	26
P_parvula	Pedosphaera parvula Ellin514	Soils	Extracellular	4870	188	637	11	75	0	862	49
G_obscuriglobus	Gemmata obscuriglobus UQM 2246	Soils, water	Extracellular	3269	34	402	3	102	2	468	21
I_pallida	Isosphaera pallida ATCC 43644	Soils, water	Extracellular	2785	72	269	12	95	0	357	30
M_infernorum	Methylacidiphilum infernorum V4	Soils, water	Extracellular	1409	28	67	37	20	0	63	74
V_bacterium	Verrucomicrobiae bacterium DG1235	Soils, water	Extracellular	3566	92	217	1	30	0	214	24
L_biflexa	Leptospira biflexa serovar Patoc	Ubiquitous	Extracellular	3505	41	308	11	47	2	240	15
L_interrogans	Leptospira interrogans serovar Lai str. 56601	Ubiquitous	Extracellular	2547	20	230	6	22	2	277	2
T_parva	Turneriella parva DSM 21527	Ubiquitous	Extracellular	3402	12	591	11	44	1	226	18
A_finegoldii	Alistipes finegoldii DSM 17242	Vertebrates	Extracellular	2916	15	227	5	18	2	116	31
A_muciniphila	Akkermansia muciniphila ATCC BAA-835	Vertebrates	Extracellular	2338	24	58	6	27	0	147	50
B_burgdorferii	Borrelia burgdorferi JD1	Vertebrates	Extracellular	0	0	0	0	0	0	0	0
B_dutonii	Borrelia dutonii BgVir	Vertebrates	Extracellular	0	0	0	0	0	0	0	0
B_intermedia	Brachyspira intermedia PWS/A	Vertebrates	Extracellular	4289	7	78	8	2	1	79	2
B_vulgatus	Bacteroides vulgatus ATCC 8482	Vertebrates	Extracellular	4454	71	179	3	8	3	121	11
B_xylanosolvens	Bacteroides xylanisolvens XB1A	Vertebrates	Extracellular	1331	34	82	0	2	0	44	0
C_orchracea	Capnocytophaga ochracea DSM 7271	Vertebrates	Extracellular	1126	11	136	3	0	0	41	16
P_asacharolytica	Porphyromonas asaccharolytica DSM 20707	Vertebrates	Extracellular	315	1	3	0	0	0	6	0
P_dentalis	Prevotella dentalis DSM 3688	Vertebrates	Extracellular	475	5	31	0	5	1	11	0
P_distasonis	Parabacteroides distasonis ATCC 8503	Vertebrates	Extracellular	1634	18	105	0	10	2	24	1
P_gingivalis	Porphyromonas gingivalis W83	Vertebrates	Extracellular	373	0	13	0	1	0	8	0
P_intermedia	Prevotella intermedia 17	Vertebrates	Extracellular	759	2	16	3	5	0	53	13
R_anatipestifer	Riemerella anatipestifer ATCC 11845 = DSM 15868	Vertebrates	Extracellular	1347	19	64	4	16	0	49	1
T_brennaborense	Treponema brennaborense DSM 12168	Vertebrates	Extracellular	2751	1	130	0	13	1	19	5
T_denticola	Treponema denticola ATCC 35405	Vertebrates	Extracellular	3365	0	118	3	8	1	58	1
T_forsythia	Tannerella forsythia ATCC 43037	Vertebrates	Extracellular	856	6	51	0	6	0	59	9
T_pallidum	Treponema pallidum subsp. pallidum DAL-1	Vertebrates	Extracellular	328	2	4	0	1	0	0	8
C_abortus	Chlamydophila abortus S26/3	Vertebrates	Intracellular	0	0	0	0	0	0	0	0
C_caviae	Chlamydophila caviae GPIC	Vertebrates	Intracellular	69	0	0	0	0	0	0	0
C_felis	Chlamydophila felis Fe/C-56	Vertebrates	Intracellular	0	0	0	0	0	0	0	0
C_muridarum	Chlamydia muridarum Nigg	Vertebrates	Intracellular	58	0	2	0	5	0	7	0
C_pecorum	Chlamydophila pecorum E58	Vertebrates	Intracellular	54	0	0	0	0	0	0	0
C_pneumoniae	Chlamydophila pneumoniae J138	Vertebrates	Intracellular	52	0	0	0	0	0	0	0
C_psitacii	Chlamydia psittaci NJ1	Vertebrates	Intracellular	48	0	2	0	0	0	0	0
C_trachomatis	Chlamydia trachomatis 434/Bu	Vertebrates	Intracellular	82	0	0	0	0	0	1	0
B_marina	Blastopirellula marina DSM 3645	Water	Extracellular	3820	94	357	7	82	1	380	36
C_akajimensis	Coraliomargarita akajimensis DSM 45221	Water	Extracellular	2309	17	73	8	14	1	181	3
C_limicola	Chlorobium limicola DSM 245	Water	Extracellular	477	6	4	0	14	0	90	0
C_parvum	Chlorobaculum parvum NCIB 8327	Water	Extracellular	541	5	14	11	8	0	57	0
C_phaeobacteroides	Chlorobium phaeobacteroides BS1	Water	Extracellular	795	2	30	21	14	0	75	8
C_tepidum	Chlorobium tepidum TLS	Water	Extracellular	464	5	16	3	6	0	28	1
C_thalassium	Chloroherpeton thalassium ATCC 35110	Water	Extracellular	1476	46	80	19	125	1	135	30
CK_stuttgartiensis	Candidatus Kuenenia stuttgartiensis	Water	Extracellular	6085	63	138	94	90	3	591	37
L_araneosa	Lentisphaera araneosa HTCC2155	Water	Extracellular	3432	8	83	17	16	2	208	11
P_aestuarii	Prosthecochloris aestuarii DSM 271	Water	Extracellular	386	0	7	1	11	0	38	1
P_brasiliensis	Planctomyces brasiliensis DSM 5305	Water	Extracellular	1379	19	80	1	11	0	155	16
P_limnophilus	Planctomyces limnophilus DSM 3776	Water	Extracellular	521	3	15	2	2	0	65	2
P_maris	Planctomyces maris DSM 8797	Water	Extracellular	3347	43	277	5	63	0	296	21
P_mikurensis	Phycisphaera mikurensis NBRC 102666	Water	Extracellular	2623	22	404	1	27	2	231	29
P_phaeoclathatiforme	Pelodictyon phaeoclathratiforme BU-1	Water	Extracellular	744	3	24	2	3	0	94	1
P_staleyi	Pirellula staleyi DSM 6068	Water	Extracellular	1455	23	192	2	31	0	145	7
R_baltica	Rhodopirellula baltica SH 1	Water	Extracellular	2126	23	222	1	14	2	302	11
S_acidiphila	Singulisphaera acidiphila DSM 18658	Water	Extracellular	2841	135	374	0	56	0	389	24
S_africana	Spirochaeta africana DSM 8902	Water	Extracellular	2119	1	210	3	44	4	90	52
S_pleomorpha	Sphaerochaeta pleomorpha str. Grapes	Water	Extracellular	5457	6	242	2	55	1	53	97

Table S4 : Partners for genes transfers (quantity)

HGT partners quantity																			
						Proteobacteria partners						Eukaryota partners						Archaea	Viruses
Synergistetes	Firmicutes	Fusobacteria	Proteobacteria	Tenericutes	Thermotogae	Alpha	Gamma	Delta	Epsilon	Beta	Eukaryota	Alveolata	Euglenozoa	Fungi	Metazoa	Stramenopiles	Viridiplantae	Archaea	Viruses
2	219	3	716	2	0	177	348	87	22	80	264	0	29	26	26	22	137	21	0
0	228	0	629	5	0	130	318	84	6	90	49	2	0	22	3	1	19	36	3
0	242	7	883	2	0	149	519	65	13	133	77	0	2	33	20	3	18	34	1
0	102	0	292	0	0	36	208	13	3	31	97	5	1	1	21	12	49	41	5
4	1013	6	3004	4	0	1020	1093	250	17	622	37	0	0	22	3	0	7	79	14
2	3385	54	1006	33	2	249	329	235	52	128	3	0	0	0	0	0	1	103	9
0	206	10	1604	2	1	484	586	183	3	345	39	1	0	12	23	1	1	41	4
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	6	0	13	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0	0
2	465	5	2057	12	0	353	771	380	19	529	10	0	1	2	2	0	3	122	0
2	505	11	2214	3	0	562	623	424	18	579	138	1	0	25	73	4	24	70	9
1	415	1	1634	6	0	478	516	339	7	288	0	0	0	0	0	0	0	121	20
2	329	0	1416	15	3	316	460	422	3	212	0	0	0	0	0	0	0	182	0
1	195	3	837	2	0	178	452	64	14	122	48	0	0	21	21	2	3	47	1
4	561	6	2254	2	0	317	1343	308	11	262	17	0	0	5	4	0	4	33	1
0	620	40	1974	9	0	275	885	317	92	352	151	3	12	41	30	5	33	97	6
1	527	11	1295	5	2	215	588	217	53	193	152	7	10	16	67	7	33	48	39
3	449	9	1827	5	0	268	795	310	37	370	108	2	2	34	27	3	13	105	13
2	1083	37	1029	13	7	156	519	223	23	105	52	0	13	3	23	2	0	73	8
3	789	27	1074	17	1	184	476	219	59	125	58	0	5	19	3	10	10	28	2
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	2670	150	1079	23	8	54	461	101	292	167	81	0	0	10	29	0	3	74	14
3	1716	34	1775	12	3	203	971	272	76	237	68	1	16	27	11	3	1	138	11
3	528	6	514	0	1	71	313	55	28	44	82	0	2	23	37	2	0	25	11
0	316	64	482	3	2	55	239	27	73	86	17	0	0	0	8	2	0	8	8
0	145	4	125	2	0	11	81	6	23	4	8	1	0	2	2	0	1	3	0
2	225	27	136	1	0	32	77	6	12	9	11	0	0	0	0	1	10	7	0
0	473	26	825	7	2	96	528	122	21	53	58	1	0	9	22	3	19	29	4
0	131	15	143	1	1	21	82	18	9	13	1	0	0	1	0	0	0	4	0
0	175	26	414	6	0	32	247	18	51	66	1	0	0	0	1	0	0	15	1
2	327	28	755	0	3	93	467	40	21	134	39	0	0	5	22	1	10	45	35
4	2032	55	399	6	5	56	145	145	6	42	35	0	0	2	10	1	21	56	0
2	2535	227	298	28	3	22	182	42	14	35	5	0	0	1	1	0	1	49	3
2	314	10	309	1	2	33	172	44	17	41	13	0	0	3	8	0	1	10	1
0	184	16	82	23	1	7	73	1	0	0	0	0	0	0	0	0	0	13	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	49	0	0	4	44	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	8	0	0	5	1	1	0	0	0	0	0	0	0	0	0	2	0
0	21	0	11	0	0	3	0	6	1	0	1	0	0	0	0	0	1	0	0
0	15	0	4	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0
0	30	0	3	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0
0	36	0	15	0	0	3	12	0	0	0	5	0	0	0	1	0	4	0	0
1	551	12	2053	13	0	434	908	342	16	341	0	0	0	0	0	0	0	126	19
4	256	4	1656	1	0	213	1080	189	9	165	60	0	0	6	16	2	34	19	6
0	30	0	309	0	0	48	121	63	0	61	10	4	0	1	5	0	0	19	0
0	53	1	345	0	1	70	131	94	2	42	22	0	0	0	5	0	16	7	0
0	58	10	502	1	0	85	208	107	7	79	4	0	0	0	3	0	0	36	6
0	29	0	337	2	0	37	122	60	7	100	9	0	0	0	0	0	9	18	0
3	173	14	664	2	0	87	253	206	26	77	22	0	0	5	5	4	4	99	2
8	1443	11	2794	33	2	268	869	1172	75	366	0	0	0	0	0	0	0	470	1
2	598	3	2195	3	0	164	1466	299	16	235	13	0	0	4	2	1	1	408	9
0	15	0	292	1	0	77	114	61	1	29	11	0	0	0	2	0	9	12	0
2	215	0	804	2	2	216	383	84	17	103	0	0	0	0	0	0	0	17	3
0	23	0	382	0	0	118	179	31	1	53	0	0	0	0	0	0	0	10	0
0	501	0	1928	7	0	564	807	259	8	284	0	0	0	0	0	0	0	49	1
0	345	4	1424	4	0	443	628	195	2	149	0	0	0	0	0	0	0	67	0
0	34	0	530	4	0	56	223	97	6	129	7	0	0	0	7	0	0	12	1
3	222	14	730	4	0	216	268	118	2	124	0	0	0	0	0	0	0	49	2
2	188	2	1286	0	3	405	595	131	5	139	0	0	0	0	0	0	0	58	0
1	240	3	1476	5	0	478	442	242	1	312	0	0	0	0	0	0	0	103	6
2	653	22	894	5	1	149	382	252	2	89	12	0	0	1	2	1	8	92	2
4	3626	87	996	44	12	271	473	127	7	103	11	0	0	0	10	0	1	103	0

	Bacterial partners									
	Bacteria	Acidobacteria	Actinobacteria	Aquificae	Chloroflexi	Chrysiogenetes	Cyanobacteria	Deinococcus	Firmicutes	Fusobacteria
Candidatus Protochlamydia amoebophila	82.8	0.3	3.1	0.0	1.0	0.0	9.1	1.4	13.2	0.2
Parachlamydia acanthamoebae UV-7	92.5	0.7	2.8	0.3	0.3	0.0	6.2	0.6	19.5	0.0
Simkania negevensis Z	93.2	0.2	8.7	0.4	0.9	0.0	5.1	0.7	14.8	0.4
Waddlia chondrophila WSU 86-1044	78.8	0.3	8.0	0.1	0.1	0.0	3.3	2.5	15.1	0.0
Opitutaceae bacterium TAV1	97.5	0.9	10.6	0.2	1.2	0.0	3.2	0.5	19.2	0.1
Treponema primitia ZAS-2	97.8	0.2	4.2	0.1	0.2	0.0	2.8	0.2	66.1	1.1
Verrucomicrobium spinosum DSM 4136	96.5	1.5	4.8	0.2	0.9	0.0	8.6	0.7	8.5	0.4
Blattabacterium sp (Blaberus giganteus)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Chthoniobacter flavus Ellin428	95.7	2.5	7.5	0.1	1.1	0.0	9.4	0.5	11.5	0.1
Opitutus terrae PB90-1	96.6	9.3	6.1	0.1	1.0	0.0	7.4	0.7	12.1	0.1
Pedosphaera parvula Ellin514	95.7	3.7	12.5	0.2	1.5	0.0	16.9	1.0	9.9	0.2
Gemmata obscuriglobus UQM 2246	95.9	1.0	11.8	0.1	3.0	0.1	13.7	0.6	12.2	0.0
Isosphaera pallida ATCC 43644	93.9	2.4	9.1	0.4	3.2	0.0	12.0	1.0	11.1	0.0
Methylacidiphilum infernorum V4	93.6	1.9	4.5	2.5	1.3	0.0	4.2	4.9	13.0	0.2
Verrucomicrobiae bacterium DG1235	98.6	2.5	6.0	0.0	0.8	0.0	5.9	0.7	15.5	0.2
Leptospira biflexa serovar Patoc	93.2	1.1	8.2	0.3	1.3	0.1	6.4	0.4	16.5	1.1
Leptospira interrogans serovar Lai str. 56601	91.4	0.7	8.3	0.2	0.8	0.1	9.9	0.1	18.9	0.4
Turneriella parva DSM 21527	93.8	0.3	16.3	0.3	1.2	0.0	6.2	0.5	12.4	0.2
Alistipes finegoldii DSM 17242	95.6	0.5	7.4	0.2	0.6	0.1	3.8	1.0	35.5	1.2
Akkermansia muciniphila ATCC BAA-835	96.4	1.0	2.4	0.2	1.1	0.0	6.1	2.1	32.5	1.1
Borrelia burgdorferi JD1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Borrelia dutonii BgVir	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Brachyspira intermedia PWS/A	96.2	0.2	1.7	0.2	0.0	0.0	1.8	0.0	59.9	3.4
Bacteroides vulgatus ATCC 8482	95.4	1.5	3.8	0.1	0.2	0.1	2.6	0.2	36.7	0.7
Bacteroides xylanisolvens XB1A	91.9	2.3	5.7	0.0	0.1	0.0	3.0	0.0	36.4	0.4
Capnocytophaga ochracea DSM 7271	97.2	0.9	11.7	0.3	0.0	0.0	3.5	1.4	27.3	5.5
Porphyromonas asaccharolytica DSM 20707	96.6	0.3	0.9	0.0	0.0	0.0	1.8	0.0	44.5	1.2
Prevotella dentalis DSM 3688	96.3	1.0	6.3	0.0	1.0	0.2	2.2	0.0	45.6	5.5
Parabacteroides distasonis ATCC 8503	94.7	1.0	6.1	0.0	0.6	0.1	1.4	0.1	27.4	1.5
Porphyromonas gingivalis W83	98.7	0.0	3.4	0.0	0.3	0.0	2.1	0.0	34.7	4.0
Prevotella intermedia 17	97.8	0.3	2.1	0.4	0.6	0.0	6.8	1.7	22.6	3.4
Riemerella anatipestifer ATCC 11845 = DSM 15868	91.9	1.3	4.4	0.3	1.1	0.0	3.3	0.1	22.3	1.9
Treponema brennaborense DSM 12168	96.8	0.0	4.6	0.0	0.5	0.0	0.7	0.2	71.5	1.9
Treponema denticola ATCC 35405	98.3	0.0	3.4	0.1	0.2	0.0	1.7	0.0	74.1	6.6
Tannerella forsythia ATCC 43037	97.3	0.7	5.8	0.0	0.7	0.0	6.7	1.0	35.7	1.1
Treponema pallidum subsp. pallidum DAL-1	96.2	0.6	1.2	0.0	0.3	0.0	0.0	2.3	54.0	4.7
Chlamydophila abortus S26/3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Chlamydophila caviae GPIC	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Chlamydophila felis Fe/C-56	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Chlamydia muridarum Nigg	96.7	0.0	3.3	0.0	8.3	0.0	11.7	0.0	0.0	0.0
Chlamydophila pecorum E58	98.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	38.2	0.0
Chlamydophila pneumoniae J138	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	28.8	0.0
Chlamydia psittaci NJ1	100.0	0.0	4.2	0.0	0.0	0.0	0.0	0.0	62.5	0.0
Chlamydia trachomatis 434/Bu	94.3	0.0	0.0	0.0	0.0	0.0	1.1	0.0	41.4	0.0
Blastopirellula marina DSM 3645	96.3	2.4	9.0	0.2	2.1	0.0	9.6	0.9	13.9	0.3
Coraliomargarita akajimensis DSM 45221	96.4	0.7	3.0	0.3	0.6	0.0	7.6	0.1	10.7	0.2
Chlorobium limicola DSM 245	94.3	1.2	0.8	0.0	2.8	0.0	17.8	0.0	5.9	0.0
Chlorobaculum parvum NCIB 8327	94.9	0.9	2.5	1.9	1.4	0.0	10.0	0.0	9.3	0.2
Chlorobium phaeobacteroides BS1	94.5	0.2	3.6	2.5	1.7	0.0	8.9	1.0	6.9	1.2
Chlorobium tepidum TLS	94.5	1.0	3.3	0.6	1.2	0.0	5.7	0.2	5.9	0.0
Chloroherpeton thalassium ATCC 35110	92.3	2.9	5.0	1.2	7.8	0.1	8.4	1.9	10.8	0.9
Candidatus Kuenenia stuttgartiensis	92.8	1.0	2.1	1.4	1.4	0.0	9.0	0.6	22.0	0.2
Lentisphaera araneosa HTCC2155	88.9	0.2	2.1	0.4	0.4	0.1	5.4	0.3	15.5	0.1
Prosthecochloris aestuarii DSM 271	94.4	0.0	1.7	0.2	2.7	0.0	9.3	0.2	3.7	0.0
Planctomyces brasiliensis DSM 5305	98.6	1.4	5.7	0.1	0.8	0.0	11.1	1.1	15.4	0.0
Planctomyces limnophilus DSM 3776	98.1	0.6	2.8	0.4	0.4	0.0	12.2	0.4	4.3	0.0
Planctomyces maris DSM 8797	98.5	1.3	8.2	0.1	1.9	0.0	8.7	0.6	14.7	0.0
Phycisphaera mikurensis NBRC 102666	97.5	0.8	15.0	0.0	1.0	0.1	8.6	1.1	12.8	0.1
Pelodictyon phaeoclathratiforme BU-1	97.4	0.4	3.1	0.3	0.4	0.0	12.3	0.1	4.5	0.0
Pirellula staleyi DSM 6068	96.6	1.5	12.7	0.1	2.1	0.0	9.6	0.5	14.7	0.9
Rhodopirellula baltica SH 1	97.3	1.1	10.2	0.0	0.6	0.1	13.8	0.5	8.6	0.1
Singulisphaera acidiphila DSM 18658	96.3	4.6	12.7	0.0	1.9	0.0	13.2	0.8	8.1	0.1
Spirochaeta africana DSM 8902	95.2	0.0	9.4	0.1	2.0	0.2	4.0	2.3	29.3	1.0
Sphaerochaeta pleomorpha str. Grapes	98.0	0.1	4.3	0.0	1.0	0.0	1.0	1.7	65.1	1.6

Table S4 : Partners for genes transfers (percent)

HGT partners percentages																	
				Proteobacteria partners					Eukaryota partners							Archaea	Viruses
Proteobacteria	Synergistetes	Tenericutes	Thermotogae	Alpha	Gamma	Delta	Epsilon	Beta	Eukaryota	Alveolata	Euglenozoa	Fungi	Metazoa	Stramenopiles	Viridiplantae	Archaea	Viruses
43.2	0.1	0.0	0.0	10.7	21.0	5.3	1.3	4.8	15.9	0.0	1.8	1.6	1.6	1.3	8.3	1.3	0.0
53.8	0.4	0.0	0.5	11.1	27.2	7.2	0.5	7.7	4.2	0.2	0.0	1.9	0.3	0.1	1.6	3.1	0.3
53.9	0.1	0.0	0.0	9.1	31.7	4.0	0.8	8.1	4.7	0.0	0.1	2.0	1.2	0.2	1.1	2.1	0.1
43.3	0.0	0.0	0.4	5.3	30.9	1.9	0.4	4.6	14.4	0.7	0.1	0.1	3.1	1.8	7.3	6.1	0.7
57.0	0.1	0.0	0.1	19.4	20.7	4.7	0.3	11.8	0.7	0.0	0.0	0.4	0.1	0.0	0.1	1.5	0.3
19.6	0.6	0.0	0.4	4.9	6.4	4.6	1.0	2.5	0.1	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.2
66.2	0.1	0.0	0.2	20.0	24.2	7.6	0.1	14.2	1.6	0.0	0.0	0.5	0.9	0.0	0.0	1.7	0.2
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
56.1	0.2	0.0	0.1	13.6	20.4	8.2	0.3	13.4	1.3	0.0	0.0	0.2	0.7	0.0	0.2	0.0	0.0
53.5	0.3	0.0	0.1	9.2	20.1	9.9	0.5	13.8	0.3	0.0	0.0	0.1	0.1	0.0	0.1	3.2	0.0
43.5	0.1	0.0	0.2	11.0	12.2	8.3	0.4	11.4	2.7	0.0	0.0	0.5	1.4	0.1	0.5	1.4	0.2
47.9	0.2	0.0	0.0	14.0	15.1	9.9	0.2	8.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.5	0.6
47.7	0.5	0.1	0.7	10.7	15.5	14.2	0.1	7.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.1	0.0
55.6	0.1	0.0	0.4	11.8	30.0	4.3	0.9	8.1	3.2	0.0	0.0	1.4	1.4	0.1	0.2	3.1	0.1
62.3	0.1	0.0	0.1	8.8	37.1	8.5	0.3	7.2	0.5	0.0	0.0	0.1	0.1	0.0	0.1	0.9	0.0
52.5	0.2	0.0	0.1	7.3	23.5	8.4	2.4	9.4	4.0	0.1	0.3	1.1	0.8	0.1	0.9	2.6	0.2
46.5	0.2	0.1	0.8	7.7	21.1	7.8	1.9	6.9	5.5	0.3	0.4	0.6	2.4	0.3	1.2	1.7	1.4
50.4	0.1	0.0	0.2	7.4	21.9	8.5	1.0	10.2	3.0	0.1	0.1	0.9	0.7	0.1	0.4	2.9	0.4
33.7	0.4	0.2	1.1	5.1	17.0	7.3	0.8	3.4	1.7	0.0	0.4	0.1	0.8	0.1	0.0	2.4	0.3
44.3	0.7	0.0	0.7	7.6	19.6	9.0	2.4	5.2	2.4	0.0	0.2	0.8	0.1	0.4	0.4	1.2	0.1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
24.2	0.5	0.2	0.6	1.2	10.3	2.3	6.6	3.7	1.8	0.0	0.0	0.2	0.7	0.0	0.1	1.7	0.3
38.0	0.3	0.1	0.2	4.3	20.8	5.8	1.6	5.1	1.5	0.0	0.3	0.6	0.2	0.1	0.0	3.0	0.2
35.5	0.0	0.1	0.9	4.9	21.6	3.8	1.9	3.0	5.7	0.0	0.1	1.6	2.6	0.1	0.0	1.7	0.8
41.6	0.3	0.2	0.0	4.7	20.6	2.3	6.3	7.4	1.5	0.0	0.0	0.0	0.7	0.2	0.0	0.7	0.7
38.3	0.6	0.0	2.1	3.4	24.8	1.8	7.1	1.2	2.5	0.3	0.0	0.6	0.6	0.0	0.3	0.9	0.0
27.6	0.2	0.0	0.0	6.5	15.6	1.2	2.4	1.8	2.2	0.0	0.0	0.0	0.0	0.2	2.0	1.4	0.0
47.8	0.4	0.1	0.5	5.6	30.6	7.1	1.2	3.1	3.4	0.1	0.0	0.5	1.3	0.2	1.1	1.7	0.2
37.8	0.3	0.3	6.1	5.6	21.7	4.8	2.4	3.4	0.3	0.0	0.0	0.3	0.0	0.0	0.0	1.1	0.0
53.4	0.8	0.0	0.0	4.1	31.8	2.3	6.6	8.5	0.1	0.0	0.0	0.0	0.1	0.0	0.0	1.9	0.1
51.5	0.0	0.2	0.0	6.3	31.9	2.7	1.4	9.1	2.7	0.0	0.0	0.3	1.5	0.1	0.7	3.1	2.4
14.0	0.2	0.2	0.2	2.0	5.1	5.1	0.2	1.5	1.2	0.0	0.0	0.1	0.4	0.0	0.7	2.0	0.0
8.7	0.8	0.1	0.6	0.6	5.3	1.2	0.4	1.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.4	0.1
35.1	0.1	0.2	0.5	3.8	19.5	5.0	1.9	4.7	1.5	0.0	0.0	0.3	0.9	0.0	0.1	1.1	0.1
24.0	6.7	0.3	0.3	2.1	21.4	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.8	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
71.0	0.0	0.0	0.0	5.8	63.8	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13.3	0.0	0.0	0.0	8.3	1.7	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.3	0.0
20.0	0.0	0.0	0.0	5.5	0.0	10.9	1.8	0.0	1.8	0.0	0.0	0.0	0.0	0.0	1.8	0.0	0.0
7.7	0.0	0.0	0.0	0.0	5.8	1.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6.3	0.0	0.0	0.0	2.1	2.1	0.0	0.0	2.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
17.2	0.0	0.0	0.0	3.4	13.8	0.0	0.0	0.0	5.7	0.0	0.0	0.0	1.1	0.0	4.6	0.0	0.0
51.8	0.3	0.0	0.4	10.9	22.9	8.6	0.4	8.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.2	0.5
69.2	0.0	0.0	0.0	8.9	45.1	7.9	0.4	6.9	2.5	0.0	0.0	0.3	0.7	0.1	1.4	0.8	0.3
61.1	0.0	0.0	0.2	9.5	23.9	12.5	0.0	12.1	2.0	0.8	0.0	0.2	1.0	0.0	0.0	3.8	0.0
60.5	0.0	0.2	2.3	12.3	23.0	16.5	0.4	7.4	3.9	0.0	0.0	0.0	0.9	0.0	2.8	1.2	0.0
59.7	0.1	0.0	0.0	10.1	24.7	12.7	0.8	9.4	0.5	0.0	0.0	0.0	0.4	0.0	0.0	4.3	0.7
68.6	0.4	0.0	0.2	7.5	24.8	12.2	1.4	20.4	1.8	0.0	0.0	0.0	0.0	0.0	1.8	3.7	0.0
41.5	0.1	0.0	0.6	5.4	15.8	12.9	1.6	4.8	1.4	0.0	0.0	0.3	0.3	0.3	0.3	6.2	0.1
42.6	0.5	0.0	0.5	4.1	13.3	17.9	1.1	5.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.2	0.0
56.8	0.1	0.0	0.1	4.2	38.0	7.7	0.4	6.1	0.3	0.0	0.0	0.1	0.1	0.0	0.0	10.6	0.2
71.4	0.2	0.0	0.2	18.8	27.9	14.9	0.2	7.1	2.7	0.0	0.0	0.0	0.5	0.0	2.2	2.9	0.0
57.5	0.1	0.1	0.2	15.4	27.4	6.0	1.2	7.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2	0.2
71.9	0.0	0.0	0.0	22.2	33.7	5.8	0.2	10.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.9	0.0
56.8	0.2	0.0	0.1	16.6	23.8	7.6	0.2	8.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.4	0.0
52.9	0.1	0.0	0.1	16.5	23.3	7.2	0.1	5.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.5	0.0
69.4	0.5	0.0	0.0	7.3	29.2	12.7	0.8	16.9	0.9	0.0	0.0	0.0	0.9	0.0	0.0	1.6	0.1
48.5	0.3	0.0	0.1	14.3	17.8	7.8	0.1	8.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.3	0.1
58.9	0.0	0.1	0.2	18.5	27.2	6.0	0.2	6.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.7	0.0
50.0	0.2	0.0	0.2	16.2	15.0	8.2	0.0	10.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.5	0.2
40.2	0.2	0.0	2.7	6.7	17.2	11.3	0.1	4.0	0.5	0.0	0.0	0.0	0.1	0.0	0.4	4.1	0.1
17.9	0.8	0.2	1.5	4.9	8.5	2.3	0.1	1.8	0.2	0.0	0.0	0.0	0.2	0.0	0.0	1.8	0.0

Conclusion

Les génomes bactériens évoluent via différents événements génétiques : mutations, pertes et acquisitions de gènes (duplications, gain de novo et transferts horizontaux). Au cours de cette thèse nous avons pu voir les relations étroites entre ces événements, l'évolution des bactéries, leurs structures cellulaires et leurs environnements.

Nous avons constaté que la présence d'une hypothétique membrane intracellulaire chez certaines bactéries du super-phylum PVC ne semblait pas être impliquée dans la protection des génomes bactériens contre les transferts de gènes horizontaux. La fonction de cette structure ne peut donc pour l'instant être identifiée, d'autres hypothèses peuvent cependant être formulées, comme par exemple un rôle dans un possible épissage qui pourrait être étudiée grâce aux données ARN. Les analyses microscopiques effectuées sur les bactéries de structure cellulaire inconnue ont permis de reconstituer, en partie, l'évolution de cette caractéristique. A contrario l'environnement des

bactéries ou leur mode de vie semblent influencer l'évolution de la composition des génomes, d'abord au niveau global mais surtout au niveau de la sélection des gènes obtenus par transferts horizontaux.

Nos résultats pâtissent de plusieurs manques : en premier lieu, il manque encore un certain nombre de génomes complets (en particulier ceux des bactéries possiblement compartimentées) qui permettraient de mieux appréhender l'évolution des différentes caractéristiques de ces bactéries ; deuxièmement, la difficulté de détecter de façon précise les transferts de gènes ou plus généralement de différencier les multiples événements génétiques : troisièmement, l'absence d'ontologie concernant les différents environnement et modes de vies qui sont étudiés. Des solutions doivent donc être apportées pour améliorer les études menées sur le super-phylum PVC ; l'aboutissement de projets de séquençage sera déjà une première étape, qui devra être associée à une amélioration des méthodes de détection génétique et à une réflexion sur l'organisation et la nature des environnements. Des

études complémentaires pourraient aussi être menées via l'étude des transcriptomes et des interactomes en rapport avec les différents phénotypes observés. L'intégration de ces données en un seul modèle synthétique nous permettrait de connaitre plus précisément les mécanismes évolutifs des bactéries PVC et de prédire en plus les fonctions et les interactions des gènes impliqués (21).

Réfèrences

1.Kielak, A., Rodrigues, J. L., Kuramae, E. E., Chain, P. S., Van Veen, J. A., & Kowalchuk, G. A. (2010). Phylogenetic and metagenomic analysis of Verrucomicrobia in former agricultural grassland soil. FEMS microbiology ecology, 71(1), 23-33.

2.Chouari, R., Le Paslier, D., Dauga, C., Daegelen, P., Weissenbach, J., & Sghir, A. (2005). Novel major bacterial candidate division within a municipal anaerobic sludge digester. Applied and environmental microbiology, 71(4), 2145-2153.

3.Shu, Q & Jiao N. (2008). Different Planctomycetes diversity patterns in latitudinal surface seawater of the open sea and in sediment. The journal of Microbiology. 10, 154-159

4.Chiapello, H., Bourgait, I., Sourivong, F., Heuclin, G., Gendrault-Jacquemard, A., Petit, M. A., & El Karoui, M. (2005). Systematic determination of the mosaic structure of bacterial genomes: species backbone versus strain-specific loops. BMC bioinformatics, 6(1), 171.

5.Van Passel, M. W., Kant, R., Palva, A., Copeland, A., Lucas, S., Lapidus, A., ... & Smidt, H. (2011) . Genome sequence of the Verrucomicrobium Opitutus terrae

PB90-1, an abundant inhabitant of rice paddy soil ecosystems. Journal of bacteriology.

6.Derrien, M., Collado, M. C., Ben-Amor, K., Salminen, S., & de Vos, W. M. (2008). The mucin degrader Akkermansia muciniphila is an abundant resident of the human intestinal tract. Applied and environmental microbiology 74(5), 1646-1648.

7.Hou, S., Makarova, K. S., Saw, J. H., Senin, P., Ly, B. V., Zhou, Z., ... & Alam, M. (2008). Complete genome sequence of the extremely acidophilic methanotroph isolate V4, Methylacidiphilum infernorum, a representative of the bacterial phylum Verrucomicrobia. Biol Direct 3(11).

8.Thrash, J. C., Cho, J. C., Vergin, K. L., Morris, R. M., & Giovannoni, S. J. (2010). Genome sequence of Lentisphaera araneosa HTCC2155T, the type species of the order Lentisphaerales in the phylum Lentisphaerae. Journal of bacteriology 192(11), 2938-2939

9.Van Niftrik, L., Van Helden, M., Kirchen, S., Van Donselaar, E. G., Harhangi, H. R., Webb, R. I., ... & Strous, M. (2010). Intracellular localization of membrane-bound ATPases in the compartmentalized anammox bacterium 'Candidatus Kuenenia stuttgartiensis'. Molecular microbiology 77(3), 701-715.

10.Budd, A., & Devos, D. P. (2012). Evaluating the evolutionary origins of unexpected character distributions within the bacterial Planctomycetes-Verrucomicrobia-Chlamydiae superphylum. Frontiers in microbiology, 3.

11.Wagner, M., & Horn, M. (2006). The Planctomycetes, Verrucomicrobia, Chlamydiae and sister phyla comprise a superphylum with biotechnological and medical relevance. Current opinion in biotechnology, 17(3), 241-249.

12.Lee, K. C., Webb, R. I., Janssen, P. H., Sangwan, P., Romeo, T., Staley, J. T., & Fuerst, J. A. (2009). Phylum Verrucomicrobia representatives share a compartmentalized cell plan with members of bacterial phylum Planctomycetes. BMC microbiology, 9(1), 5.

13.Lindsay, M. R., Webb, R. I., Strous, M., Jetten, M. S., Butler, M. K., Forde, R. J., & Fuerst, J. A. (2001). Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. Archives of Microbiology, 175(6), 413-429.

14.Lindsay, M. R., Webb, R. I., & Fuerst, J. A. (1997). Pirellulosomes: a new type of membrane-bounded cell compartment in planctomycete bacteria of the genus Pirellula. Microbiology, 143(3), 739-748.

15.Fieseler, L., Horn, M., Wagner, M., & Hentschel, U. (2004). Discovery of the novel candidate phylum “Poribacteria” in marine sponges. Applied and environmental microbiology, 70(6), 3724-3732

16.Merhej, V., Royer-Carenzi, M., Pontarotti, P., & Raoult, D. (2009). Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biology direct 4(1), 13.

17.Moliner, C., Fournier, P. E., & Raoult, D. (2010). Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. FEMS microbiology reviews 34(3), 281-294

18.Fuerst, J. A., & Webb, R. I. (1991). Membrane-bounded nucleoid in the eubacterium Gemmatata obscuriglobus. Proceedings of the National Academy of Sciences, 88(18), 8184-8188.

19.Devos, D. P. (2014). PVC bacteria: variation of, but not exception to, the Gram-negative cell plan. Trends in microbiology, 22(1), 14-20

20.Speth, D. R., van Teeseling, M. C., & Jetten, M. S. (2012). Genomic analysis indicates the presence of an asymmetric bilayer outer membrane in Planctomycetes and Verrucomicrobia. Frontiers in microbiology, 3.

21.Medina, M. (2005). Genomes, phylogeny, and evolutionary systems biology.

Proc Natl Acad Sci USA. 102 Suppl 1, 6630-6635

Remerciements

Je tiens à remercier tous les gens qui ont, de prés ou de loin, soutenu mon travail de thèse : mon directeur de thèse, le professeur Raoult et mon co-directeur de thèse, le docteur Pierre Pontarotti pour son aide, ses conseils et toutes nos discussions passionnantes, ainsi que Vicky Merhej pour toute l'aide qu'elle m'a apportée. Je remercie aussi les membres du jury qui ont accepté d'évaluer mon travail de thèse, le professeur Greub et le docteur Simonet, tous les membres de l'EBM et de l'URMITE qui ont été présents durant ces trois années et mes collègues d'Hippo'Thése, de la CJC et de DynAMU. Un grand merci aussi à ma famille pour avoir supporté mes humeurs durant ces trois ans (particulièrement mon père, mes frère et Nicole) mes amis pour avoir été là dans les moments les plus difficiles (un merci spécial à Zachia pour son soutient infaillible). Enfin je remercie aussi les gens qui ne sont plus là aujourd'hui mais qui ont toujours cru en moi, spécialement ma mère sans qui je ne serais jamais arrivée jusqu'ici.

Résumé

La compréhension de l'évolution des génomes est un des enjeux clé de la biologie actuelle. Cela induit l'identification des événements génétiques affectant les génomes au cours du temps, mais aussi la détermination des relations entre les génomes, les phénotypes et l'environnement des organismes.
Nous avons rédigé une revue de la littérature consacrée à la contribution de la génomique dans la compréhension de la diversité, de l'évolution et des causes génétiques de phénotypes d'un super-phylum bactérien appelé super-phylum PVC (pour Planctomycetes, Verrucomicrobiae et Chlamydiae). Ces bactéries proviennent d'environnements variés et présentent des caractéristiques phénotypiques particulières, dont certaines ne sont retrouvées que dans ce groupe. Les analyses génomiques ont révélé la grande diversité de ces espèces, mais ont aussi permis de reconstruire l'évolution de leurs génomes et dans certains cas, d'expliquer l'apparition de phénotypes particuliers. Plusieurs caractéristiques morphologiques ou physiologiques restent cependant mal expliquées, que ce soit leur origine ou leur évolution, entre autre à cause du manque de données disponibles. Une partie de notre travail était consacré à l'étude de l'évolution et de l'impact de la présence d'un plan cellulaire particulier chez les bactéries PVC. Ce plan cellulaire est sujet a différentes interprétations et induirait la compartimentation des cellules en deux régions distinctes dont l'une contiendrait le nucléoside. Les résultats obtenus semblent indiquer que cette caractéristique n'induit pas une protection des génomes bactériens vis à vis des transferts de gènes horizontaux, comme on pourrait le supposer, sa fonction reste donc pour l'instant indéfinie. En revanche les observations microscopiques réalisées sur deux espèces ont permis de mieux appréhender l'évolution de ce plan cellulaire. Nous avons, de plus, détecté une contribution de l'environnement concernant la sélection des gènes transférés. Il semblerait que les gènes transférés soient en effet sélectionnés selon leurs fonctions par les différents environnements.
Nos travaux ont donc permis d'améliorer la compréhension des relations entre l'évolution, les phénotypes et l'environnement, en particulier chez les bactéries du super-phylum PVC, fournissant ainsi de nombreuses pistes de travail qui pourront être approfondies dans le futur.

Abstract

The comprehension of genomes evolution is a key issue of modern biology. This induces the identification of genetic events occurred in history of genomes and also the determination of relations between the genomes, phenotypes and environments. We wrote a review dedicated to the genomic contribution in comprehension of diversity, evolution and genetic causes of phenotypic features, in a bacterial super-phylum named PVC (for Planctomycetes, Verrucomicrobiae and Chlamydiae). These bacteria are distributed in varied environments and present specific phenotypic characteristics, whom some of them are identified only in this group. Genomic analyzes revealed the important diversity of these species and allow also to reconstruct the genomes evolution and, in some cases, to explain the presence of specific phenotypes. Due to the lack of information, it is difficult to define the origins and evolution of all specific phenotypes. One part of our work was dedicated to the study of evolution and impact of one of this phenotype, the special cell plan detected in PVC bacteria. This original cell plan is subject to different interpretations and induces the compartmentalization of cells in two different regions, whom one containing the nucleoid. Our results indicate that this feature has probably no role in the protection of bacterial genomes against horizontal genes transfers, so, its function is still unknown. Microscopic observations of two species from PVC super-phylum permit to better understand the evolution of the special cell plan. The environment seems to contribute in the genomes evolution, by selection of genes transferred. Genes transferred are probably selected according to their functions by the different environments.
Our works allowed to improve the knowledge about relations between evolution, genomes, phenotypes and environment, especially in bacteria from PVC super-phylum, providing new information for future works.