

Université du Québec
Institut National de la Recherche Scientifique
Centre Institut Armand-Frappier

Suivi de la succession des communautés microbiennes et de l'oxydation du CO d'un sol forestier afin d'identifier des bactéries carboxydotrofes inconnues

Isabelle Lalonde
11 janvier 2016

Mémoire présenté pour l'obtention du grade de
Maître ès sciences (M.Sc.)
Microbiologie Appliquée

Jury d'évaluation

Président du jury et
Examinateur interne

Frédéric Veyrier
INRS – Institut Armand-Frappier

Examinateur externe

Julien Tremblay
CNRC-Montréal

Directeur de recherche

Philippe Constant
INRS – Institut Armand-Frappier

RÉSUMÉ

Le monoxyde de carbone (CO) influence la chimie atmosphérique de plusieurs autres gaz dans l'atmosphère. Les radicaux hydroxyles ($\text{OH}\cdot$) et certaines bactéries aérobies du sol sont les deux grands puits atténuant les émissions globales de CO. Le puits biologique étant le moins connu et le plus incertain rend difficile d'évaluer l'impact du changement global sur la distribution et l'activité métabolique des bactéries capturant le CO et ainsi sur la chimie atmosphérique de plusieurs gaz. Quelques bactéries aérobies pouvant consommer le CO atmosphérique, appelées carboxydovores, ont été isolées jusqu'à présent mais ne sont pas suffisantes pour comprendre l'ensemble de ce puits de CO. Pour les étudier, le gène *coxL*, codant pour la CO-déshydrogénase (CODH) permettant la réaction d'oxydation du CO, est utilisé comme point de départ. Une étude précédente a observé un enrichissement de séquences *coxL* atypiques dans un sol de forêt de feuillus. Ces séquences ont été assignées à des *Deltaprotéobacteries*, selon leur seul représentant le plus près, *Haliangium ochraceum*. Cette assignation à de nouvelles espèces de *Deltaprotéobacteries* aérobies ou d'autres espèces inconnues est fort incertaine en raison du transfert latéral des gènes codant pour la CODH. Le présent projet a été réalisé afin de tenter d'identifier les bactéries possédant le gène *coxL* atypique en colonisant le sol de forêt de feuillus stérile avec sa flore indigène. Ainsi, la consommation du CO, la diversité des gènes *coxL* et la diversité des gènes d'ARNr 16S ont pu être suivis en fonction du temps d'incubation du sol. Une méthode innovatrice impliquant le calcul de réseaux de corrélation a été utilisée pour intégrer les données. Les ribotypes démontrant une forte covariation durant la colonisation du sol ont d'abord été groupés en 5 modules différents. La diagonalisation des modules en vecteurs a permis d'identifier un module fortement corrélé à l'activité de consommation de CO du sol. Des analyses de corrélations ont été effectuées entre le vecteur de ce dernier module et les génotypes *coxL* détectés durant l'incubation. La très faible abondance de ribotypes de *Deltaprotéobacteries* retrouvés dans l'étude permet de conclure qu'il est peu probable que le génotype *coxL* atypique provienne de bactéries appartenant à cette classe. Selon la forte corrélation au CO et l'abondance de certains OTU ARNr 16S, les génotypes atypiques *coxL* pourrait provenir de bactéries de la famille des *Microbacteriaceae*, *Actinospicacae* ou *Acetobacteraceae*. Le suivi des séquences *coxL* par qPCR laisse croire que les bactéries possédant le gène *coxL* atypique seraient des stratégistes K. Leur domestication nécessitera des milieux de culture formulés pour défavoriser les bactéries à croissance rapide et la cellulose sera la source de carbone à privilégier.

Mots-clés : CO atmosphérique, bactéries carboxydovores, *coxL*, sol

Étudiante

Directeur de recherche

ABSTRACT

Surveys of the *coxL* gene, encoding the large subunit of the CO-dehydrogenase, are used as a standard approach in carboxydovore bacteria scavenging atmospheric CO ecological studies. Recent soil surveys unveiled that the distribution of *coxL* sequences encompassing the atypical genotype *coxL*-type I(group x) was correlated to the CO oxidation activity. Based on phylogenetic analysis including the available *coxL* reference genome sequences, this unusual genotype was assigned to an unknown *Deltaproteobacteria* with the *coxL* sequence from *Haliangium ochraceum* being the sole and closest reference sequence. Here we seek to challenge the proposed taxonomic assignation of *coxL*(group x) genotype through the monitoring of CO consumption activity and microbial community successions during the colonization of sterile soil microcosms inoculated with indigenous microorganisms. In our study, we established that the estimated population density of *Deltaproteobacteria* was too small to account for the abundance of *coxL*(group x) genotype detected in soil. Furthermore, we computed a correlation network to relate 16S rRNA gene profiles with the succession of *coxL* genotypes and CO uptake activity in soil. We found that most of the *coxL* genotypes for which the colonization profile displayed covariance with CO uptake activity were related to potential carboxydovore bacteria belonging to *Actinobacteria* and *Alphaproteobacteria*. Our analysis did not provide any evidence that *coxL*(group x) genotypes belonged to *Deltaproteobacteria*. Considering the colonization profile of CO-oxidizing bacteria and the theoretical energy yield of measured CO oxidation rates in soil microcosms, we propose that unknown carboxydovore bacteria harboring the atypical *coxL*(group x) genotype are mixotrophic K-strategists.

Keywords: Carbon monoxide, forest soil, carboxydovore bacteria, correlation network.

REMERCIEMENTS

Premièrement, je voudrais remercier mon directeur de recherche, Philippe Constant, pour son aide, son support et pour tout ce qu'il m'a appris au cours de ma maîtrise. Je n'aurais pas pu avoir un directeur aussi compréhensif, présent et à l'écoute de ses étudiants. Merci pour toutes nos longues conservations, qu'elles soient scientifiques ou pas, et surtout merci pour ton positivisme, ta joie de vivre et tes taquineries, qui font de toi une personne tellement agréable à côtoyer!

Ensuite, je voudrais remercier mes collègues et amies Amélie Bergeron et Audrey-Anne Durand. Sans vous, mon expérience à l'Institut Armand-Frappier n'aurait pas été la même. Nous avons eu tellement de plaisir ensemble! En plus, vous m'avez grandement aidé et conseillé quand j'en avais besoin. Vous êtes des amies en or! Ce n'est que le début d'une grande amitié.

Merci à tous les membres de mon laboratoire, Sarah, Quentin, Mondher et Liliana ainsi que les autres membres du laboratoire Guertin, Narin, Jean-Philippe et Claude pour votre aide et pour tous les bons moments. En gros, merci à vous tous et à tous les autres collègues du GRME pour toutes les belles sorties de groupe au cours de ma maîtrise. Je n'oublierai jamais le *surprise party* que vous m'aviez organisé pour célébrer mon anniversaire. La présence et la contribution de tous et chacun à cette belle soirée m'ont énormément touchée. Je garderai un bon souvenir des gens extraordinaire que j'ai rencontré ici à l'Institut.

Enfin, je ne peux terminer sans remercier mes parents qui m'ont supporté tout au long de mes études. Merci de m'avoir écouté parler de science même si ce sujet est si loin de vos domaines respectifs! Merci pour tout.

TABLE DES MATIÈRES

Résumé	i
Abstract.....	iii
Remerciements.....	iv
Table des matières.....	v
Liste des tableaux	viii
Liste des figures.....	ix
Liste des sigles et abréviations	x
Synthèse	1
Chapitre 1 : Introduction.....	2
1.1 Les gaz à effet de serre	2
1.2 Le monoxyde de carbone (CO).....	3
1.2.1 Distribution du CO dans l'atmosphère.....	5
1.2.2 Sources et puits du CO (bilan)	6
1.3 Consommation du CO par le sol	9
1.3.1 Bactéries oxydant le CO.....	12
1.3.2 Carboxydovores et carboxydotrophes.....	12
1.4 La CO-déshydrogénase (CODH)	15
1.4.1 Métabolisme	15
1.4.2 Structure	17
1.4.3 Gènes structuraux et accessoires	20
1.4.4 Gène <i>coxL</i>	22
1.4.5 Diversité du gène <i>coxL</i> dans l'environnement.....	25
1.5 Problématique : Effet du changement global	27

1.51 Biomarqueur	28
1.6 Hypothèse	30
1.7 Objectifs	30
Article	31
Chapitre 2 : Article	32
2.1 Contributions à l'article.....	33
2.2 Résumé	33
2.3 Abstract.....	34
2.4 Introduction	35
2.5 Material and Methods.....	36
2.5.1 Soil sample and incubation	36
2.5.3 CO oxidation rate in soil	37
2.5.4 DNA extraction and qPCR assays	38
2.5.5 <i>coxL</i> genotyping	38
2.5.6 Bacterial 16S rRNA gene profiling.....	39
2.5.7 Statistical analysis	39
2.6 Results	40
2.6.1 Maturation of soil CO uptake activity.....	40
2.6.2 Abundance of bacterial 16S rRNA and <i>coxL</i> gene	41
2.6.3 Bacterial 16S rRNA genotypes.....	43
2.6.4 <i>coxL</i> genotypes	44
2.6.5 Correlation network analysis	46
2.7 Discussion.....	51
2.8 Acknowledgments	54
Conclusion et Perspectives	55
Références	57

Annexe	75
Annexe 1: Matériel supplémentaire (ARTICLE)	76
Annexe 2: Commandes mothur	84
Annexe 3: Commandes R - WGCNA	92
Annexe 3: Commandes R - Homogeneity Multivariate Dispersion	101

LISTE DES TABLEAUX

Tableau 1 : Sources de CO (Tg/année) provenant de différentes analyses.....	8
Tableau 2 : Consommation de CO d'un sol selon son type de végétation.....	11
Tableau 3 : Présences des <i>coxL</i> de forme I et forme II chez différents carboxydotrophes.....	24
Table 1 : Specific richness and evenness of <i>coxL</i> and 16S rRNA gene OTUs in soil microcosms.....	43
Table 2 : Composition of module 1 comprising 16S rRNA OTUs displaying colonization profile related to the maturation of CO oxidation activity in soil.....	48
Table S1 : Two case studies supporting the taxonomic inference of <i>coxL</i> sequences for which the colonization profile was correlated to the maturation of CO uptake activity in soil microcosms.....	82

LISTE DES FIGURES

Figure 1 : Rôles des radicaux hydroxyles (OH) dans les réactions photochimiques atmosphériques.....	4
Figure 2 : Chaîne de transport d'électrons simplifiée des bactéries aérobies insensibles au CO.....	16
Figure 3 : La CO-déshydrogénase.....	18
Figure 4 : Mécanisme d'oxydation du CO de la CO déshydrogénase.....	20
Figure 1 (Article) : Time series of the CO oxidation activity measured in inoculated and sterile soil microcosms.....	41
Figure 2 (Article) : Time series of the relative abundance of (A) <i>coxL</i> and (B) <i>coxL</i> (group x) genotypes in soil microcosms as estimated by qPCR.....	42
Figure 3 (Article) : Maximum-likelihood phylogenetic analysis of <i>coxL</i> -inferred amino acid sequences of 136 OTUs and reference sequences from NCBI (147 residues).....	45
Figure 4 (Article) : Taxonomic composition of the five modules defined by the covariance of 16S rRNA OTUs time series in the soil.....	47
Figure 5 (Article) : Spearman correlations between the <i>coxL</i> OTUs with the eigengene of the five modules comprising 16S rRNA OTUs displaying similar soil colonization profile.....	50
Figure S1 : Abundance of total <i>coxL</i> , <i>coxL</i> (group x) and 16S rRNA genes during the incubation, as determined by qPCR.....	74
Figure S2 : Homogeneity multivariate dispersion of (A) soil microbial community based of 16S rRNA gene profiles and (B) carboxydovore bacteria based on <i>coxL</i> gene profiles.....	78
Figure S3 : Detailed view of the clusters (A) <i>coxL</i> (α -Proteobacteria), (B) <i>coxL</i> (Actinobacteria), and (C) <i>coxL</i> (group x) in the <i>coxL</i> phylogenetic tree presented in the article (Fig. 3).....	79

LISTE DES SIGLES ET ABRÉVIATIONS

GES	Gaz à effet de serre
IR	Infrarouge
CO	Monoxyde de carbone
CH ₄	Méthane
H ₂ O	Eau
CO ₂	Dioxyde de carbone
N ₂ O	Protoxyde d'azote
O ₃	Ozone
SF ₆	Hexafluorure de soufre
CFCs	Chlorofluorocarbures
NOx	Oxyde d'azote
PRG	Potentiel de réchauffement global
OH·	Radical hydroxyle
GMD	Global Monitoring Division
ESRL	Earth System Research Laboratory
NOAA	National Oceanic and Atmospheric Administration
GFED	Global Fire Emissions Database
COVNM	Composé organique volatil non méthanique
GIEC	Groupe d'experts intergouvernementaux sur l'évolution du climat
Tg	Téragramme
nm	Nanomètre
µm	Micromètre
ppmv	Partie par million par volume
ppbv	Partie par milliard par volume

K_m	Constante de Michaelis (affinité d'un enzyme pour un substrat)
CODH	Monoxyde de carbone déshydrogénase
CBB	Calvin-Bassham-Benson
RubisCO	Ribulose-1,5-diphosphate carboxylase/oxygénase
MO	Molybdène
MCD	Cofacteur Molybdoptérine cytosine dinucléotide
a.a.	Acides aminés
kDa	Kilodalton

SYNTHÈSE

CHAPITRE 1 : INTRODUCTION

1.1 Les gaz à effet de serre

La Terre reçoit de la chaleur par l'entremise des rayons visibles et ultraviolets du soleil. Une partie de ces rayons sont immédiatement renvoyés vers l'espace par certains aérosols par exemple, mais la surface de la Terre absorbe approximativement les deux tiers de cette énergie incidente. Par la suite, cette énergie est renvoyée sous forme de rayonnement infrarouge (IR) et certains gaz dans l'atmosphère ont la capacité d'absorber ce rayonnement. Ainsi, ces gaz empêchent en partie la perte de cette énergie et réchauffe la planète. Ces gaz, nommés gaz à effet de serre (GES), exercent donc un forçage radiatif positif puisqu'ils permettent de réchauffer la Terre, comparativement aux aérosols qui exercent plutôt un forçage radiatif négatif. Les GES sont nécessaires afin que la Terre maintienne des températures pouvant soutenir les différentes formes de vie. À la base, l'effet de serre procuré par ces gaz, est un processus tout à fait naturel. La vapeur d'eau (H_2O), le dioxyde de carbone (CO_2), le méthane (CH_4), l'oxyde nitreux (N_2O) et l'ozone (O_3) sont tous des GES. Toutefois, l'activité anthropogénique a provoqué un surplus d'émissions de ces gaz en plus d'en avoir créé de nouveaux pour les biens de l'industrie comme l'hexafluore de soufre (SF_6), les chlorofluorocarbures (CFCs) et autres halocarbones. En conséquence, cela augmente le processus d'effet de serre de la planète, menant à un changement global du climat (Forster *et al.*, 2007).

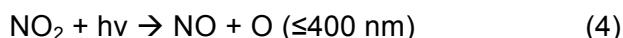
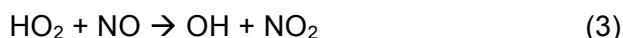
Un indice de potentiel de réchauffement global (PRG) est attribué aux GES selon leur forçage radiatif direct et leur durée de vie dans l'atmosphère. Le PRG est exprimé en watts par mètre carré (W/m^2) et est la plupart du temps divisé par le PRG du dioxyde de carbone (CO_2), le GES le plus abondant après la vapeur d'eau (H_2O). Le CO_2 est plus abondant que le CH_4 et le N_2O , mais le CH_4 a un PRG 24 fois plus grand que le CO_2 sur une période de 100 ans, et près de 300 fois plus grand pour le N_2O (Ramaswamy *et al.*, 2001). Ces gaz pouvant absorber les rayons infrarouges ne sont pas les seuls pouvant contribuer à l'effet de serre. Certains gaz n'absorbent pas les rayons infrarouges, mais influencent directement la chimie des GES. Ces gaz sont considérés comme des GES indirects. Les oxydes d'azotes (NOx), hydrogène moléculaire (H_2) et le monoxyde de carbone (CO) en sont des exemples.

1.2 Le monoxyde de carbone (CO)

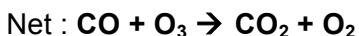
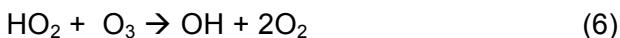
Le CO, est un gaz inodore, incolore et toxique, retrouvé dans la troposphère, les premiers 10-18km de l'atmosphère (Monks, 2005). Le CO est présent à l'état de traces, soit à des concentrations inférieures à 1%. Sa durée de vie se limite de 2-3 mois, comparativement à 8-9 ans pour le CH₄ (Monson *et al.*, 2001). Le CO n'est pas un GES direct car son forçage radiatif est très faible, soit de 0,024 W/m² (Ramanathan *et al.*, 1987). Cependant, le CO contribue significativement au forçage du climat global (Prather, 1996) de manière indirect, avec un forçage radiatif indirect d'environ 0,2 W/m² (Forster *et al.*, 2007), faisant de lui un GES indirect important.

Le CO affecte la chimie des GES de plusieurs façons : par la production de CO₂, la production d'ozone (O₃) avec ses sous-produits de réactions et en limitant le nombre de radicaux hydroxyles disponibles (OH·), le puits principal du CH₄ (Fuglestvedt *et al.*, 1996). En effet, il y a production de CO₂ lors de l'oxydation du CO mais cette source de CO₂ est négligeable compte tenu des faibles quantités produites lors de cette réaction comparativement aux autres sources de CO₂. Le CO affecte aussi la concentration de O₃ en fonction de la concentration de NOx présent lors de son attaque par les OH·. À haute concentration (e.g. milieu urbain), le CO engendre une augmentation importante de O₃ troposphérique (réactions 1-5) tandis qu'à faible concentration de NOx (e.g. milieu rural), le CO engendre une diminution de O₃ (réactions 1, 2, 6) (Crutzen *et al.*, 1983) :

À haute concentration de NO :



À faible concentration de NO :



Toutefois, le CO doit principalement son titre de GES indirect par son contrôle de la concentration des OH·, des molécules très réactives connues comme étant les « molécules de ménages » de l'atmosphère (Logan *et al.*, 1981). Ces molécules réagissent avec plusieurs autres gaz dont le C₂H₆, H₂, CH₃Cl, CH₃CCl₃, CH₃Br, H₂S, SO₂ mais principalement avec le CH₄ et le CO (Figure 1). Environ 50% des OH· sont éliminés seulement par le CH₄ et le CO (Prather, 1996). La grande compétition entre le CO et le CH₄ pour les OH· fait en sorte que la concentration de CO dans l'atmosphère influence directement la concentration de CH₄ (Daniel *et al.*, 1998, Prather, 1996), qui lui est un GES important. En prenant compte des effets indirects du CO, son PRG est près de trois fois plus grand que celui du CO₂ sur une période de 20 ans, mais devient équivalent si le calcul du PRG est fait sur une période de 100 ans (Daniel *et al.*, 1998).

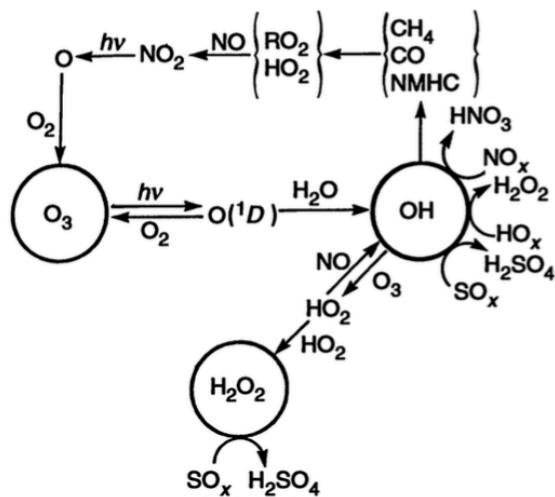


Figure 1. : Rôles des radicaux hydroxyles (OH·) dans les réactions photochimiques atmosphériques (Anne M. Thompson, 1992).

1.2.1 Distribution du CO dans l'atmosphère

Il existe plusieurs instruments et programmes pour la prise de données de CO à l'échelle mondiale. Par exemple, SCHIAMACY et MOPITT sont des instruments retrouvés sur des satellites pouvant mesurer des concentrations de CO dans l'atmosphère à l'échelle planétaire selon la réflexion des ondes dans l'atmosphère. SCHIAMACY peut mesurer plusieurs gaz mais mesure la réflexion du CO, spécifiquement à une longueur d'onde de 2,3 µm, tandis que MOPITT est un instrument spécifique pour le CO qui le mesure à une longueur d'onde de 4,7µm, étant plus sensible à la mi-troposphère (Kopacz *et al.*, 2010) (<http://www.sron.nl/sciamachy/instrument-eos-missions-1497>). Entre autres, des prises de données sont aussi réalisées par le *Global Monitoring Division* (GMD) du laboratoire *Earth System Research* (ESRL) du *National Oceanic and Atmospheric Administration* (NOAA) mesurant les concentrations de plusieurs gaz à plusieurs sites à travers le monde à la surface de la Terre (http://www.esrl.noaa.gov/gmd/about/cmdl_operating_plan.pdf).

La charge de CO dans l'atmosphère demeure approximative car on estime 20-30% d'incertitude dans la distribution globale du CO dû aux erreurs possibles dans la mesure des concentrations et aux différences de distribution du CO dans la troposphère généralement dues à sa courte durée de vie (Logan *et al.*, 1981). Il y a deux fois plus de CO dans l'hémisphère Nord que dans l'hémisphère Sud (Heidt *et al.*, 1980). De plus, les concentrations de CO varient selon les saisons. Dans l'hémisphère Nord, les pointes de hautes concentrations sont observées vers la fin de l'hiver, soit mars-avril en raison d'augmentation de combustion de matières fossiles et de biomasse ainsi qu'en raison de la diminution des vitesses de réactions avec les OH· à basse température (Evans *et al.*, 1995). Dans l'hémisphère Sud, les pointes de hautes concentrations sont plutôt observées au début de l'automne en septembre-octobre (Worden *et al.*, 2013). Ces pointes éphémères illustrent bien le fait que la durée de vie du CO est variable : passant de plus d'un an l'hiver à moins de 0,1 an l'été (Badr *et al.*, 1995, Stevens *et al.*, 1972). Toutefois, des concentrations moyennes 60-150 ppbv sont retrouvées selon l'hémisphère (Seiler, 1974) et il est possible de retrouver jusqu'à 1 ppmv de CO en milieu urbain (Burnett *et al.*, 1998).

Une augmentation globale d'environ 80 ppbv de CO a été mesuré depuis le temps préindustriel (Thompson *et al.*, 1993), mais une diminution de la concentration de CO a été remarquée au début des années 1990. Elle serait expliquée à 70% par la baisse des émissions dû à la mise en place de programmes de contrôles d'émissions, et à 30% par la dégradation de

O_3 de la stratosphère qui aurait eu lieu après l'éruption du mont Pinatubo en juin 1991, produisant plus de OH· (Zhang *et al.*, 2001). La concentration globale de CO aurait diminué d'environ 0,52 ppbv/année de 1991 à 2001 (Novelli *et al.*, 2003). Les tendances de CO mesurées pour la dernière décennie, soit de 2001 à 2010 par Yoon et Pozzer ont démontré une diminution de $35,5 \pm 5,8$; $59,6 \pm 9,1$; et $13,7 \pm 9,5$ ppbv/décennie pour l'Europe, l'est des États-Unis et le nord de l'Australie mais une augmentation de $8,9 \pm 4,8$ ppbv/décennie pour le sud de l'Asie (Yoon *et al.*, 2014). Des diminutions des concentrations globales de CO autour de 1% par année ont été observées pour l'hémisphère nord. Les données pour l'hémisphère Sud sont moins significatives mais une tendance est quand même observée (Worden *et al.*, 2013). Une diminution importante de CO a été remarquée entre 2007 à 2009 et serait directement reliée à un taux plus faible d'émissions issues de la combustion de combustibles fossiles et de biomasse (Fortems-Cheiney *et al.*, 2011), ce qui concorderait avec la période de récession mondiale (Yurganov *et al.*, 2010). Il est rassurant de constater que les émissions de CO ne sont pas à la hausse, mais elles doivent continuer à être surveillées puisque leur augmentation perturberait la chimie de plusieurs autres gaz.

1.2.2 Sources et puits du CO (bilan)

La chimie atmosphérique est très complexe mais il faut comprendre le système en entier afin d'évaluer un procédé atmosphérique en particulier. Les modèles de transport chimique sont des modèles mathématiques en trois dimensions de la composition atmosphérique permettant d'intégrer et d'étudier un procédé atmosphérique tout en prenant compte de toutes ses interactions. Ces modèles permettent d'établir les liens entre les différentes émissions et les concentrations de ce gaz retrouvées dans l'atmosphère en prenant compte des facteurs affectant le gaz en question comme le transport, la chimie, les émissions et les processus de déposition. Par exemple, afin d'établir des budgets de CO, des modèles de transport chimique comme GEOS-Chem et MOZART sont utilisés.

Les bilans atmosphériques peuvent être construits par analyses ascendantes « Bottom-up » ou à partir d'analyses inverses « top-down ». Les analyses « bottom-up » génèrent des budgets en utilisant des données géographiques et statistiques d'émissions pour ensuite extrapoler au niveau global. Les valeurs obtenues par ces analyses sous-estiment parfois les différents types d'émissions en raison des grandes variations spatio-temporelles qui leur sont associées. Les analyses « top-down » dépendent des données de concentrations retrouvées

dans l'atmosphère pour ensuite calculer les émissions des différentes sources. Dépendamment du type d'analyses réalisées, différentes bases de données peuvent être utilisées pour obtenir des données de concentrations, comme MOPITT, ou des données plus spécifiques d'émissions comme par exemple, GFED (Global Fire Emissions Database) pour les émissions provenant de feu ou EDGAR pour les émissions anthropogéniques. Il est possible d'optimiser les budgets en appliquant les signatures isotopiques. Les différentes sources de CO ont chacune des signatures isotopiques bien précises de ^{13}C et de ^{18}O , exprimé en pour mille (\textperthousand) permettant de les retracer.

Puisqu'il existe différentes façons d'établir les bilans de CO à partir des différents modèles, données et nombreux paramètres à tenir compte, les valeurs obtenues d'une étude à une autre peuvent varier grandement et demeurent des estimations. Toutefois, ces bilans atmosphériques contenant plusieurs incertitudes sont en constante amélioration. Le tableau 1 affiche des bilans atmosphériques du CO obtenus selon quelques études utilisant différents modèles, donnés et types d'analyses. Par exemple, les valeurs des équipes de Duncan et Holloway ont été obtenues par « bottom-up », et les valeurs des équipes de Muller, Arellano et Park par « top-down ».

L'activité anthropogénique est en grande partie responsable des émissions de CO par l'entremise de combustion de combustibles fossiles et de biomasse mais l'oxydation du CH₄ et des composés organiques volatiles non méthaniques (COVNM_s) sont aussi de grandes sources de CO. L'oxydation du CH₄ contribue à environ 25% des sources de CO (Logan *et al.*, 1981). De plus, les océans produisent de faibles émissions de CO, prisent parfois en compte dans les budgets de CO. Selon le rapport du groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) de 2001, 700 Tg/année de CO est émis par la combustion de biomasse, 650 par la combustion de matières fossiles et 800 Tg/année par l'oxydation du CH₄. Les sources totales de CO s'élèvent à 2780 Tg/année comparativement à 2400 Tg/année rapporté du rapport du GIEC de 1990 (Ehhalt *et al.*, 2001, Watson *et al.*, 1990). C'est valeurs sont relativement près des valeurs obtenues par d'autres études démontrées dans le tableau 1.

Pour ce qui est des puits de CO, ils sont beaucoup moins nombreux. Les OH· comptent pour le plus grand puits de CO en étant responsable de son élimination à environ 80% (Levy, 1971, Weinstock, 1969). Ces molécules sont très réactives et on les retrouve à moins de 0,1 ppbv dans l'atmosphère (Monks, 2005). Environ $2,1 \times 10^{15}$ g CO par an est éliminé via les OH· (Crutzen *et al.*, 1983). Le second puits du CO provient du sol, oxydant environ 15% des émissions de CO annuel par l'entremise de bactéries spécialisées (Conrad, 1996), qui représente un puits de 16-636 Tg CO⁻¹ année⁻¹ (King, 1999b).

Tableau 1. Sources de CO (Tg/année) provenant de différentes analyses.

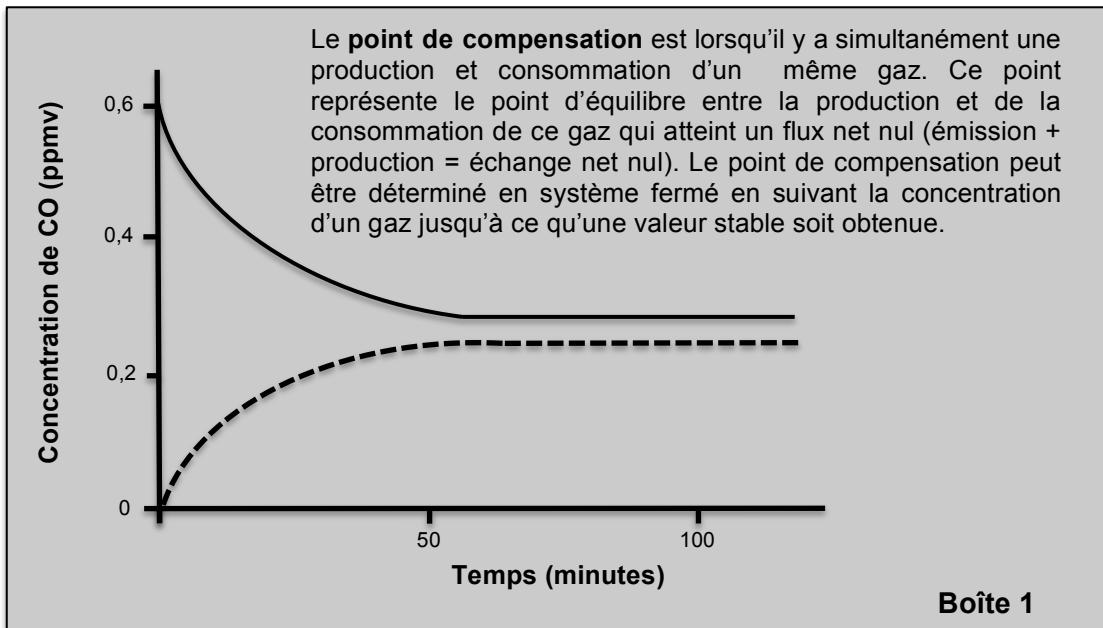
	Duncan et al., 2007	Holloway et al., 2000	Muller et Stavrakou, 2005	Arellano et al., 2006	Park et al., 2015
Modèles et données	GEOS-Chem 3D GEOS-GMAO	GFDL-GCTM NOAA-GMD	IMAGES NOAA-GMD	GEOS-Chem MOPITT	MOZART-4 NOAA-GMD
Année	1988-1997	-	1997	2000-2001	2004
Combustibles fossiles et biocarburants	653-676	300	760	841	716
Combustion de biomasse	451-573	748	359	501	377
Océans	-	-	23	-	20
Oxydation du CH ₄	778-861	760	870	820	919
Oxydation NMHC total	-	-	774	-	454
Biogénique HC	354-379	683	-	394	-
Biogénique	-	-	142	-	97
Sources totales	2236-2489	2491	2928	2556	2583

1.3 Consommation du CO par le sol

Le sol est le deuxième plus grand puits de CO mais est toutefois le puits le plus incertain puisqu'il est contrôlé par des processus biotiques, soit par l'entremise de bactéries. Initialement, il était loin d'être évident que les bactéries étaient responsables de cette consommation de CO et qu'il n'était pas question de processus abiotiques. Plusieurs études ont confirmé l'implication des bactéries par les résultats suivants :

- La température de réaction optimale est de 30° et l'activité de consommation du CO est perdue en chauffant le sol. Ce qui démontre que la réaction n'est pas produite par des processus chimiques, car dans ce cas, la vitesse de réaction augmenterait avec la température (Ingersoll *et al.*, 1974). De plus, les températures trop élevées éliminent les bactéries, affectant la consommation de CO du sol.
- L'ajout d'antibiotiques dans le sol, inhibe la consommation de CO, indiquant que les bactéries sont responsables de l'activité (Conrad *et al.*, 1980).
- Aucune activité n'est observée dans un sol stérile et l'ajout de sol non stérile permet de restaurer l'activité (Inman *et al.*, 1971a).

Dans certains sols et selon différentes périodes de l'année, le sol peut aussi être une source nette de CO. Cette production de CO dans les sols provient de processus chimiques comme la dégradation des acides humiques et autres composés phénolés présent dans le sol (Conrad *et al.*, 1985, Moxley *et al.*, 1998a). Le taux d'oxydation de CO observé dans un sol est toujours le résultat d'une production et d'une consommation simultané, notamment il existe un point de compensation (voir boîte 1) entre le flux de production et de consommation de CO qui varie selon les sols (Conrad, 1994). Globalement, la production de CO par les sols est négligeable puisqu'au total le sol est un puits de CO avec un taux d'oxydation net de 7-40 Tg/année (Potter *et al.*, 1996).



Plusieurs facteurs peuvent influencer le potentiel de consommation de CO d'un sol. Le contenu en eau est un facteur important. Un sol trop sec ou avec un trop grand contenu en eau aura une faible consommation de CO. Un sol sec réduit l'activité des bactéries et un sol contenant trop d'eau limite (ou ralenti) la diffusion du CO dans le sol permettant d'être assimilé par les bactéries (Heichel, 1973). Le contenu en eau optimal d'un sol n'est pas universel mais se trouve entre ces deux extrêmes selon la capacité de rétention d'eau du sol (King *et al.*, 2002, Moxley *et al.*, 1998b). La température influence aussi la consommation du CO dans le sol. Comme mentionné plus haut, la température optimale se retrouve souvent entre 20-30° et des températures plus élevées peuvent engendrer une production nette de CO (King, 1999a).

L'activité anthropogénique affecte les cycles et les échanges de gaz entre l'atmosphère et la biosphère notamment par les pratiques de gestion des terres qui influencent la composition des sols. Ces modifications ont un impact sur le potentiel de consommation des sols. Le tableau 2 démontre bien la consommation moyenne de CO par année par mètre carré d'un sol selon son type de végétation et composition. Les sols agricoles sont parmi les terres ayant la plus faible consommation de CO par unité de surface. Plusieurs études ont démontré que les sols agricoles ont un taux d'oxydation du CO plus faible que des sols de forêts (Chan *et al.*, 2006, Ingersoll *et al.*, 1974). Les traitements chimiques utilisés sur les terres agricoles comme les pesticides et les fongicides peuvent avoir un certain effet sur la consommation de CO tout comme l'application de fertilisants. Ces traitements ont démontré une diminution de l'activité de consommation de CO non significative mais tout de même présente (Ingersoll *et al.*, 1974, Moxley *et al.*, 1998b). Toutefois, la diminution d'activité dans les sols agricoles serait

probablement plus expliquée par la diminution de matière organique à la surface du sol. Par contre, les résultats de King et Sanhueza contredisent la littérature en démontrant qu'un sol de terres agricoles consomme plus de CO comparativement à un sol de forêt à proximité (King, 2000, Sanhueza *et al.*, 1994). Les conditions d'humidité et de température des différents sols étudiés pourraient expliquer ces contradictions, mais peu d'études ont tenté de clarifier cette notion.

Selon le tableau 2, les sols de forêts sont largement plus efficaces pour éliminer le CO, même s'ils peuvent produire plus de CO que les sols agricoles en raison de leur fort contenu en matière organique (Gödde *et al.*, 2000). Le point de compensation est ainsi plus élevé mais en somme, la vitesse consommation est largement supérieure. Cela est justement expliqué en grande partie par la grande quantité de matière organique de ces sols, ce qui stimule l'activité bactérienne mais qui permet plus de dégradation photochimique des acides humiques. Bref, les sols de forêt consistent en un élément important du puits de CO. Il est primordial de mieux comprendre ce puits important de CO jouant ainsi un effet sur le budget global du CO, qui exerce une grande influence sur la chimie atmosphérique d'autres gaz impliqués dans la régulation du climat.

Tableau 2. Surface globale et consommation de CO de différents biomes terrestres (Modifié de Badr et Probert, 1995).

Biomes terrestres	Aire (10^{12} m^2)	Consommation moyenne de CO (g année $^{-1} \text{ m}^2$)
Terres agricoles	11,91	33,20
Prés	4,77	69,23
Prairies tropicales	9,01	342,43
Prairies tempérées	8,03	33,40
Steppe	8,94	29,61
Forêts de montagne	12,07	67,84
Forêts boréale	12,20	49,23
Forêts mixtes et feuillus	3,42	98,38
Forêts de pins du Sud	0,75	195,29
Forêts tropicales à feuilles caduques	4,61	426,99
Forêt humides tropicales	16,96	311,04
Toundra	2,98	47,49
Désert	28,13	20,15
Couvert d'eau, de glace, de routes ou de structures	23,83	0,00

1.3.1 Bactéries oxydant le CO

Le CO est consommé dans le sol par des bactéries spécialisées. Des bactéries aérobies et anaérobies ont la capacité de consommer le CO, mais ces deux groupes utilisent des enzymes complètement différentes. L'enzyme utilisé en aérobie possède une CO-déshydrogénase contenant du molybdène (MoFeS), contrairement à la CO-déshydrogénase contenant du nickel (NiFeS) utilisée en anaérobie (Dobbek *et al.*, 2002, Dobbek *et al.*, 2001). Puisque les bactéries aérobies se retrouvent dans les premiers centimètres du sol exposé à l'atmosphère, celles-ci sont responsables de la consommation de CO atmosphérique (King, 1999a). Ainsi, le puits biologique du CO est contrôlé seulement par les bactéries aérobies. Les prochaines sections aborderont de leur physiologie, biochimie, génétique et distribution dans l'environnement afin de mieux comprendre ce puits important de CO.

1.3.2 Carboxydovores et carboxydotrophes

Le CO est reconnu pour sa toxicité chez les organismes vivants due à sa forte affinité pour l'hémoglobine, soit d'environ 250 fois plus grande que l'oxygène (Davenport, 1949, Rodkey *et al.*, 1974), diminuant ainsi la capacité de transport d'oxygène du sang tout en affectant la dissociation de l'oxygène de l'hémoglobine dans les tissus menant à l'hypoxie (Roughton *et al.*, 1944). Chez les microorganismes, le CO se lie à l'oxydase terminale des organismes aérobies et inhibe la croissance de la majorité d'entre eux (Cypionka *et al.*, 1982). Pourtant, certaines bactéries sont insensibles au CO et ont la capacité de le consommer. Les bactéries pouvant utiliser le CO ont une chaîne respiratoire ramifiée. L'embranchement se fait au niveau du cytochrome de type b ou de l'ubiquinone. Le cytochrome a_1 est l'oxydase terminale pour la chaîne respiratoire sensible au CO tandis que le cytochrome o (cytochrome b_{563}) est l'oxydase terminale pour la chaîne respiratoire insensible au CO (Cypionka *et al.*, 1983) car elle n'a que peu ou aucune affinité pour le CO. Ainsi, c'est en grande partie à ce niveau que la tolérance au CO d'une bactérie est déterminée. Une oxydase terminale ayant une faible affinité pour le CO permettra à la bactérie de supporter des concentrations de CO beaucoup plus élevées. D'ailleurs, l'inhibition des bactéries par le CO n'est pas seulement une question de concentration de CO mais plutôt du rapport de CO et O₂(Cypionka *et al.*, 1982).

Zavarzin et Nozhevnikova, ont introduit le terme carboxydobactéries pour les bactéries pouvant consommer le CO et ont proposé qu'il existerait des carboxydobactéries avec différentes tolérances et affinités pour le CO (Zavarzin *et al.*, 1977). King a ensuite introduit les termes carboxydrotrophes et carboxydovores pour désigner deux groupes de carboxydobactéries avec des affinités faible ou forte pour le CO. Les carboxydrotrophes sont les bactéries aucunement ou très peu inhibées par le CO, qui tolèrent des concentrations très élevées. Ces bactéries ont une faible affinité pour le CO et peuvent le consommer à des concentrations de plus de 1% (10 000 ppmv), même jusqu'à plus de 90%, mais ne peuvent pas le consommer à faibles concentrations. Elles peuvent utiliser le CO comme seule source de carbone et d'énergie. Elles utilisent les électrons générés au cours de la réaction pour former de l'énergie et assimilent une partie du CO₂ formé par la réaction de la CODH pour croître (Meyer *et al.*, 1983). *Oligotropha carboxidovorans* est l'une des premières carboxydrotrophes à avoir été étudiée, tout comme *Hydrogenophaga pseudoflava*.

Plusieurs autres souches ont été identifiées comme carboxydrotrophes mais le K_{m(app)} de ces bactéries se retrouve entre 400-950 nM (Conrad, 1996). Le K_m est une valeur attribuée pour représenter l'affinité d'une enzyme pour son substrat (une forte affinité est associée à une faible valeur de K_m). Les valeurs de K_m de ces carboxydrotrophes de 50 à 200 plus élevées que ceux du sol, ont semé le doute sur l'existence d'autres bactéries présentent dans le sol, consommant le CO à plus grande affinité (Conrad *et al.*, 1981).

Les carboxydovores sont les bactéries aérobies pouvant consommer le CO seulement à faibles concentrations (< 1%), se rapprochant des concentrations atmosphériques. Pour se faire, ces bactéries ont une très haute affinité pour leur substrat avec un K_{m(app)} inférieur à 100 ppmv pour le CO (Hardy *et al.*, 2001). Ainsi, c'est ce groupe de carboxydobactéries qui sont importantes dans le bilan global de CO. Toutefois, ces bactéries sont inhibées par des concentrations trop élevées de CO, supérieures à 1% (King, 2003a). Puisqu'elles utilisent le CO qu'à très faible concentration, il est assumé que ce substrat est utilisé seulement comme source d'énergie supplémentaire pour ces bactéries, même s'il fournit moins de 1% de l'énergie totale que peut fournir du glucose (Cunliffe, 2013).

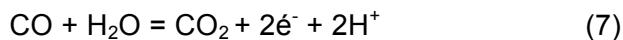
Il est plus ardu d'isoler une souche carboxydovore en raison du manque de pression de sélection induit par les traces de CO utilisées par ces bactéries. Pour isoler une bactérie carboxydrotrophe, de fortes concentrations de CO peuvent être ajoutées et seulement les bactéries pouvant croître en présence de CO à ces concentrations seront obtenues. En utilisant du CO à concentration atmosphérique, ce qui est nécessaire pour l'isolement de

carboxydovores, aucune pression de sélection n'est exercée et différentes bactéries qui ne consomment pas nécessairement le CO atmosphérique seront également isolées. Jusqu'à présent, seulement quelques souches de carboxydovores ont été isolées et étudiées. Ce trait est assez diversifié, puisqu'il a été retrouvé dans divers embranchements bactériens. Chez les *Protéobactéries*, des carboxydovores ont été retrouvés dans toutes les différentes classes. Les *Alphaprotéobactéries* comprennent entre autres, *Ruegeria pomeroyi* DSS-3, plusieurs souches de *Bradyrhizobium* et plusieurs souches de *Stappia* mais aucun $K_{m(app)}$ n'a été déterminé pour ces souches (Cunliffe, 2013, King, 2003a, Weber *et al.*, 2007). La souche *Aminobacter* sp. COX consomment aussi le CO atmosphérique et un $K_{m(app)}$ de 15 ppmv a été mesuré pour cette souche. Cette valeur de $K_{m(app)}$ se rapproche des valeurs obtenues des sols (Hardy *et al.*, 2001, King, 2003a). Weber et King ont démontré que les bactéries carboxydovores étaient communes chez le genre *Burkholderia* (*Betaprotéobactéries*), pouvant consommer le CO se retrouvant à moins de 100 ppmv (Weber *et al.*, 2012). Chez les *Deltaprotéobactéries*, rares sont les souches aérobies, toutefois la souche aérobie *Haliangium ochraceum* dont la présence du gène *coxL* de type I est observée, a récemment été étudiée pour sa capacité à consommer le CO atmosphérique, qui s'est avérée positive avec une utilisation du CO à moins de 5 ppmv (Quiza *et al.*, 2014). Pour ce qui est des *Gammaprotéobactéries*, la souche *Alkalilimnicola ehrlichii* MLHE-1 peut consommer le CO de 5 ppmv à 0,1 ppmv (Hoeft *et al.*, 2007). Une autre souche de *Gammaprotéobactéries* ayant démontré une consommation de CO atmosphérique est la souche *Streptomonas* sp. LUP (King, 2003a). Chez les *Actinobactéries*, *M. tuberculosis* H37Ra et *M. smegmatis* consomment le CO à hautes concentrations (20-30%) et à faible concentration (<de 50 ppmv) mais ne peuvent pas croître à partir du CO à faible concentration (King, 2003b). La souche *Mycobacterium* JC1 a été capable de consommer le CO à partir de 100 ppmv mais ne peut pas consommer le CO présent dans l'air ambiant. Toutefois, elle peut aussi consommer le CO comme seule source de carbone et d'énergie (Cho *et al.*, 1985). De ce fait, l'attribution de cette souche au groupe de bactéries carboxydovores est ambiguë. La bactérie carboxydovore la plus récemment isolée et étudié a été assignée taxonomiquement au nom de *Thermogemmatispora carboxidovorans* PM5 de l'embranchemen *Chloroflexi*. Elle peut consommer le CO à des concentrations de < 0,5 ppmv. D'autres souches de la classe *Ktedonobacteria*, dont fait partie cette souche, ont aussi démontré la capacité de consommer le CO à faibles concentrations < 50 ppmv (King *et al.*, 2014).

1.4 La CO-déshydrogénase (CODH)

1.4.1 Métabolisme

Afin de pouvoir consommer le CO, les bactéries aérobies du sol possèdent une enzyme, communément appelée la CO-déshydrogénase (CODH), catalysant la réaction suivante :



Ainsi, une molécule de CO₂, deux électrons et deux protons sont issues de la réaction. Chez certaines bactéries, le CO₂ peut ensuite être assimilé comme source de carbone. La capacité d'utiliser le CO₂, une molécule inorganique, comme seule source de carbone est très répandue chez les bactéries. Le terme autotrophie est utilisé afin de décrire ce type d'apport en énergie. Différentes voies métaboliques permettant d'assimiler le CO₂ existent chez les bactéries autotrophes. En aérobiose, le CO₂ peut être utilisé dans le cycle de Calvin-Bassham-Benson (CBB) par l'entremise de l'enzyme RubisCO (ribulose biphosphate carboxylase). Le cycle de Krebs inverse (ou cycle de l'acide citrique inverse) utilise le CO₂ et l'H₂O pour former des molécules plus complexes au lieu du cycle de Krebs/cycle de l'acide citrique originaux oxydant les molécules complexes en CO₂ et H₂O (Evans *et al.*, 1966). L'utilisation de ce cycle a été observé chez *Hydrogenobacter thermophilus* (Shiba *et al.*, 1985) et *Aquifex pyrophilus* (Beh *et al.*, 1993) en aérobiose facultative. Enfin, le cycle 3-hydroxypropionate est une autre voie métabolique utilisée pour fixer le CO₂ pour la production de pyruvate mais a aussi été observé seulement en aérobiose facultative (Strauss *et al.*, 1993).

Bien que le CO₂ résultant de l'oxydation du CO peut être fixé, ce n'est pas tout le CO₂ produit lors de la réaction qui finit par l'être. Environ 2% à 16% du CO₂ seulement serait assimilé dans la biomasse. Les souches avec la plus faible assimilation étant *Azomonas* sp. ainsi que *Azobacter* sp., et la souche *O. carboxidovorans* ayant la plus grande assimilation (Kim *et al.*, 1983, Spratt *et al.*, 1981, Whalen *et al.*, 2001). Certaines bactéries n'utilisent pas le CO₂ comme source de carbone, mais utilisent seulement les électrons et les protons produits lors de la réaction comme source d'énergie supplémentaire par l'entremise de la chaîne de transport d'électrons, qui forme un gradient de protons à travers la membrane, servant à synthétiser de l'ATP (Kiessling *et al.*, 1982).

Certaines bactéries n'oxydent pas le CO en croissance hétérotrophique (consommation de composés organiques exogènes). Le CO est seulement utilisé comme source d'énergie supplémentaire quand les ressources sont limitées. D'autres bactéries ont la capacité d'utiliser le CO simultanément que leurs substrats utilisés en croissance hétérotrophique, comme le pyruvate. Dans ces conditions, il est cas de mixotrophie. Ainsi, lors de ce type de croissance, la CODH fournit des électrons supplémentaires dans la chaîne respiratoire, en plus des électrons reçus par le NADH provenant du cycle de Krebs (Figure 3). La figure 3 résume comment l'énergie et le carbone sont fournis par la consommation de CO, présente la structure de la chaîne respiratoire chez les bactéries résistantes au CO et finalement l'assimilation du CO₂ par la voie Calvin-Bassham-Benson.

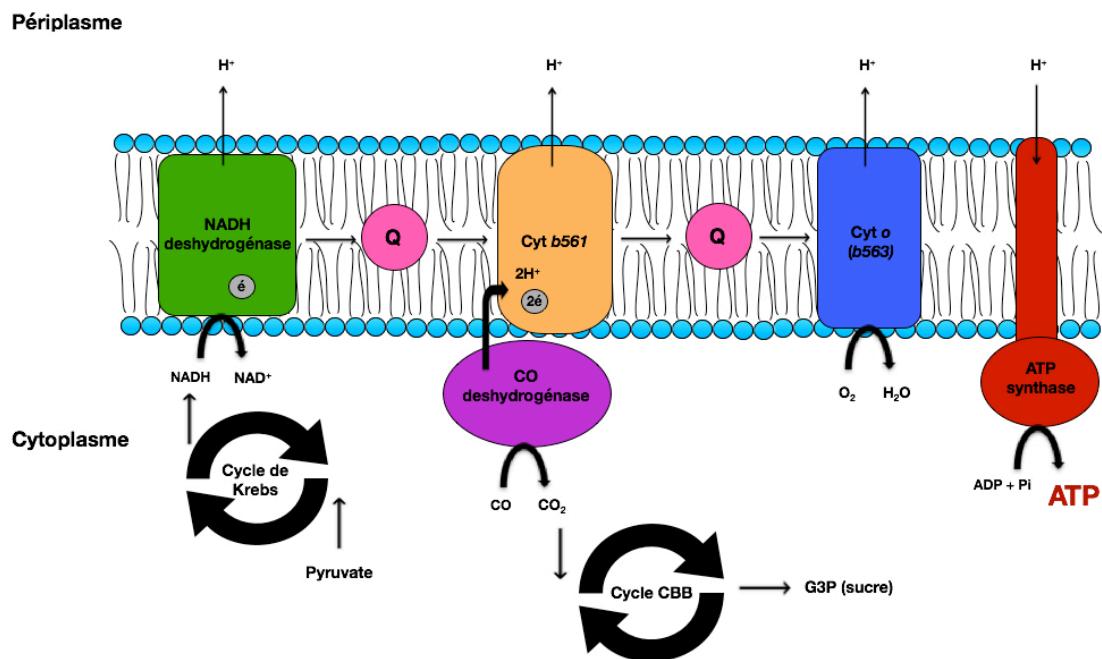


Figure 2. Chaîne de transport d'électrons simplifiée des bactéries aérobies insensibles au CO.

Certaines bactéries, comme *P. carboxydoflava*, ont démontré une croissance mixotrophique. En croissance hétérotrophique, elle utilise le CO comme source d'énergie supplémentaire, lui permettant d'assimiler encore plus de substrat organique (Kiessling *et al.*, 1982). La consommation de glucose est aussi stimulée en présence de CO chez *Mycobacterium* sp. JC1 (Kim *et al.*, 1989). *Stappia aggregata* (*Labrenzia aggregata*), quant à elle, est une bactérie carboxydovore qui peut utiliser le CO en présence de glucose. Une diminution de la consommation de CO est observée à des concentrations de glucose de plus de 0,5 mM mais le CO et le glucose sont consommés en mode mixotrophique (Weber *et al.*, 2007).

Les concentrations atmosphériques de CO peuvent soutenir des populations bactériennes de 10^2 à 10^7 bactéries/g de sol_(poids sec) (Conrad, 1999), pourtant on peut retrouver jusqu'à 10^{11} copies de séquences *coxL*/g de sol_(poids sec) (Quiza *et al.*, 2014). Cela suggère que la majorité des carboxydovores dans le sol utilisent obligatoirement d'autres sources de carbone et d'énergie en plus, afin de pouvoir soutenir une aussi grande population. Ainsi, ces bactéries utiliseraient un mode de croissance mixotrophique et sont d'intérêt majeur pour comprendre le puits biologique de CO.

1.4.2 Structure

La CODH est une enzyme de la famille des hydroxylases contenant du molybdène (Hille, 2005). Elle a été largement étudiée chez *O. carboxidovorans* (Dobbek *et al.*, 1999). C'est d'ailleurs chez cette bactérie que l'enzyme a été purifiée et cristallisée pour la première fois. Par la suite, elle fut aussi cristallisée chez *H. pseudoflava* (Hänelmann *et al.*, 2000). Ces deux CODH sont toujours les seules à avoir été cristallisées à ce jour. La CODH, est un dimère de trois sous-unités structurales de tailles différentes, la petite (CoxS), la moyenne (CoxM) et la grande (CoxL), formant une structure symétrique ($L_2M_2S_2$) en forme de papillon (Figure 2) (Meyer *et al.*, 1986). Chez *O. carboxidovorans*, les sous-unités sont de 166 a.a, 288 a.a et 809 a.a respectivement avec un poids moléculaire de 18 kDa, 30 kDa et 89 kDa. Le poids moléculaire de l'enzyme fonctionnelle avec ses cofacteurs, fait environ 277 kDa (Schubel *et al.*, 1995). La taille des sous-unités varie selon les souches mais demeurent près des valeurs obtenues chez d'*O. carboxidovorans*.

La petite sous-unité est une protéine fer-soufre qui porte un centre [2Fe-2S] de type I et II. La moyenne sous-unité est une flavoprotéine qui comporte la flavine adénine dinucléotide (FAD), un cofacteur lié de façon non-covalente à la flavoprotéine (Kim *et al.*, 1981) pour un rapport de 1 : 1 (Meyer *et al.*, 2000). La grande sous-unité est une molybdoprotéine comprenant le cofacteur molybdoptérine cytosine dinucléotide (MCD) lié aussi de façon non-covalente, qui lui comprend le molybdène (Mo), formant le site actif de l'enzyme (Johnson *et al.*, 1990). Initialement, la présence de sélénium avait été rapportée dans le site actif de la protéine par cristallographie. Une résolution plus élevée de 1,09 Å du cristal de la CODH a révélé la présence d'un atome de cuivre dans la molybdoprotéine où se retrouvait le sélénium (Dobbek *et al.*, 2002). Ceci serait dû à une utilisation d'une longueur d'onde à faible énergie inappropriée lors de la première analyse (Dobbek *et al.*, 1999). Ainsi, le site actif est composé d'un centre bimétallique de Mo et de Cu lié par l'entreprise d'un atome de soufre (S) (Gnida *et al.*, 2003).

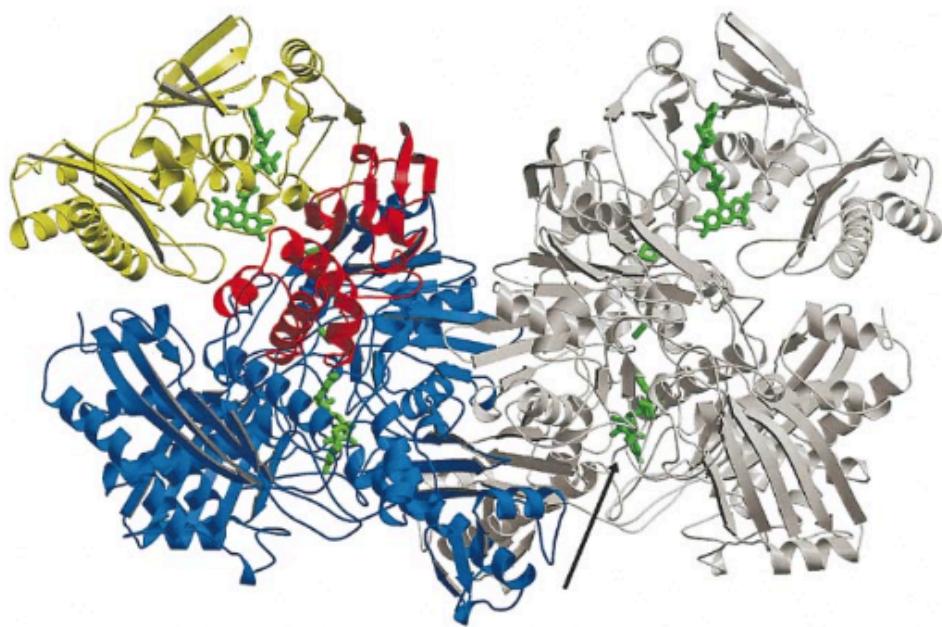


Figure 3. La CO-déshydrogénase. Les sous unités sont représentées en bleu (CoxL), rouge (CoxS) et jaune (CoxM). Le Mo, les deux centres [2Fe-2S] et le FAD sont représentés en vert dans leur sous-unité respective. La flèche démontre le canal menant au site actif (Tiré de Hanzelmann *et al.*, 2000).

Le site actif de l'enzyme est reconnu par son motif d'acides aminés AYXCSFR conservé. La cystéine présente dans le site actif est très importante car la participation de son atome de soufre est cruciale dans la réaction enzymatique oxydant le CO (Figure 3). L'oxydation du CO se produit au site du Mo (VI) en se liant à l'atome de cuivre, lié au soufre de la cystéine³⁸⁸ du site actif (Siegbahn *et al.*, 2005). La liaison du CO au cuivre le rend alors susceptible à une attaque nucléophile par l'oxygène du molybdène (Mo=O). La liaison du carbone du CO avec le soufre liant le molybdène et le cuivre est thermodynamiquement non-favorable, laissant le Mo (VI)-S-Cu (I) intact. Une molécule d'eau vient libérer le CO₂ et le Mo est réduit pour former Mo (IV). Le Mo (IV) libère deux électrons et deux protons pour régénérer le Mo(VI) initial (Jeoung *et al.*, 2014). Les électrons libérés du Mo passe pour les deux centres [2Fe-2S], sont transférés au FAD, pour ensuite être envoyés aux quinones dans la membrane cytoplasmique et ainsi dans la chaîne respiratoire insensible au CO avec l'O₂ comme accepteur d'électron final. De ce fait, le passage des électrons dans la chaîne respiratoire forme un gradient de proton produisant de l'énergie pour la cellule sous forme d'ATP (Gremer *et al.*, 2000). Pour générer ce gradient dans la membrane, la CODH doit y être attachée. Chez *O. carboxidovorans*, 87% des CODH sont associés à la membrane interne en phase exponentielle mais y seraient liées que faiblement puisque cette valeur diminue à 50% en phase stationnaire (Meyer *et al.*, 1986, Rohde *et al.*, 1985). Cypionka et son équipe, quant à eux, ont déterminé que la CODH serait une protéine soluble, appuyant le fait que l'enzyme serait faiblement liée à la membrane (Cypionka *et al.*, 1980).

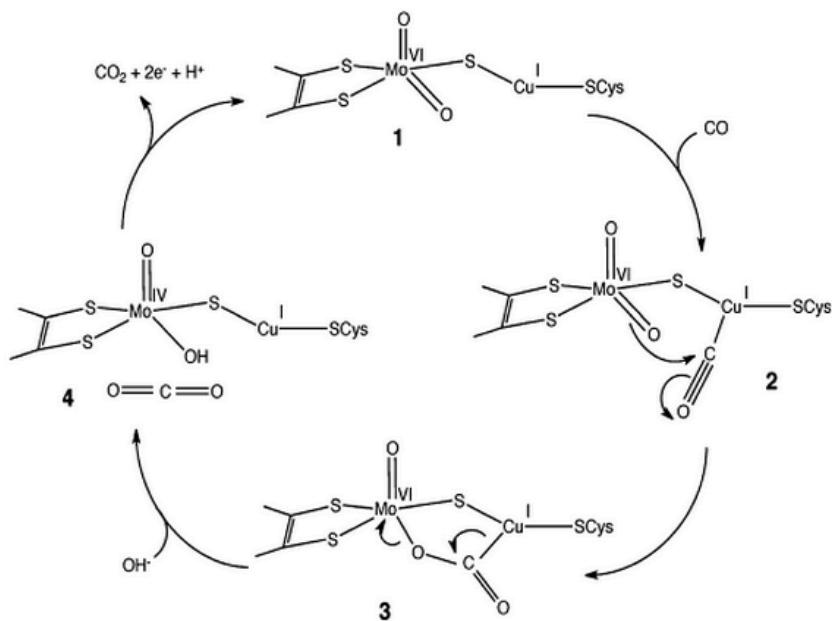


Figure 4. Mécanisme d'oxydation du CO de la CO déshydrogénase. (1) Le centre bimétallique Mo(VI)-Cu(I) est prêt pour recevoir le CO. (2) Le CO se lie au Cuivre, l'activant pour une attaque nucléophile par le liguant oxo lié au Mo. (3) Le site actif dans son état intermédiaire se réarrange et une molécule d'eau engendre la libération de CO_2 . (4) Le centre Mo(IV)-Cu(I) relâche deux électrons et deux protons pour régénérer le centre Mo(VI)-Cu(I) initiale. (Tiré de Jeoung et al., 2014).

1.4.3 Gènes structuraux et accessoires

Les gènes codant pour la CODH en aérobie ont été assigné au nom de *cox* (carbon monoxide oxidases genes) (Hugendieck et al., 1992). Les gènes structuraux codant pour les trois sous-unités de la protéine se nomment *coxS*, *coxM*, et *coxL* pour la petite, moyenne et grande sous-unité, respectivement. Les trois gènes font partie du même opéron ainsi que les gènes accessoires nécessaires pour la maturation de l'enzyme (Kang et al., 1999). Chez *O. carboxidovorans*, les gènes structuraux sont transcrits suivant l'ordre *coxMSL* (Schubel et al., 1995). Cette souche contient les gènes accessoires suivants : *coxB*, *coxC*, *coxD*, *coxE*, *coxF*, *coxG*, *coxH*, *coxI* et *coxK*. Les gènes sont tous transcrits du même sens, à l'exception du gène *coxK*, transcrits dans le sens opposé. Chez les autres souches, les gènes *coxD*, *coxE* et *coxF* sont habituellement toujours présent à la base mais la présence des autres gènes accessoires peut varier. L'ordre dans lequel les gènes se retrouvent sur l'opéron peut aussi varier selon les souches (King et al., 2007).

Quelques études ont été réalisées afin de déterminer l'implication des gènes accessoires sur l'opéron *cox* pour la maturation de la CODH. Des mutations des gènes fonctionnels ont été réalisées. Le mutant *coxD* est non fonctionnel. L'enzyme contient le molybdène mais il y a absence du soufre permettant la liaison μ_2 . L'hypothèse est que *coxD* servirait de chaperonne pour déplier la grande sous-unité afin de rendre le centre accessible pour l'insertion du soufre et du cuivre (Pelzmann *et al.*, 2009). Les mutants *coxE* et *coxF* forment, tous les deux, des CODH non fonctionnelles avec l'absence de l'atome de Cu. Les gènes *coxEF* auraient donc une fonction dans l'ajout du cuivre au site actif de l'enzyme (Gnida *et al.*, 2003). Les mutants *coxG* produisent tout de même une CODH fonctionnelle et le gène. Le gène *coxG* jouerait un rôle dans l'encrage de la CODH à la membrane cytoplasmique (Fuhrmann *et al.*, 2003). Les gènes *coxHI* ne sont pas impliqués dans la synthèse d'une CODH fonctionnelle. Le gène *coxI* coderait pour une protéine cytoplasmique, mais dont le rôle exact est inconnu.

Chez *O. carboxidovorans*, l'opéron se retrouve sur un méga plasmide de 128 kb, pHCG3. Pour les autres bactéries oxydant le CO, l'opéron peut aussi se retrouver sur un plasmide, ou directement sur le chromosome, comme chez *Pseudomonas thermocarboxydovorans* (Black *et al.*, 1990). Fuhrmann et son équipe ont émis l'hypothèse que *O. carboxidovorans* serait l'hôte original du plasmide pHCG3 contenant les gènes structuraux et accessoires de la CODH puisque le contenu en GC du plasmide (60,6%) est très similaire au contenu en GC du chromosome (62,8%) (Fuhrmann *et al.*, 2003). Les CODH des bactéries à Gram négatif sont des enzymes inductibles, à l'exception de *H. pseudoflava* et de *Mycobacterium* sp. JC1 dont l'enzyme est exprimée constitutivement, qu'il y ai des composés organiques disponibles ou non (Kiessling *et al.*, 1982, Ro *et al.*, 1993).

1.4.4 Gène *coxL*

Des analyses phylogénétiques du gène *coxL* ont révélé la présence de deux principaux génotypes *coxL* : les formes I et II. La forme I, présente chez les bactéries *Oligotropha*, *Mycobacterium* et *Pseudomonas*, fut nommé OMP tandis que la forme II présente chez les bactéries *Bradyrhizobium*, *Mesorhizobium* et *Sinorhizobium*, fut nommé BMS. Certaines souches peuvent posséder les deux versions du gène *coxL*, comme *Burkholderia xenovorans* LB400, seulement *coxL* de forme I, comme *O. carboxidovorans*, ou seulement *coxL* de forme II, comme *Bradyrhizobium japonicum* USDA 110. Certaines souches peuvent même posséder deux versions du même gène, comme c'est le cas pour *Burkholderia xenovorans* LB400. Ainsi, cette souche possède deux gènes *coxL* de forme I et un gène *coxL* de forme II (King *et al.*, 2007).

La forme I a principalement été étudiée chez *O. carboxidovorans*, et elle est très bien caractérisée. Comme mentionné, sa CODH a été cristallisée et le mécanisme de réaction de l'enzyme est bien connu ainsi que son motif d'acides aminés AYXCSFR conservé au site actif. Cette forme code pour une CODH fonctionnelle. La forme II n'a pas été autant étudiée. En fait, la CODH de forme II n'a jamais été cristallisée. Toutefois, son motif d'acides aminés au site actif a été déterminé et diffère de la forme I. Le motif AYRGAGR est retrouvé dans la forme II. L'absence de la cystéine au site actif de l'enzyme de la forme II, soulève plusieurs questions au niveau de la fonctionnalité de cette enzyme. Puisque la cystéine au site actif comprend l'atome de S, un élément majeur pour soutenir la réaction oxydant du CO, le CO pourrait ne pas être le substrat principal de cette forme de l'enzyme. De plus, certains gènes accessoires nécessaires à la maturation de l'enzyme sont absents sur l'opéron des gènes codant pour la CODH de forme II (Cunliffe, 2011, King *et al.*, 2007).

Selon le résultat de PCR ciblant les génotypes *coxL* de forme I et II, les souches *Mesorhizobium* sp. NMB et *Aminobacter* sp. COX, *Xanthobacter* sp. COX et *Stenotrophomonas* sp. LUP possèdent seulement le gène *coxL*-BMS, codant la CODH hypothétique, et ont la capacité de consommer le CO (King, 2003a). Par contre, comme les génomes de ces bactéries n'ont pas été séquencés, il est possible que le gène *coxL* de forme I n'ait pas réussi à être amplifié en utilisant les oligonucléotides disponibles mais qu'il est réellement présent. Par exemple, selon la PCR de King en 2003, *Burkholderia* sp. LUP possèderait seulement la forme II du gène *coxL*. Pourtant, dans l'étude de Weber et King en 2012, reprenant les mêmes amorces utilisées pour *coxL* de forme I en 2003, la séquence du gène *coxL* de forme I de

Burkholderia sp. LUP a été obtenue et séquencée (King, 2003a, Weber *et al.*, 2012). Dans la littérature, l'exemple de la consommation de CO chez la souche *Bradyrhizobium japonicum* USDA 110 qui contiendrait seulement la CODH de forme II est souvent repris. Toutefois, l'utilisation de cet exemple pour prouver l'implication de la CODH de forme II est controversée. Le génome de la souche *Bradyrhizobium japonicum* USDA 110, maintenant connu sous le nom de *Bradyrhizobium diazoefficiens* USDA 110 a été séquencé (Kaneko *et al.*, 2002). Il fut mentionné que cette souche consomme aussi le CO et son activité serait associé à la CODH de forme II (King *et al.*, 2007). Par contre, en retournant aux sources bibliographiques d'origine (référencées par King et al. 2007), c'est la souche *Bradyrhizobium japonicum* USDA 110spc4 qui a été étudiée précisément pour sa capacité d'utiliser le CO comme seul source de carbone et d'énergie (Lorite *et al.*, 2000). Cette souche est un mutant spontané obtenu de *Rhizobium japonicum* 311bl10 après avoir été exposé à 200mg/ml de spectinomycine (Regensburger *et al.*, 1983) dont le génome n'a pas été séquencé afin de déterminer l'effet exacte des mutations sur le génome. L'affirmation que l'activité de consommation du CO de *Bradyrhizobium japonicum* USDA 110 associée à la CODH de forme II par King et al (2007) est donc à prendre avec précaution. Enfin, tous ces éléments réunis (*i.e.* rôle de *coxL* de forme II reposant sur la PCR seulement et extrapolations faites chez *Bradyrhizobium japonicum* USDA 110) sèment le doute sur la capacité de la CODH de forme II de consommer le CO, à elle seule. Cette enzyme, aurait probablement une autre fonction au sein de la cellule. Par conséquent, le gène *coxL* de forme I est le gène à cibler pour l'étude des bactéries aérobies consommant le CO. Le tableau 3 affiche les versions des gènes *coxL* que possèdent certains carboxydovores et carboxydotrophes énumérés dans le texte.

Tableau 3. Présences des séquences *coxL* forme I et forme II chez différents carboxydovores et carboxydotrophes

	Souche	Taxonomie	<i>coxL</i> forme I	<i>coxL</i> forme II	Références
Carboxydovores	<i>Ruegeria pomeroyi</i> DSS-3	α-Protéobactérie	+	+	Cunliffe, 2013
	<i>Aminobacter</i> sp. COX	α-Protéobactérie	?	+	Hardy <i>et al.</i> , 2001; King, 2003a
	<i>Haliangium ochraceum</i>	δ-Protéobactérie	+	-	Quiza <i>et al.</i> , 2014
	<i>Burkholderia xenovorans</i> LB400	β-Protéobactérie	+	+	King 2003a
	<i>Alkalilimnicola ehrlichii</i> MLHE-1	γ-Protéobactérie	+	-	Hoeft <i>et al.</i> , 2007
	<i>Stenotrophomonas</i> sp. LUP	γ-Protéobactérie	-	+	King 2003a
	<i>Thermogemmatispora carboxidovorans</i> PM5	Chloroflexi	+	-	King et King, 2014
Carboxydotrophes	<i>Oligotropha carboxidovorans</i>	α-Protéobactérie	+	-	Meyer et Schlegel, 1983
	<i>Bradyrhizobium japonicum</i> USDA 110	α-Protéobactérie	-	+	King 2003a; Kaneko 2002
	<i>Hydrogenophaga pseudoflava</i>	β-Protéobactérie	+	?	Kiessling et Meyer, 1982
	<i>Pseudomonas thermocarboxydovorans</i>	γ-Protéobactérie	+	?	Black <i>et al.</i> , 1990
	<i>Mycobacterium</i> sp. JC1	Actinobactérie	+	?	Park <i>et al.</i> , 2003

1.4.5 Diversité du gène *coxL* dans l'environnement

Le séquençage de gènes *coxL* issus d'échantillons environnementaux et de souches isolées a démontré la grande diversité des séquences *coxL*, mais aussi sa grande distribution sur l'ensemble du domaine bactérien. Cependant, la diversité du gène *coxL* de forme I est plus répandue que celle du gène *coxL* de forme II. Ce dernier est majoritairement retrouvé chez les *Protéobactéries* comparativement au gène *coxL*-OMP aussi bien présent chez les *Protéobactéries* que dans plusieurs autres différents embranchements comme *Actinobactérie*, *Acidobactérie*, *Chloroflexi*, *Firmicutes*, *Verrucomicrobia* et autres. En général, la distribution du gène *coxL* suit la distribution du gène de l'ARNr 16S par transfert vertical, mais il a aussi des événements de transfert horizontal (latéral) qui ont été observés. Certaines bactéries très proches phylogénétiquement peuvent posséder des gènes *coxL* très distincts autant que deux bactéries plus éloignées phylogénétiquement, peuvent posséder des gènes *coxL* très similaires, même identiques dans le cas de *M. bovis* et *M. smegmatis* (King, 2003b). Lors d'analyses de diversité de séquences dans un sol, les arbres phylogénétiques regroupent les séquences *coxL* connues de souches cultivées comme séquences de références et les séquences des clones. L'assignation taxonomique des séquences environnementales est ensuite attribuée selon son regroupement avec les séquences de référence. Cependant, puisque le transfert latéral du gène *coxL* est possible, ces assignations peuvent être erronées.

Seulement une dizaine d'années se sont écoulées depuis la première étude sur la diversité des séquences *coxL* dans l'environnement (Dunfield *et al.*, 2004). Depuis cette étude, seulement quelques autres études ont tenté d'évaluer la distribution du gène *coxL* dans l'environnement, dans le but de déterminer comment la diversité se rapporte à l'oxydation du CO et quels sont les facteurs environnementaux affectant la distribution des gènes *coxL*. Toutes ont été réalisées à des conditions plus extrêmes, notamment dans des sols de dépôts volcaniques à Hawaii et au Japon avec peu de matières organiques (Dunfield *et al.*, 2004, Gomez-Alvarez *et al.*, 2007) King *et al.*, 2008b, (Weber *et al.*, 2010a), des environnements avec de fortes émissions de CO comme des sources thermales (Yang *et al.*, 2015), des lacs à haute salinité (Yang *et al.*, 2013) et des sols minéraux à haute altitudes (Lynch *et al.*, 2012). Certes, ces endroits sont plus propices à l'utilisation du CO par les bactéries qui sont soumises à des conditions de famine mais la diversité du gène *coxL* dans différents types de sols doit aussi être déterminée afin de mieux comprendre ce puits biologiques de CO. En étudiant leur diversité dans plusieurs sols comportant des propriétés physicochimiques différentes et des activités de

consommation de CO variables il est possible d'observer certains profils génétiques et mieux cibler les bactéries responsable de l'activité de consommation du CO des sols. Des sols de forêt établis sur d'anciens dépôts volcaniques ont tout de même été échantillonnés dans ces études. Les séquences *coxL* de forme I provenant de *Protéobactéries* étaient les plus abondantes dans ces sols (Dunfield *et al.*, 2004, Gomez-Alvarez *et al.*, 2007, King *et al.*, 2008b). King a attribué de l'importance aux *Betaprotéobactéries*, plus précisément les *Burkholderia*, car le sol ayant démontré la plus grande vitesse de consommation de CO possédait 33% de séquences *coxL* de forme I se rapprochant des séquences de *Burkholderia* (Weber *et al.*, 2010a, Weber *et al.*, 2010b, Weber *et al.*, 2012). Toutefois, ces études ne permettent pas de tirer plus de conclusions quant à l'importance relative des différents génotypes *coxL* de forme I pour expliquer l'activité de consommation du CO observée.

1.5 Problématique : Effet du changement global

Les bactéries sont au cœur des cycles biogéochimiques. Différentes bactéries du sol se spécialisent dans la régulation de la concentration des divers GES. En effet, les bactéries du sol consomment en moyenne 15% des émissions globales de CO₂. Elles jouent un rôle important dans le budget global du CO₂ afin de le maintenir à faibles concentrations non toxiques et qui n'affecte pas le cycle du CH₄. Cependant, l'effet du changement global sur ces bactéries du sol est très peu connu et pourrait affecter grandement la chimie atmosphérique. Plus d'importance devraient leur être accordé puisqu'il s'agit d'une réelle problématique. D'ici 2100, il est prévu que la température de la surface globale augmentera entre 1,1-6,4°C (Meehl *et al.*, 2007). De plus, l'usage des terres, les fertilisants et la pollution peuvent affecter les propriétés du sol. Cela pourrait induire l'élimination ou l'introduction de groupes fonctionnels de bactéries, un changement fonctionnel dans une population particulière ou n'apporter aucun changement si cela induit un dééquilibre de populations entre bactéries d'un même groupe fonctionnel (*i.e.* redondance fonctionnelle). Toutefois, cela doit être déterminé.

Le but ultime de ces études serait de construire des modèles mathématiques introduisant les données moléculaires *coxL* pour prédire les changements au niveau de la consommation du CO₂ d'un sol face au changement global. Pour ce faire, il faut d'abord comprendre la physiologie des différentes populations participant dans la consommation du CO₂ atmosphérique et déterminer comment leur abondance et activité peuvent être altérées selon différentes conditions.

1.51 Biomarqueur

Pour mieux comprendre la physiologie des populations de bactéries jouant un rôle comme puits de CO atmosphérique, un biomarqueur doit être identifié afin de suivre cette population dans les sols. Un biomarqueur est un élément biologique appartenant à un système permettant de donner des informations sur ce système fondée sur des liens existant entre le biomarqueur et le système. Le système dont il est question ici est la capacité de consommation de CO atmosphérique du sol. Puisque le gène *coxL* code pour une sous unité de la protéine assurant l'oxydation du CO chez les bactéries aérobies, ce gène semble, à première vue, un bon biomarqueur pour l'étude de la consommation du CO des sols. Puisque les bactéries possédant le gène *coxL* de forme II ne consomment pas le CO, mis à part quelques exceptions douteuses, le gène *coxL* à lui seul ne peut pas être un bon biomarqueur. Il a été proposé que le gène *coxL* de forme I était un bon biomarqueur, du moins, meilleur que *coxL* de forme II (Quiza *et al.*, 2014). Par contre, ce n'est pas toutes les bactéries possédant le gène *coxL* de forme I qui ont la capacité à consommer le CO atmosphérique, soit les bactéries carboxydovores, et seulement quelques-unes ont été identifiées. À la base, quelques bactéries ne sont pas suffisantes pour créer un modèle. C'est pourquoi il est nécessaire d'identifier d'autres carboxydovores afin de cibler les joueurs clés dans la consommation du CO atmosphérique et développer le meilleur biomarqueur pour suivre cette population. Idéalement, chaque génotype *coxL* devrait être associé à une cinétique d'oxydation du CO, ce qui rendrait les modèles de biogéochimie moléculaire extrêmement puissants.

L'abondance des gènes *coxL* de certains groupes de bactéries en particulier, au lieu de cibler l'ensemble des gènes *coxL* de forme I, pourrait être un meilleur biomarqueur en ciblant seulement les candidats responsables de la consommation du CO atmosphérique. Des résultats obtenus précédemment dans notre laboratoire ont démontré un enrichissement de séquences *coxL* de *Deltaprotéobactéries* dans un sol de forêt de feuillus comparativement à un sol d'une monoculture de maïs et d'une plantation de mélèzes. De plus, le sol de forêt de feuillus a obtenu une vitesse d'oxydation du CO plus élevée que les deux autres sols (Quiza *et al.*, 2014). La corrélation de l'enrichissement de séquences *coxL* et de l'oxydation du CO dans le sol de forêt de feuillus, laisse croire qu'un groupe de *Deltaprotéobactéries* pourrait être fortement impliqué dans l'oxydation du CO du sol. Par contre, l'assignation taxonomique de ces séquences *coxL* n'est basée que sur la présence de la séquence du gène *coxL* d'*Haliangium ochraceum*, la séquence la plus similaire à ces séquences *coxL* environnementales retrouvée sur les bases de

données publiques. Il est à déterminer si la présence du gène *coxL* et la capacité de consommer le CO atmosphérique sont communes chez d'autres bactéries inconnues de la classe *Deltaprotéobactéries*, où *H. ochraceum* est la seule bactérie représentative de cette classe de *coxL* atypique. Des séquences se classant dans les *Deltaprotéobactéries* ont aussi été retrouvées dans les autres études de diversité *coxL*. Environ 25% des séquences *coxL* d'une source thermale des Philippines étaient des *Deltaprotéobactéries* et un clone seulement possédait une séquence de *Deltaprotéobactéries* dans un dépôt volcanique d'Hawaii (Weber *et al.*, 2010b, Yang *et al.*, 2015). Bref, la présence de ces séquences inconnues retrouvées aussi dans d'autres études indique qu'elles peuvent être présentes dans divers environnements et pourraient contribuer en partie à la consommation de CO des sols. Enfin, leur identification devrait définitivement être considéré pour mieux comprendre leur écophysiologie.

1.6 Hypothèse

Selon les résultats obtenus précédemment, l'hypothèse est que le groupe de séquences *coxL* de forme I atypiques appartient soit à des bactéries aérobies inconnues de la classe des *Deltaprotéobactéries*, soit *H. ochraceum* aurait reçu ce gène *coxL* par transfert latéral et qu'ainsi ces séquences n'appartiendrait pas aux *Deltaprotéobactéries* mais plutôt à un autre groupe de bactéries inconnues. D'une façon ou d'une autre, ces séquences doivent être identifiées.

1.7 Objectifs

1. Coloniser un sol de forêt stérile par sa flore indigène, et suivre sur une période temporelle les paramètres suivants :
 - L'oxydation du CO dans les microcosmes
 - Les séquences *coxL* par pyroséquençage 454 des amplicons et par qPCR
 - Les séquences du gène codant pour l'ARNr 16S des bactéries par séquençage Illumina et qPCR
2. Établir des corrélations entre le développement de l'oxydation du CO et l'émergence de certains ribotypes et certains génotypes *coxL* à l'aide de réseaux de corrélations.
3. Identifier les bactéries impliquées dans la consommation du CO atmosphérique du sol

Succession of bacterial communities, *coxL* genotypes and carbon monoxide oxidation activity in soil microcosms to identify unknown carboxydovore bacteria dominating deciduous forest soil

ARTICLE

CHAPITRE 2 : ARTICLE

Titre

Succession of bacterial communities, *coxL* genotypes and carbon monoxide oxidation activity in soil microcosms to identify unknown carboxydovore bacteria dominating deciduous forest soil

Auteurs

Isabelle Lalonde and Philippe Constant

Running Title

Identification of unknown carboxydovore bacteria

Journal

Applied and Environmental Microbiology

État actuel

Accepté (numéro de suivi de la soumission : AEM03595-15)

2.1 Contributions à l'article

L'échantillonnage du sol a été réalisé par Liliana Quiza et Philippe Constant. L'analyse des séquences d'ARNr 16S ont été réalisés par Mondher Khdhiri. Toutes les manipulations en laboratoire, l'analyse des séquences *coxL*, les analyses statistiques et la rédaction de l'article ont été réalisées par Isabelle Lalonde. La conception et la supervision du projet ont été réalisées par Philippe Constant. Isabelle Lalonde et Philippe Constant ont rédigé l'article.

2.2 Résumé

La détection du gène *coxL*, codant pour la grande sous-unité de la CO-déshydrogénase catalysant la réaction d'oxydation du CO, est utilisée comme approche standard pour l'étude de l'écologie des bactéries carboxydovores consommant le CO atmosphérique. Une étude récente a révélé que la distribution d'un génotype atypique de séquences *coxL*, *coxL*(groupe x), était corrélé à l'activité d'oxydation du CO. En se basant sur une analyse phylogénique de ces séquences *coxL* avec des séquences *coxL* de références provenant de génomes, ces séquences ont été assignées à des *Deltaprotéobactéries*. Nous cherchons à confirmer ou infirmer l'assignation de ce génotype *coxL*(groupe x) par le suivi de l'activité de consommation de CO et de la succession des communautés microbiennes lors de la colonisation des microcosmes de sols stériles inoculés avec leur microflore indigène. La densité de population estimée de *Deltaprotéobactéries* était trop faible pour tenir compte de l'abondance du génotype *coxL*(groupe x) détecté dans le sol. Un réseau de corrélation a été calculé pour relier le profil des gènes d'ARNr 16S avec la succession des génotypes *coxL* et l'activité de consommation de CO dans le sol. La plupart des génotypes *coxL* pour laquelle le profil de colonisation affichait une covariance avec l'activité de consommation de CO étaient liés à des bactéries appartenant à des *Actinobactéries* et *Alphaprotéobactéries*. L'analyse n'a fourni aucune preuve que le génotype *coxL*(groupe x) appartient aux *Alphaprotéobactéries*. Considérant le profil de la colonisation de bactéries oxydant le CO et le rendement énergétique théorique du taux d'oxydation de CO mesuré dans les microcosmes de sol, nous proposons que les bactéries carboxydovores inconnues possédant le génotype atypique *coxL*(groupe x) sont des stratégistes K mixotrophes. Ces traits métaboliques doivent être pris en compte dans les futures tentatives d'isolation de ces bactéries inconnues.

2.3 Abstract

Surveys of the *coxL* gene, encoding the large subunit of the CO-dehydrogenase, are used as a standard approach in carboxydovore bacteria scavenging atmospheric CO ecological studies. Recent soil surveys unveiled that the distribution of *coxL* sequences encompassing the atypical genotype *coxL*-type I(group x) was correlated to the CO oxidation activity. Based on phylogenetic analysis including the available *coxL* reference genome sequences, this unusual genotype was assigned to an unknown *Deltaproteobacteria* with the *coxL* sequence from *Haliangium ochraceum* being the sole and closest reference sequence. Here we seek to challenge the proposed taxonomic assignation of *coxL*(group x) genotype through the monitoring of CO consumption activity and microbial community successions during the colonization of sterile soil microcosms inoculated with indigenous microorganisms. In our study, we established that the estimated population density of *Deltaproteobacteria* was too small to account for the abundance of *coxL*(group x) genotype detected in soil. Furthermore, we computed a correlation network to relate 16S rRNA gene profiles with the succession of *coxL* genotypes and CO uptake activity in soil. We found that most of the *coxL* genotypes for which the colonization profile displayed covariance with CO uptake activity were related to potential carboxydovore bacteria belonging to *Actinobacteria* and *Alphaproteobacteria*. Our analysis did not provide any evidence that *coxL*(group x) genotypes belonged to *Deltaproteobacteria*. Considering the colonization profile of CO-oxidizing bacteria and the theoretical energy yield of measured CO oxidation rates in soil microcosms, we propose that unknown carboxydovore bacteria harboring the atypical *coxL*(group x) genotype are mixotrophic K-strategists.

2.4 Introduction

Carbon monoxide (CO) is a trace gas in the atmosphere, with a typical mixing ratio of 60–150 ppbv (Seiler, 1974). Combustion of fossil fuels and biomass burning are the main sources of this gas (Park *et al.*, 2015). The hydroxyl radicals ($\text{OH}\cdot$) are responsible for 85% of the global sink of CO while specialized aerobic microorganisms in the soil consume the remaining 15% (Inman *et al.*, 1971b, Monks, 2005). These CO-oxidizing bacteria represent the most uncertain term for the budget of atmospheric CO. The diversity and ecophysiology of this ubiquitous functional group present in the environment are poorly documented and impair the fate prediction of the biological sink and the atmospheric burden CO in response to global change.

CO-oxidizing bacteria possess a CO-dehydrogenase enzyme (CODH) catalyzing the following reaction: $\text{CO} + \text{H}_2\text{O} \rightarrow \text{CO}_2 + 2\text{H}^- + 2\text{e}^+$. The enzyme is composed of a dimer of heterodimers with CoxS, CoxM and CoxL as small, medium and large subunits respectively (Dobbek *et al.*, 2002). The CoxL subunit contains the active site of the enzyme and is used as a molecular marker to assess CO-oxidizer diversity in the environment (Dunfield *et al.*, 2004). Furthermore, the *coxL* gene sequences encompass two different phylogenetic groups: the type I that has been genetically and biochemically characterized, and the putative type II. There is compelling evidence suggesting that type II-CODH is not functional or use another substrate rather than CO (Cunliffe, 2011, Lorite *et al.*, 2000) and thus *coxL* sequences encompassing this group are not considered in an environmental survey of CO-oxidizing bacteria. CO-oxidizing bacteria harboring *coxL*-type I comprise two different physiological groups. They are composed of the carboxydrophic bacteria displaying low affinity for CO and the ability to grow using CO as a sole carbon and energy source, and the carboxydovore bacteria displaying high affinity for CO enabling the capacity to scavenge the trace amounts present in the atmosphere (King *et al.*, 2007). Only few carboxydovore bacteria have been extensively characterized, encompassing *Alphaproteobacteria* (*Bradyrhizobium* spp., *Stappia* spp., *Aminobacter* sp., *Ruegeria pomeroyi* DSS-3)(Cunliffe, 2013, King, 2003a, Weber *et al.*, 2007), *Betaproteobacteria* (*Burkholderia* spp.)(Weber *et al.*, 2012), *Gammaproteobacteria* (*Alkalilimnicola ehrlichii*, *Streptomonas* sp. LUP)(Hoeft *et al.*, 2007, King, 2003a), *Actinobacteria* (*Mycobacterium* spp.) (King, 2003a) and *Chloroflexi* (*Ktedonobacteria*) (King *et al.*, 2014). However, the environmental factors determining their activity and distribution still remains elusive (King *et al.*, 2007).

Previously published environmental surveys of *coxL* sequences have been done in volcanic deposits (Dunfield *et al.*, 2004, King *et al.*, 2008a, Weber *et al.*, 2010a), hot springs (Yang *et al.*, 2015) and aquatic environments (Cunliffe *et al.*, 2008). Even though forest ecosystems exert a dominant role in the global sink of atmospheric CO (Badr *et al.*, 1995), the carboxydovore bacteria thriving in forest soil have received little attention. Recently, a soil survey of *coxL*-type I sequences in deciduous forest have led to the identification of a new *coxL* group, referred as *coxL*(group x) in this paper, for which the distribution was correlated with CO soil uptake activity. A phylogenetic analysis identified the reference *coxL* sequence from *Haliangium ochraceum*, as the closest relative of *coxL*(group x) genotype, suggesting the presence of unknown CO-oxidizing *Deltaproteobacteria* in soil (Quiza *et al.*, 2014). However, this assumption is highly uncertain since the potential lateral transfer of the *coxL* gene and the relatively low identity score (less than 75%) between *coxL*(group x) sequences retrieved from soil and the reference *coxL* sequence in *H. ochraceum*. To address that problematic our study aimed to challenge the hypothetical affiliation of the *coxL*(group x) genotype to *Deltaproteobacteria*. The succession of bacterial communities, *coxL* genotypes, and the CO oxidation activity were monitored in sterile deciduous forest soil microcosms inoculated with the indigenous microorganisms. Also, we verified the hypothesis that the maturation of the CO oxidation activity would be related to the concomitant emergence of *coxL*(group x) genotype and 16S rRNA gene sequences from *Deltaproteobacteria*.

2.5 Material and Methods

2.5.1 Soil sample and incubation

Soil sample was collected from a deciduous forest located about 40 km away from Montreal on the south shore of the St. Lawrence River (45°67'N; 73°32'W). Soil texture was determined with the hydrometer method and particle size distribution assigned soil samples to the silt loam textural class (Elghamry *et al.*, 1962). The same site was visited to investigate the impact of land-use change, soil nutrients and physicochemical properties on CO-oxidizing bacteria (Quiza *et al.*, 2014). Soil sample was homogenized (2-mm sieve) and used for the preparation of the colonization substrates and inocula. For the colonization substrate, soil was sterilized two times separated by a 24-h interval. Sterilization was done in 500 ml Wheaton® glass bottles closed with foam cap. Three bottles were prepared containing 72 g of soil. After the

sterilization procedure, the bottles were kept on the laboratory bench with a foam cap for a 5-day period in order to evacuate CO produced during autoclaving. Three sterile soil microcosms were inoculated with 8 g of homogenized non-sterile soil (*i.e.* indigenous soil microorganisms), corresponding to a 10% (v/v) inoculum. Three bottles were treated with the same manner with 80 g of sterile soil without inoculation for the controls. Sterile foam cap was kept during the incubation at 20°C to allow gaseous exchanges in the microcosms. After 0, 4, 9, 12, 15, 22, 36, and 55 incubation days, sub-samples of 5.5 g were collected in the six microcosms for DNA extraction (0.5 g) and determination of moisture content (5 g) using standard gravimetric method. Soil water content was maintained at approximately 35% during the incubation period. This water level was selected to avoid diffusion limitation of the CO uptake activity, while supporting bacterial metabolism and growth (King *et al.*, 2002, Moxley *et al.*, 1998b). Soil samples dedicated to DNA extraction were frozen in 2 ml microtube containing 700 µl TPM buffer (50 mM Tris-HCl (pH 7), 1.7% polyvinylpolypyrrolidone, 20 mM MgCl₂), 35 µl 20% SDS and 0.5 g glass beads (0.1 mm diameter).

2.5.3 CO oxidation rate in soil

Before measurement of CO oxidation activity, foam cap was replaced by butyl septum cap. CO gas mixture (508 ± 10 ppmv CO GST-Welco, PA, USA) was injected to get approximately 4 ppmv initial concentration in the static headspace. Headspace samples (10 ml) were collected with a Pressure Lok® gastight glass syringe (VICI® Precision Sampling Inc., Baton Rouge, Louisiana, USA) and injected in a gas chromatograph equipped with a reduction gas detector (ta3000R, Ametek Process Instruments®, Delaware, USA). The first-order oxidation rates were calculated by integrating the CO mole fraction time series measured over a 2-hour period, using at least four concentration points for data integration. After each oxidation rate measurement, a foam cap was put back to the bottles for incubation. Reproducibility of the CO analyses was assessed before each set of measurements by repeated analyses of certified CO standard gas (2.05 ± 0.10 ppmv, GST-Welco, PA, USA), and standard deviations were lower than 5%.

2.5.4 DNA extraction and qPCR assays

Genomic DNA was extracted from using a combination of chemical and mechanical cell lysis procedure (Constant *et al.*, 2008). DNA was precipitated with 2 volumes 96% ethanol and polyvinylpolypyrrolidone spin column was used for final purification (Berthelet *et al.*, 1996). Purified DNA extracts were eluted in 200 µl nuclease-free water. Purified DNA was used for qPCR of bacterial 16S rRNA gene, *coxL*-type I, and *coxL*-type I(*Delta*proteobacteria) [represented as *coxL*(group x) in this paper] using previously described procedures (Quiza *et al.*, 2014). The reactions were performed using 5 µl of diluted genomic DNA (1:50).

2.5.5 *coxL* genotyping

Genomic DNA extracted after 4, 22, 36, and 55 incubation days were used for *coxL* pyrosequencing. PCR-amplification of *coxL* genes using the universal assay (Quiza *et al.*, 2014), libraries preparation, and 454 pyrosequencing were performed at McGill University and Génome Québec Innovation Centre. A total of 45,000 raw sequences were obtained from the 12 libraries. The sequences were analyzed using the software mothur (Schloss *et al.*, 2009). Sequences with qaverage lower than 25, with more than 8 consecutive homopolymers and with ambiguous bases were all removed. Average length of sequences was of 336 bases. Sequences were aligned against a *coxL* sequences database previously built (Quiza *et al.*, 2014). An arbitrary identity cut-off of 0.04 was used for the clusterization of nucleotide sequences into operational taxonomic units (OTU) representing *coxL* sequences belonging to the same bacterial species (Quiza *et al.*, 2014). Libraries were standardized to the sequencing effort of the smallest *coxL* gene library (*i.e.* 641 sequences) to avoid biases in comparative analyses introduced by the sampling depth. At this stage, hypothetical *coxL*-type II OTUs identified by the canonical amino acids signature of their active site (AYRGAGR) were removed from the dataset. The resulting OTU table was used for rarefaction and diversity analyses. OTUs with less than 7 sequences (0.1% of total sequences) were eliminated, and 136 OTUs were left for phylogenetic and correlation network analysis. Nucleic acid sequences were imported in the software MEGA (Tamura *et al.*, 2007) for phylogenetic analysis using the reference database and procedure previously described (Quiza *et al.*, 2014) in which sequences were added. Raw sequences were deposited to the Sequence Read Archive of the National Center for Biotechnology Information under the Bioproject PRJNA299500.

2.5.6 Bacterial 16S rRNA gene profiling

Genomic DNA extracted after 4, 22, 36, and 55 incubation days were used for 16S rRNA gene profiling. PCR-amplification of the V6-V8 regions of 16S rRNA, libraries preparation and Illumina MiSeq 250 bp paired-ends sequencing reactions were performed by the technical staff of McGill University and Génome Québec Innovation Centre. The two paired-end reads (forward and reverse) were merged using Flash (Magoč *et al.*, 2011) with a minimum overlap length between the two reads of 20 bases and a maximum overlap of 250. The maximum allowed ratio between the number of mismatched base pairs and the overlap length was set to 0.3. Reads were truncated to a uniform length of 420, corresponding to the length of more than 97.5% of the total reads. Reads with low-quality score were removed using a maximum expected error value of 2.0 (Edgar *et al.*, 2011). The remaining reads (4,058,447) were processed using the software QIIME version 1.8.0 (Caporaso *et al.*, 2010). Reads were dereplicated and then sorted by abundance. Unique reads (1,214,627) were then clustered at 97% identity using the USEARCH (Edgar *et al.*, 2011) as the OTU picking method with a minimum of two sequences for the OTU to be considered. The resulting OTUs were extra checked for chimeras using the gold standard reference database and those that were flagged as chimeric (10,339) were removed. The remaining 2,060 non chimeric OTUs were assigned using the assign taxonomy command (QIIME) run against the Greengenes database (version gg_13_8) (McDonald *et al.*, 2012). Reads were rarefied to the sequencing effort of the smallest 16S rRNA gene library (206,201 reads) to avoid biases in comparative analysis introduced by the sampling depth, giving a final number of 613 OTUs. Raw sequences were deposited to the Sequence Read Archive of the National Center for Biotechnology Information under the Bioproject PRJNA299500.

2.5.7 Statistical analysis

Statistical analyses were all performed using the software R (Team, 2008). The package “Hmisc” was used to measure Spearman correlation between *coxL*:16S rRNA gene ratio estimated by qPCR and soil CO uptake (Harrell, 2015). The package “Vegan” was used to compute alpha diversity indexes and beta-diversity with homogeneity multivariate dispersion (Anderson *et al.*, 2006, Oksanen *et al.*, 2015). The package “Weighted Gene Co-Expression Network Analysis” (WGCNA) was used to identify modules of highly correlated 16S rRNA OTUs (Langfelder *et al.*, 2008). Data matrix used for this analysis was expressed in 16S rRNA gene

copies per gram of soil_(dw). For this purpose, the relative frequency of individual 16S rRNA gene OTU as determined through sequencing was multiplied by the absolute copy number of 16S rRNA gene as determined by qPCR. A soft-power of 12 was used to raise the Spearman correlation matrix into an adjacency matrix. This matrix was then converted into a topological overlap (TO) matrix by the TOM similarity algorithm and genes were hierarchically clustered based on TO similarity. Modules comprising less than 10 OTUs were not considered and modules having eigengene dissimilarity <0.45 were merged. Spearman correlations were computed between modules eigengene and CO uptake rate and the absolute abundance of *coxL* OTUs. The absolute abundance of *coxL* OTUs (expressed in copies per gram of soil_(dw)) was computed by multiplying the relative abundance of individual OTU obtained from pyrosequencing profiles by the absolute abundance of *coxL* genes determined by qPCR to identify the module the most correlated with CO uptake and identify related *coxL* OTUs.

2.6 Results

2.6.1 Maturation of soil CO uptake activity

The maturation of CO uptake activity was detected after a lag phase of 10 days (Fig. 1). We observed that the activity then increased continuously, reaching $3466 \pm 800 \text{ pmol g}_{(\text{dw})}^{-1} \text{ h}^{-1}$ after 55 days. This activity is comparable to the oxidation rate of $3243 \text{ pmol g}_{(\text{dw})}^{-1} \text{ h}^{-1}$ measured in the native soil (Quiza *et al.*, 2014), indicating that activity was restored at the end of the incubation period. Furthermore, we found that no significant CO uptake activity was observed in the control experiment comprising non-inoculated sterile soil (Fig. 1).

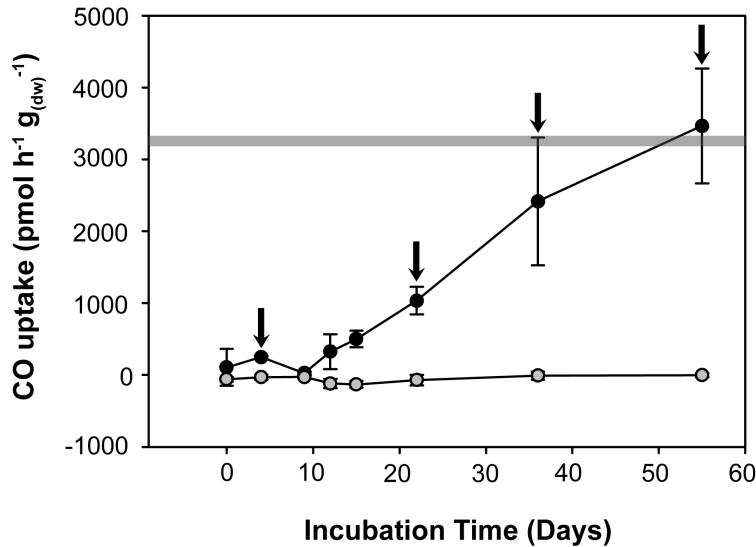


Figure 1. : Time series of the CO oxidation activity measured in inoculated (black dots; ●) and sterile (grey dots; ○) soil microcosms. Averages are presented with standard deviations derived from three independent replicates. The grey line represents the CO oxidation rate measured in native soil. The black arrows indicate selected soil samples for high-throughput sequencing of PCR-amplified *coxL* and 16S rRNA genes.

2.6.2 Abundance of bacterial 16S rRNA and *coxL* gene

Using qPCR we determined that the abundance of 16S rRNA gene increased from $3.3 \pm 0.5 \times 10^9$ copies $\text{g}_{(\text{dw})}^{-1}$ at the beginning of the experiment to $4.4 \pm 1.4 \times 10^{10}$ copies $\text{g}_{(\text{dw})}^{-1}$ after 4 incubation days (Fig. S1). No significant increase was observed between 22 and 55 incubation days, with $2.6 \pm 0.5 \times 10^{11}$ copies $\text{g}_{(\text{dw})}^{-1}$ at the end of the incubation period. Quantification of the *coxL* gene followed the same trend than 16S rRNA gene, while *coxL*(group x) increased exponentially after a lag phase of 15 days (Fig. S1). The relative abundance of *coxL* genotypes in soil was estimated by computing the *coxL* to 16S rRNA gene ratio (Fig. 2). We established that the relative abundance of *coxL* increased by $5.0 \pm 0.94\%$ after 4 days and then decreased exponentially, reaching $2.2 \pm 0.37\%$ at the end of the incubation. On the other hand, the relative abundance of *coxL*(group x) increased exponentially from $0.05 \pm 0.01\%$ at day 15 to $0.7 \pm 0.09\%$ after 55 incubation days. The relative abundance of *coxL*(group x) was positively correlated with CO uptake rate (Spearman, $p = <0.001$), while neither the abundance of *coxL* nor 16S rRNA genes showed a significant relation with CO oxidation activity.

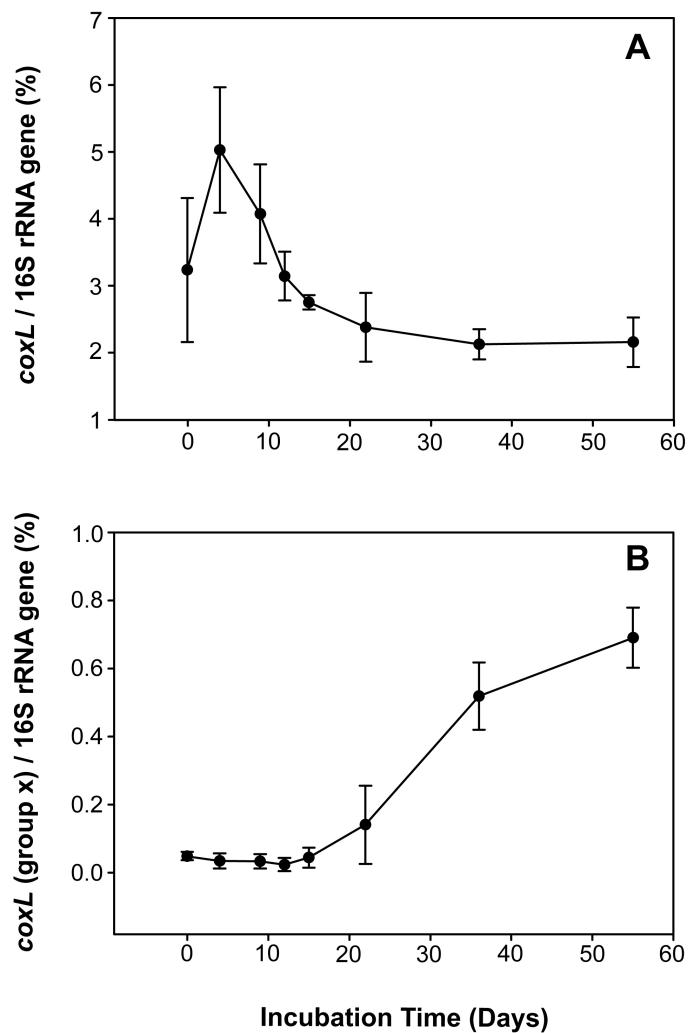


Figure 2. : Time series of the relative abundance of (A) *coxL* and (B) *coxL*(group x) genotypes in soil microcosms as estimated by qPCR.

2.6.3 Bacterial 16S rRNA genotypes

Sequencing effort was sufficient to cover most of the diversity of 16S rRNA genes, as assessed by rarefaction curves that reached a saturation plateau (data not shown). In total, 613 OTUs were identified. Interestingly, the incubation of soil microcosms resulted in an overall decrease in species richness and species evenness (Table 1), indicating that a fraction of the bacteria present in the native soil inoculum demonstrated the ability to colonize sterilized soil with some disproportionate single-species contribution to diversity. The beta diversity analysis showed that the structure of microbial communities was significantly more heterogeneous at the beginning of the incubation than observed in the last three subsamples, indicating reproducible enrichment of the subset of the initial bacterial community in replicated microcosms (Fig. S2a). At the end of the incubation period, we found that soil microbial communities were dominated by the phyla *Actinobacteria* ($65 \pm 4\%$), *Acidobacteria* ($17 \pm 1\%$), and *Proteobacteria* ($16 \pm 3\%$). *Proteobacteria* were dominated by *Alphaproteobacteria* ($16 \pm 2.6\%$), followed by *Betaproteobacteria* ($0.44 \pm 0.43\%$), *Gammaproteobacteria* ($0.12 \pm 0.12\%$), and *Deltaproteobacteria* ($0.008 \pm 0.002\%$). Considering the abundance of 16S rRNA gene determined by qPCR and assuming three rRNA gene copies per genome (Vetrovsky *et al.*, 2013), we evaluated that the population density of *Deltaproteobacteria* reached 7.0×10^4 cell g_(dw)⁻¹ in the soil microcosm. This is well below the absolute abundance of *coxL*(group x) sequences determined by qPCR (1.8×10^9 copies g_(dw)⁻¹).

Table 1. Specific richness and evenness of *coxL* and 16S rRNA gene OTUs in soil microcosms.

Incubation time (days)	16S rRNA gene		<i>coxL</i>	
	Shannon	Simpson	Shannon	Simpson
4	4.57 ± 0.10	0.980 ± 0.001	4.43 ± 0.14	0.978 ± 0.004
22	3.45 ± 0.08	0.920 ± 0.007	4.07 ± 0.21	0.953 ± 0.004
36	3.66 ± 0.04	0.930 ± 0.006	3.87 ± 0.23	0.924 ± 0.018
55	3.84 ± 0.12	0.937 ± 0.006	3.91 ± 0.23	0.930 ± 0.018

2.6.4 *coxL* genotypes

Rarefaction analysis showed that the sequencing effort was sufficient to cover *coxL* diversity (data not shown), with 136 OTUs identified. As observed in the 16S rRNA gene profile, *coxL* richness and evenness decreased during soil colonization (Table 1). Homogenisation of *coxL* genotype profiles in replicated microcosms was slower than 16S rRNA genes, and exhibited more heterogeneity in soil subsamples collected after 4 and 22 days than observed in the replicated microcosms sampled at days 36 and 55 (Fig. S2b). The *coxL* OTUs were assigned to one of the three reference genotypes in a phylogenetic analysis (Fig. 3). At the end of the incubation, $53 \pm 3\%$ of *coxL* sequences were assigned to *coxL*(group x) while *Actinobacteria* and *Alphaproteobacteria* represented $14 \pm 3\%$ and $31 \pm 3\%$ of retrieved *coxL* sequences. We noticed that few OTUs dominated the three genotypes. One OTU closely related to *Bradyrhizobium japonicum* USDA (*Alphaproteobacteria*; OTU 1) represented 70% of total sequences encompassing *Alphaproteobacteria* genotype and 18% of total *coxL* sequences (Fig. S3a). The most abundant OTU belonging to *Actinobacteria* was closely related to *Arthrobacter* sp. FB24 (OTU 4) and represented 24% of the sequences comprising this genotype (Fig. S3b). OTU 2 and OTU 3 represented 21 and 11% of total sequences encompassing *coxL*(group x) genotype, respectively (Fig. S3c). The only reference bacterium present in the cluster *coxL*(group x) defined by the phylogenetic analysis is *H. ochraceum*. The percentage of identity between this reference sequence and the closest OTU was 80%. All the other OTUs assigned to *coxL*(group x) were more distantly related to *H. ochraceum*, with an identity score between 46 and 59%.

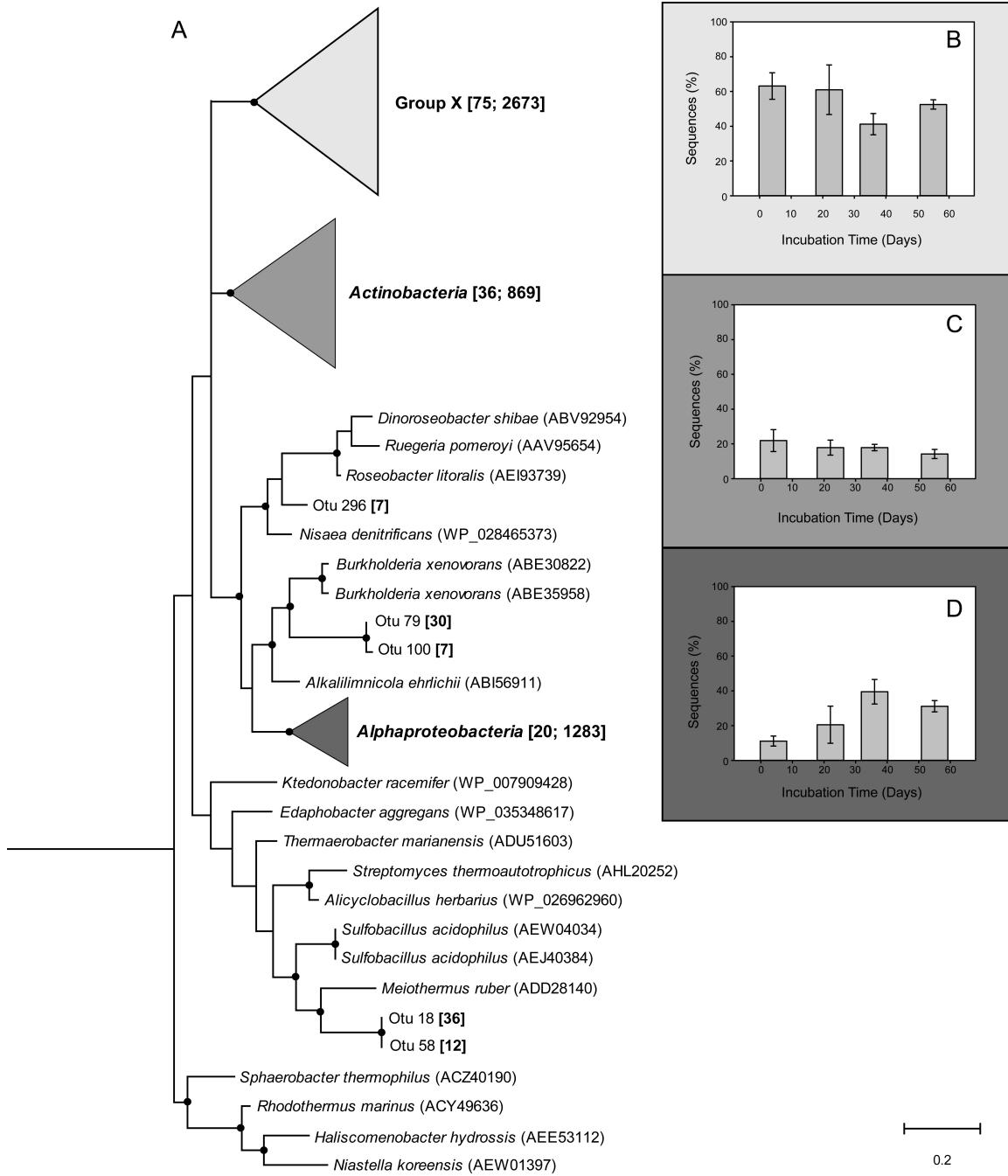


Figure 3. : Maximum-likelihood phylogenetic analysis of *coxL*-inferred amino acid sequences of 136 OTUs and reference sequences from NCBI (147 residues). The three main subclass of sequences belonging to *coxL* are represented in the three clusters. The number of sequences of each OTU or a cluster is indicated in the square brackets [OTU; number of sequences]. The percentage of replicated trees in which the associated *coxL* sequences clustered together in the bootstrap test (1000 replicates) are shown for nodes supported by $\geq 50\%$ of the replicates. (B) Time series of *coxL* sequences relative abundance in *coxL*(group x) (C) *coxL*(*Actinobacteria*) and (D) *coxL*(*Alphaproteobacteria*) clusters.

2.6.5 Correlation network analysis

The CO oxidation activity, *coxL* genotypes, and 16S rRNA gene profile databases were combined to correlate soil genetic profile with the CO oxidation activity and test the hypothesis that *coxL*(group x) sequences positively correlated to activity belonged to *Deltaproteobacteria*. For this purpose, a correlation network was first constructed using the time series of 16S rRNA profiles. This analysis led to a classification of the OTUs into five modules according to the covariance of their soil colonization profile (Fig. 4). The first module (Module 1) was dominated with OTUs encompassing *Actinobacteria* (62%), *Acidobacteria* (22%), and *Proteobacteria* (15%). The second module (Module 2) was mostly composed of *Firmicutes* OTUs (90%). Module 3 was the most taxonomically diverse, comprising 12 different phyla. Module 4 was mainly represented by OTUs affiliated to *Actinobacteria* (60%), *Proteobacteria* (26%), and *Actinobacteria* (9%). Finally, module 5 was composed mostly of *Proteobacteria* (35%), *Actinobacteria* (26%), and *Chlamydia* (16%). Correlation of modules eigengene with the CO oxidation rates measured during the incubation of soil microcosms unveiled that relative abundance of OTUs encompassing module 1 was significantly correlated to the maturation of CO uptake activity in soil microcosms (Spearman, $p = <0.001$). Thus, the OTUs constituting module 1 were from presumptive carboxydovore bacteria and other bacteria displaying a similar colonization profile but incapable of CO oxidation activity.

The 16S rRNA gene OTUs assigned as *Acidobacteriaceae*, *Actinospicaceae*, *Microbacteriaceae*, *Bradyrhizobiaceae*, and *Acetobacteraceae* dominated the composition of module 1 with 22, 36, 21, 7 and 3% of total sequences in the module respectively (Table 2). Most of the taxonomic groups (family level) comprising module 1 are represented by at least reference genome or bacterium possessing *coxL* sequence (Table 2). Even though no *coxL* genes were found in reference genome from *Actinospicaceae*, *Microbacteriaceae*, and *Acetobacteraceae*, the colonization profile of OTUs assigned to these taxonomic groups was significantly correlated with CO oxidation activity (Table 2).

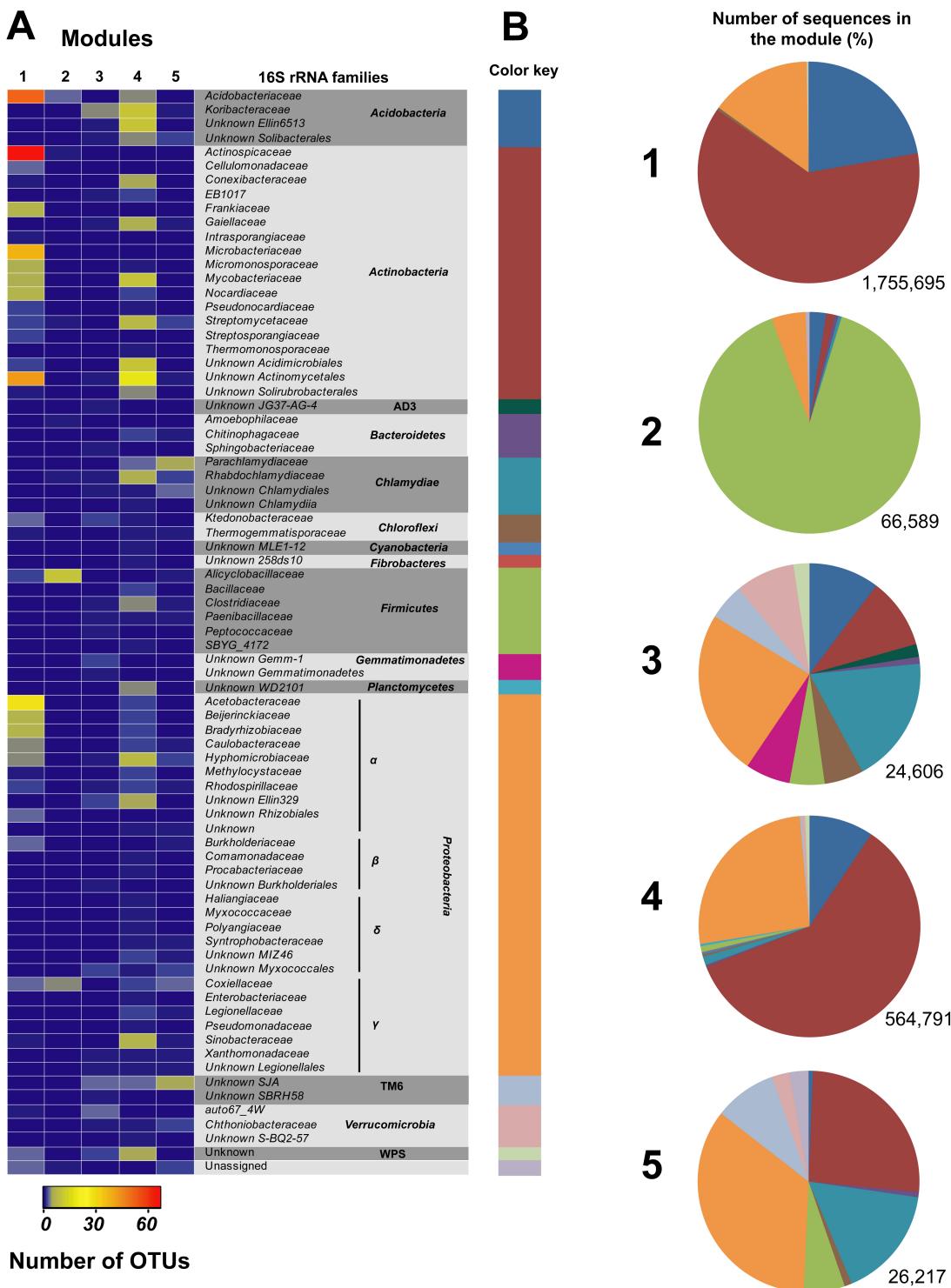


Figure 4. : Taxonomic composition of the five modules defined by the covariance of 16S rRNA OTUs time series in the soil. (A) The heatmap shows the number of OTUs representing each family detected in the modules. (B) The OTUs assigned to the same bacterial phyla were clustered together, and the number of sequences comprised in each group was computed to report their relative abundance in the five modules. The color key used to identify each bacterial phylum in the pie charts is defined along the side of the heatmap.

Table 2. Composition of module 1 comprising 16S rRNA OTUs displaying colonization profile related to the maturation of CO oxidation activity in soil. The relative abundance of sequences encompassing each family and the presence or the absence of *coxL* gene in reference genome of each taxonomic group are presented. Spearman correlations between the absolute abundance of OTUs grouped at the family level and CO oxidation rate are presented. The symbols *, ** and *** denote $p = <0.05$, $p = <0.01$ and $p = <0.001$ correlation significance levels, respectively. NA = non-applicable.

Families	Abundance in module 1 (%)	Correlation	Presence of <i>coxL</i>	Example (accession number)
<i>Acidobacteriaceae</i>	22.23	0.21*	+	WP_035348617
Unknown <i>Acidimicrobiales</i>	0.02	0.37	+	WP_035348617
<i>Actinospicaceae</i>	35.44	0.71**	-	NA
<i>Cellulomonadaceae</i>	0.08	0.88***	-	NA
<i>Conexibacteraceae</i>	0.03	0.83***	+	ADB51446
<i>Frankiaceae</i>	0.42	0.29	-	NA
<i>Intrasporangiaceae</i>	0.01	0.09	-	NA
<i>Microbacteriaceae</i>	21.12	0.82***	-	NA
<i>Micromonosporaceae</i>	0.13	0.68*	+	WP_033364885
<i>Mycobacteriaceae</i>	0.97	-0.22	+	AFP37210
<i>Nocardiaceae</i>	0.31	0.22	+	WP_031936578
<i>Pseudonocardiaeae</i>	0.09	0.78**	+	WP_028847129
<i>Streptomycetaceae</i>	0.19	0.05	+	WP_033825254
<i>Streptosporangiaceae</i>	0.46	0.59*	+	WP_020541099
Unknown <i>Actinomycetales</i>	2.70	0.61	+	WP_033364885
Unknown <i>Solirubrobacterales</i>	0.33	-0.19	+	ADB51446
<i>Ktedonobacteraceae</i>	0.29	-0.33	+	WP_007909428
<i>Thermogemmatisporaceae</i>	0.10	0.54	+	WP_052889641
<i>Alicyclobacillaceae</i>	0.03	0.28	+	WP_026962960
<i>Acetobacteraceae</i>	2.76	0.76**	-	NA
<i>Beijerinckiaceae</i>	2.39	0.45	+	WP_020174914
<i>Bradyrhizobiaceae</i>	7.43	0.79**	+	WP_014491996
<i>Caulobacteraceae</i>	1.27	0.36	-	NA
<i>Hyphomicrobiaceae</i>	0.13	0.85***	-	NA
<i>Methylocystaceae</i>	0.03	-0.63*	-	NA
<i>Rhodospirillaceae</i>	0.06	0.42	+	WP_028465373
Unknown <i>Rhizobiales</i>	0.07	0.93***	+	WP_014491996
<i>Burkholderiaceae</i>	0.50	0.60*	+	ABE35958
<i>Coxiellaceae</i>	0.05	0.22	-	NA
<i>Sinobacteraceae</i>	0.04	0.59*	-	NA
<i>auto67 4W</i>	0.03	0.31	NA	NA
Unknown <i>WPS-2</i>	0.20	0.84***	NA	NA
Unassigned	0.06	0.80**	NA	NA

* Found by sequence identity with Blastp (Altschul *et al.*, 1990) using the *coxL* sequence in *Burkholderia xenovorans* LB400 (Accession number ABE35958).

Correlation of modules eigengene with time series of *coxL* genotypes led to the identification of 42 *coxL* OTUs whose colonization profile were significantly correlated with those of 16S rRNA gene OTUs comprising module 1 (Fig. 5). Among these OTUs, 27 encompassed *coxL*(group x), 5 belonged to *Actinobacteria*, 8 comprised *Alphaproteobacteria*, and 2 were outside these groups and affiliated to a *coxL* sequence from *Deinococcus-Thermus*. Convergence of the taxonomic affiliation and the relative abundance of some rRNA gene OTUs in module 1 and *coxL* OTUs (*i.e.* *Bradyrhizobium* spp.) supported the phylogenetic analysis of *coxL* genes (Table S1). No *Deltaproteobacteria* was present in module 1, impairing the assignation of *coxL*(group x) genotype to this taxonomic group. At the end of the incubation, the absolute abundance of the 27 *coxL*(group x) OTUs correlated with CO uptake rate was $2.4 \pm 0.34 \times 10^9$ bacteria g_(dw)⁻¹, as estimated by multiplying the relative abundance of the *coxL* OTUs obtained from pyrosequencing profiles by the absolute abundance of *coxL*(group x) genes determined by qPCR. Considering the specific richness and the abundance of OTUs encompassing *coxL*(group x) genotype, the module 1 was searched for potential 16S rRNA gene OTUs candidates to infer a taxonomic affiliation to unknown presumptive carboxydovore bacteria. 16S rRNA OTUs belonging to *Actinospicaceae*, and *Microbacteriaceae* were the best *coxL*(group x) candidates as predicted using the correlation network analysis. These two bacterial families were the sole candidates for which the absolute abundance and species richness of 16S rRNA OTUs as congruent with those of *coxL*(group x) OTUs in module 1 (Table S1).

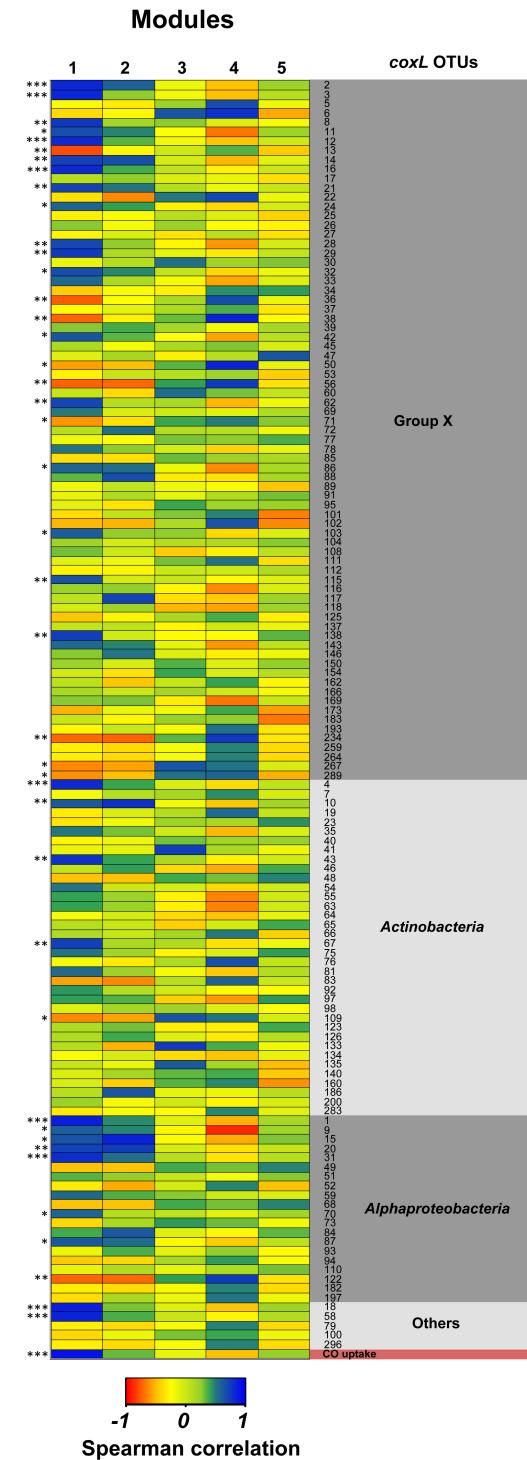


Figure 5. : Spearman correlations between the *coxL* OTUs with the eigengene of the five modules comprising 16S rRNA OTUs displaying similar soil colonization profile. Only the eigengene of module 1 was correlated with CO uptake rate measured in soil microcosms. Significant correlations between the time series of individual *coxL* OTU and the eigengene of module 1 are marked with asterisks along the side of the heatmap ($*p = <0.05$, $**p = <0.01$ and $***p = <0.001$).

2.7 Discussion

Enrichment of microorganisms utilizing atmospheric trace gas is challenging because the energy potential of these substrates usually supports mixotrophic survival metabolism, where reduced inorganic and organic compounds are simultaneously used as energy source for maintenance or survival. Furthermore, increase of energy potential through increased gas concentration is not recommended because it will result in the enrichment of microbes that are incapable of scavenging atmospheric trace gas owing to their low substrate affinity. Most high affinity CO-oxidizing bacteria were identified through screening of isolates obtained from environmental samples or public collections after detection on *coxL* in sequenced genome. One notable exception was the isolation of the first carboxydovore bacterium *Aminobacter* sp. *cox1* after long-term soil exposure to 40-400 ppmv CO (Hardy *et al.*, 2001). Recent investigations into the diversity of soil carboxydovore bacteria led to the identification of an atypical *coxL*(group x) genotype for which the distribution was significantly related to CO uptake rate in soil (Quiza *et al.*, 2014). Before investment of cultivation efforts to isolate these unknown carboxydovore bacteria, we sought to challenge the hypothetical taxonomic affiliation of atypical *coxL* genotype to *Delta proteobacteria* through the monitoring of microbial succession and maturation of CO uptake activity in soil microcosms.

Successions of 16S rRNA and *coxL* genes analyzed by qPCR (Fig. 2) unveiled that *coxL* carboxydovore bacteria generally are opportunistic *r*-strategist bacteria using labile nutrients in soil. They also increase their relative abundance early in the incubation period. Our data showed that their enrichment was transient, suggesting that their populations were not maintained upon soil nutrient depletion. Interestingly, the CO-oxidizers represented by *coxL*(group x) displayed *K*-strategists colonization profile, with a slow growth of their population sustained by the efficient use of limited resources in soil. At the end of the incubation, total *coxL* and *coxL*(group x) genotypes represented 2.2% and 0.7% of the total bacterial population, respectively. It is believed that atmospheric CO can support mixotrophic metabolism of these carboxydovore bacteria (King *et al.*, 2007). To address that question, we combined our qPCR time series with thermodynamic models. The theoretical population size of high-affinity CO-oxidizing bacteria in soil for which CO oxidation fully supply maintenance energy requirements was estimated based on the theoretical maintenance energy consumption rate and the free energy of atmospheric CO oxidation, using the following model formulated by Conrad (Conrad, 1999):

$$N = [1.4 \times 10^{14} (-\Delta G) d] / mE,$$

where N is the theoretical population size of CO-oxidizing bacteria in soil (expressed as the number of cells per g of soil_(dw)), 1.4×10^{14} is a constant expressing the density of bacterial cells containing 1 mol carbon (number of cells per mol of C biomass), ΔG is the Gibbs free energy of atmospheric CO oxidation (-235 kJ per mol of CO), d is the measured CO oxidation rate (mol of CO per g of soil_(dw)), and mE (4.5 kJ per mol of C biomass) is the energy maintenance requirement of the population (Tijhuis *et al.*, 1993). This model estimated that measured CO uptake activity would provide sufficient energy to support the survival of a maximal population 2.5×10^7 carboxydovore bacteria g_(dw)⁻¹. According to our qPCR results, the abundance of total *coxL* and *coxL*(group x) sequences reached $5.5 (\pm 0.97) \times 10^9$ and $1.8 (\pm 0.20) \times 10^9$ copies g_(dw) in soil at the end of the incubation, respectively. Taken together, these observations imply that oligotrophic cultivation media comprising both organic carbon and CO would be necessary to isolate unknown carboxydovore bacteria harboring atypical *coxL*(group x) genotype that are expected mixotrophic K-strategists.

Although microbial succession was shaped by stochasticity in colonial growth dynamic of individual cells (Alonso *et al.*, 2014) and alteration of the chemical structure and bioavailability of soil organic carbon due to autoclaving (Berns *et al.*, 2008), the taxonomic affiliation of *coxL* genotypes inferred from phylogenetic analysis resulted in similar distributions when compared with the observations made using native deciduous soil (Quiza *et al.*, 2014). *Actinobacteria*, *Alphaproteobacteria*, and *coxL*(group x) were the dominating *coxL* genotypes, representing $31 \pm 3\%$, $14 \pm 3\%$ and $53 \pm 3\%$ of the total sequences, respectively. Very few studies have reported the taxonomic composition of carboxydovore bacteria in soil. Soil survey along a succession of volcanic deposits unveiled an enrichment of *coxL* sequences closely related to *Burkholderia* isolates known to consume atmospheric CO (King, 2003a) as a function of vegetation regeneration and CO uptake activity (Weber *et al.*, 2010a, Weber *et al.*, 2010b). In this study two *coxL* OTUs (OTU 79 and 100) related to *Burkholderia xenovorans* LB400 *coxL* sequence were detected but their temporal distribution profile was not related to the maturation of CO uptake activity in soil. It has been reported that some species of mycobacterium from *Mycobacteriaceae* family in the *Actinobacteria* phylum are capable to consume atmospheric CO (Kim *et al.*, 2012, King, 2003b). In this study, 14 *coxL* OTUs closely related to *Mycobacterium* sequences were detected, but none displayed significant correlation with CO oxidation rate in soil. The high abundance of *coxL* sequences closely related to *Bradyrhizobium* spp. and the correlation of their distribution profile with CO uptake activity suggest an implication of these *Alphaproteobacteria* in the activity measured in the microcosms. Although they were not dominant in previous soil surveys, their capacity to consume atmospheric CO have already been reported using

environmental isolates (King, 2003a). Sequences encompassing *coxL*(group x) were abundant, with *H. ochraceum* as the only representative. CO uptake activity have been confirmed in this bacterium (Quiza *et al.*, 2014) but no *coxL* sequence have been noticed in sequenced genome of the other aerobic *Deltaproteobacteria* encompassing the orders *Myxococcales* and *Bdellovibrionales*. Combining qPCR and high-throughput sequencing of 16S rRNA and *coxL* genes led to the conclusion that it is very unlikely that *Deltaproteobacteria* harbor *coxL*(group x) genotype. If unknown *Deltaproteobacteria* harbored the atypical *coxL* genotype, then a strong primer bias have impaired their detection or considerably have underestimated their abundance.

Network correlation analysis led to the identification of taxonomic groups of bacteria for which the distribution was significantly correlated with the distribution of OTUs encompassing *coxL*(group x). Although the analysis does not imply any causal effect, it supported the suggestion that unknown carboxydovore bacteria are mixotrophic *K*-strategists, which is critical for the establishment of the best isolation strategy. Among the taxonomic groups detected in module 1 for which the eigengene was correlated to CO uptake rate, bacteria encompassing *Actinospicaceae* and *Microbacteriaceae* families were the most probable taxonomic assignation unknown carboxydovore bacteria harboring *coxL*(group x) genotype. This taxonomic inference will need validation through the isolation efforts since no representative of these two families was reported as CO-oxidizing bacteria. Interestingly, microorganisms that belong to these taxonomic groups are generally *K*-strategists, with some of them having the ability to use recalcitrant carbon such as cellulose (Lynd *et al.*, 2002). Furthermore, soil amendments experiments unveiled that cellulose stimulates CO uptake activity in soil while glucose exerts the inverse effect (Isabelle Lalonde, unpublished). Bioprospection efforts are currently invested by the authors to identify and characterize unknown carboxydovore bacteria harboring *coxL*(group x) genotype. Alternatively, single-cell sequencing (Ottesen *et al.*, 2006) could be envisaged if cultivation efforts are unsuccessful.

2.8 Acknowledgments

This work was supported by a Natural Sciences and Engineering Research Council of Canada Discovery grant to PC. IL is grateful to the Fondation Universitaire Armand-Frappier INRS for her M.Sc. scholarship. The authors are grateful to the personnel staff of McGill University and Génome Québec Innovation Centre for preparation of 16S rRNA gene libraries and sequencing services. Mondher Khdhiri is acknowledged for his technical assistance in quality control, classification and taxonomic affiliation of 16S rRNA gene sequences.

CONCLUSION ET PERSPECTIVES

La découverte de nouveaux carboxydovores pour une meilleure compréhension du puits biologique du CO, un important GES indirect, est essentielle. Une approche innovatrice a été utilisée dans le but d'identifier des bactéries à haute affinité responsable de la consommation atmosphérique de CO. Le séquençage du gène codant pour l'ARNr 16S des bactéries et du gène fonctionnel *coxL* n'est rien de nouveau, mais l'utilisation du séquençage en parallèle avec le suivi de l'activité de consommation du CO sur une série temporelle de la colonisation d'un sol stérile spécifique pour identifier les génotypes *coxL* est une approche qui se distingue des autres techniques. Par exemple, le séquençage du gène *coxL* de divers sols et l'association de l'abondance des gènes avec la consommation de CO sont utilisés afin d'identifier des groupes spécifiques de bactéries lorsque l'identification de bactéries est encore au niveau exploratoire.

Toutefois, les résultats obtenus avec cette approche permettent de ressortir certaines bactéries comprenant des souches inconnues potentiellement carboxydovores, mais les assignations proposées devront être vérifiée expérimentalement. Premièrement, les résultats démontrent que les OTU *coxL* atypiques corrélés à l'activité de consommation de CO ne proviennent probablement pas de *Deltaprotéobactéries* tel que proposé dans une étude passée (Quiza *et al.*, 2014). Avec seulement $0.008 \pm 0.002\%$ de ribotypes de *Deltaprotéobactéries* dont l'abondance au cours de l'incubation ne corrèlent pas avec l'activité de consommation de CO, et $53 \pm 3\%$ d'OTU *coxL* du groupe x, il est très peu probable que ces séquences proviennent de *Deltaprotéobactéries*. La grande majorité des séquences des OTU corrélés à l'activité de consommation de CO sont assignés à différentes familles de l'embranchement des *Actinobactéries*, suivies des *Acidobactéries* et de *Protéobactéries*. Plusieurs des OTU *coxL* qui ont ressortis au cours de l'étude peuvent facilement être associés à des bactéries connues avec la capacité de consommer le CO et possédant le gène *coxL*. Il est plus logique de s'attarder sur les membres des familles dont aucune séquence *coxL* n'a encore été identifiée, qui pourrait plus probablement représenter les séquences *coxL* inconnues corrélé à la consommation de CO comme les *Microbacteriaceae*, *Cellulomonadaceae* et *Hyphomicrobiaceae*.

Le SIP (stable isotope probing) est une bonne technique pour identifier les bactéries utilisant un substrat précis marqué au ^{13}C , donc marquant l'ADN au ^{13}C des populations actives ayant consommé ce substrat. Dans cette technique, l'ADN plus lourd des bactéries actives peut être séparé de l'ADN non marqué par centrifugation en gradient de densité. Par contre, le SIP

ne peut malheureusement pas être envisagé pour identifier les bactéries carboxydovores pour deux principales raisons : (1) le pourcentage de CO₂ provenant de l'oxydation du CO qui est assimilé dans la cellule est très faible. Les bactéries ne seraient que faiblement marquées et la séparation de l'ADN des bactéries marquées au ¹³C des autres bactéries non marquées serait impossible. (2) Le CO est retrouvé qu'à l'état de traces dans l'atmosphère donc le marquage de l'ADN à partir d'une source de carbone si peu abondante n'est pas suffisante. Bref, ces deux points réunis éliminent le SIP comme stratégie de choix pour identifier les bactéries possédant le génotype *coxL* atypique.

Plusieurs perspectives du projet sont envisagées dans le but de mieux cibler les bactéries possédant le génotype *coxL* atypique. L'isolement de ces bactéries pourrait être tenté puisqu'on sait maintenant que les bactéries à cibler seraient des stratégistes K. Des milieux de cultures avec des sources de carbones complexes, comme de la cellulose, et du CO à faible concentration pourraient permettre de les enrichir et les isoler. Des tests préliminaires dans un sol avec ajout de cellulose ont déjà démontré une consommation de CO de 111%, comparativement à un sol contrôle, dès le début de l'incubation et jusqu'à 120% après une semaine. L'ajout de glucose, une source de carbone simple et très labile, a induit une diminution de consommation avec 86% de la consommation du contrôle au début de l'incubation et une consommation de 109% après une semaine. Cela démontre bien l'implication des stratégistes K dans la consommation du CO comparativement aux stratégistes R. Toutefois, leur domestication pourrait être très difficile ou impossible. Une autre possibilité serait d'utiliser la technique CARD-FISH (catalyzed reporter deposition fluorescence *in situ* hybridization) pour cibler le génotype *coxL* atypique l'aide d'oligonucléotides spécifiques. Par contre, l'identification de ces bactéries pourrait se faire en utilisant la cytométrie en flux permettant de trier les bactéries marquées (celles dont la sonde y serait fixée) pour éventuellement séquencer leur génome. Suivant cette approche, leur identité serait révélée et aiderait grandement à l'avancement de la compréhension du puits biologique de CO.

RÉFÉRENCES

- Alonso AA, Molina I & Theodoropoulos C (2014) Modeling bacterial population growth from stochastic single-cell dynamics. *Appl. Environ. Microbiol.* 80(17):5241-5253.
- Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. *J. Mol. Biol.* 215(3):403-410.
- Anderson MJ, Ellingsen KE & McArdle BH (2006) Multivariate dispersion as a measure of beta diversity. *Ecol. Lett.* 9(6):683-693.
- Badr O & Probert SD (1995) Sinks and environmental impacts for atmospheric carbon monoxide. *Appl. Energy* 50(4):339-372. Doi: 10.1016/0306-2619(95)98803-A
- Beh M, Strauss G, Huber R, Stetter K-O & Fuchs G (1993) Enzymes of the reductive citric acid cycle in the autotrophic eubacterium *Aquifex pyrophilus* and in the archaeabacterium *Thermoproteus neutrophilus*. *Arch. Microbiol.* 160(4):306-311. Doi: 10.1007/BF00292082
- Berthelet M, Whyte LG & Greer CW (1996) Rapid, direct extraction of DNA from soils for PCR analysis using polyvinylpolypyrrolidone spin columns. *FEMS Microbiol. Lett.* 138(1):17-22. Doi: 10.1111/j.1574-6968.1996.tb08128.x
- Black GW, Lyons CM, Williams E, Colby J, Kehoe M & O'Reilly C (1990) Cloning and expression of the carbon monoxide dehydrogenase genes from *Pseudomonas thermocarboxydovorans* strain C2. *FEMS Microbiol. Lett.* 58(3):249-254.
- Burnett RT, Cakmak S, Raizenne ME, Stieb D, Vincent R, Krewski D, Brook JR, Philips O & Ozkaynak H (1998) The association between ambient carbon monoxide levels and daily mortality in Toronto, Canada. *J. Air Waste Manag. Assoc.* 48(8):689-700.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK & Gordon JI (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7(5):335-336.
- Chan ASK & Steudler PA (2006) Carbon monoxide uptake kinetics in unamended and long-term nitrogen-amended temperate forest soils. *FEMS Microbiol. Ecol.* 57(3):343-354.
- Cho JW, Yim HS & Kim YM (1985) Acinetobacter isolates growing with carbon monoxide. *Kor. J. Microbiol.* 23:1-8.
- Conrad R (1994) Compensation concentration as critical variable for regulating the flux of trace gases between soil and atmosphere. *Biogeochemistry* 27(3):155-170.
- Conrad R (1996) Soil microorganisms as controllers of atmospheric trace gases (H_2 , CO, CH_4 , OCS, N_2O , and NO). *Microbiol. Rev.* 60(4):609-640.
- Conrad R (1999) Soil microorganisms oxidizing atmospheric trace gases (CH_4 , CO, H_2 , NO). *Indian J. Microbiol.* 39(4):193-204.
- Conrad R, Meyer O & Seiler W (1981) Role of carboxydotobacteria in consumption of atmospheric carbon monoxide by soil. *Appl. Environ. Microbiol.* 42(2):211-215.
- Conrad R & Seiler W (1980) Role of microorganisms in the consumption and production of atmospheric carbon monoxide by soil. *Appl. Environ. Microbiol.* 40(3):437-445.

- Conrad R & Seiler W (1985) Characteristics of abiological carbon monoxide formation from soil organic matter, humic acids, and phenolic compounds. *Environ. Sci. Technol.* 19(12):1165-1169. Doi: 10.1021/es00142a004
- Constant P, Poissant L & Villemur R (2008) Isolation of Streptomyces sp. PCB7, the first microorganism demonstrating high-affinity uptake of tropospheric H₂. *ISME J.* 2(10):1066-1076.
- Crutzen PJ & Gidel LT (1983) A two-dimensional photochemical model of the atmosphere. 2: The tropospheric budgets of the anthropogenic chlorocarbons CO, CH₄, CH₃Cl and the effect of various NO_x sources on tropospheric ozone. *J. Geophys. Res. Oc. Atm.* 88(Nc11):6641-6661. Doi: 10.1029/Jc088ic11p06641
- Cunliffe M (2011) Correlating carbon monoxide oxidation with cox genes in the abundant Marine Roseobacter Clade. *ISME J.* 5(4):685-691. Doi: 10.1038/ismej.2010.170
- Cunliffe M (2013) Physiological and metabolic effects of carbon monoxide oxidation in the model marine bacterioplankton Ruegeria pomeroyi DSS-3. *Appl. Environ. Microbiol.* 79(2):738-740. Doi: 10.1128/AEM.02466-12
- Cunliffe M, Schafer H, Harrison E, Cleave S, Upstill-Goddard R & Murrell JC (2008) Phylogenetic and functional gene analysis of the bacterial and archaeal communities associated with the surface microlayer of an estuary. *ISME J.* 2(7):776-789.
- Cypionka H & Meyer O (1982) Influence of carbon monoxide on growth and respiration of carboxydobacteria and other aerobic organisms. *FEMS Microbiol. Lett.* 15(3):209-214. Doi: 10.1111/j.1574-6968.1982.tb00069.x
- Cypionka H & Meyer O (1983) Carbon monoxide-insensitive respiratory chain of Pseudomonas carboxydovorans. *J Bacteriol.* 156(3):1178-1187.
- Cypionka H, Meyer O & Schlegel HG (1980) Physiological-characteristics of various species of strains of carboxydobacteria. *Arch. Microbiol.* 127(3):301-307. Doi: 10.1007/Bf00427208
- Daniel JS & Solomon S (1998) On the climate forcing of carbon monoxide. *J Geophys. Res.* 103(D11):13249. Doi: 10.1029/98jd00822
- Davenport HE (1949) The haemoglobins of Ascaris lumbricoides. *Proc. R. Soc. Lond., B, Biol. Sci.* 136(883):255-270.
- Dobbek H, Gremer L, Kiefersauer R, Huber R & Meyer O (2002) Catalysis at a dinuclear [CuSMo(=O)OH] cluster in a CO dehydrogenase resolved at 1.1-A resolution. *Proc. Natl. Acad. Sci. U.S.A.* 99(25):15971-15976. Doi: 10.1073/pnas.212640899
- Dobbek H, Gremer L, Meyer O & Huber R (1999) Crystal structure and mechanism of CO dehydrogenase, a molybdo iron-sulfur flavoprotein containing S-selanyl cysteine. *Proc. Natl. Acad. Sci. U.S.A.* 96(16):8884-8889.
- Dobbek H, Svetlitchnyi V, Gremer L, Huber R & Meyer O (2001) Crystal structure of a carbon monoxide dehydrogenase reveals a [Ni-4Fe-5S] cluster. *Science* 293(5533):1281-1285. Doi: 10.1126/science.1061500
- Dunfield KE & King GM (2004) Molecular analysis of carbon monoxide-oxidizing bacteria associated with recent Hawaiian volcanic deposits. *Appl. Environ. Microbiol.* 70(7):4242-4248. Doi: 10.1128/aem.70.7.4242-4248.2004
- Edgar RC, Haas BJ, Clemente JC, Quince C & Knight R (2011) UCHIME improves sensitivity and speed of chimer detection. *Bioinformatics* 27(16):2194-2200.

Doi: 10.1093/bioinformatics/btr381

- Ehhalt DH, Prather MJ, Dentener F, Derwent R, Dlugokencky E, Holland EA, Isaksen I, Katima J, Kirchhoff VWJH, Matson P, Midgley P & Wang M (2001) Atmospheric chemistry and greenhouse gases contents. *Climate Change 2001: Working Group I: The Scientific Basis*, Cambridge University Press, Cambridge, UK.
- Elghamry W & Elashkar M (1962) Simplified textural lassification triangles. *Soil Sci. Soc. Am. J.* 26(6):612-613. Doi: 10.2136/sssaj1962.03615995002600060028x
- Evans MC, Buchanan BB & Arnon DI (1966) A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *Proc. Natl. Acad. Sci. U.S.A.* 55(4):928-934.
- Evans WFJ & Puckrin E (1995) An observation of the greenhouse radiation associated with carbon monoxide. *Geophys. Res. Lett.* 22(8):925-928. Doi: 10.1029/95GL00606
- Forster P, Ramaswamy V, Artaxo P, Berntsen T, Betts R, Fahey DW, Haywood J, Lean J, Lowe DC, Myhre G, Nganga J, Prinn R, Raga G, Schulz M & Van Dorland R (2007) Changes in atmospheric constituents and in radiative forcing. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, UK.
- Fortems-Cheiney A, Chevallier F, Pison I, Bousquet P, Szopa S, Deeter MN & Clerbaux C (2011) Ten years of CO emissions as seen from Measurements of Pollution in the Troposphere (MOPITT). *J. Geophys. Res. Atmos.* 116(D5):n/a-n/a.
Doi: 10.1029/2010JD014416
- Fuglestvedt JS, Isaksen ISA & Wang WC (1996) Estimates of indirect global warming potentials for CH₄, CO and NO_x. *Clim. Change* 34(3-4):405-437. Doi: 10.1007/Bf00139300
- Fuhrmann S, Ferner M, Jeffke T, Henne A, Gottschalk G & Meyer O (2003) Complete nucleotide sequence of the circular megaplasmid pHCG3 of Oligotropha carboxidovorans: Function in the chemolithoautotrophic utilization of CO, H₂ and CO₂. *Gene* 322(1-2):67-75.
- Gnida M, Ferner R, Gremer L, Meyer O & Meyer-Klaucke W (2003) A novel binuclear [CuSMo] cluster at the active site of carbon monoxide dehydrogenase: Characterization by x-ray absorption spectroscopy. *Biochemistry* 42(1):222-230.
- Gödde M, Meuser K & Conrad R (2000) Hydrogen consumption and carbon monoxide production in soils with different properties. *Biol. Fertil. Soils* 32(2):129-134.
- Gomez-Alvarez V, King GM & Nusslein K (2007) Comparative bacterial diversity in recent Hawaiian volcanic deposits of different ages. *FEMS Microbiol. Ecol.* 60(1):60-73.
Doi: 10.1111/j.1574-6941.2006.00253.x
- Gremer L, Kellner S, Dobbek H, Huber R & Meyer O (2000) Binding of flavin adenine dinucleotide to molybdenum-containing carbon monoxide dehydrogenase from Oligotropha carboxidovorans. *J. Biol. Chem.* 275(3):1864-1872.
- Hänelmann P, Dobbek H, Gremer L, Huber R & Meyer O (2000) The effect of intracellular molybdenum in Hydrogenophaga pseudoflava on the crystallographic structure of the seleno-molybdo-iron-sulfur flavoenzyme carbon monoxide dehydrogenase. *J. Mol. Biol.* 301(5):1221-1235. Doi: 10.1006/jmbi.2000.4023
- Hardy KR & King GM (2001) Enrichment of high-affinity CO oxidizers in Maine forest soil. *Appl. Environ. Microbiol.* 67(8):3671-3676.

- Harrell FEJ (2015) *Hmisc: Harrell Miscellaneous*, <http://CRAN.R-project.org/package=Hmisc>
- Heichel GH (1973) Removal of carbon monoxide by field and forest soils. *J. Environ. Qual.* 2(4):419-423. Doi: 10.2134/jeq1973.00472425000200040001x
- Heidt LE, Krasnec JP, Lueb RA, Pollock WH, Henry BE & Crutzen PJ (1980) Latitudinal distributions of CO and CH₄ over the Pacific. *J. Geophys. Res. Oc.* 85(C12):7329-7336. Doi: 10.1029/JC085iC12p07329
- Hille R (2005) Molybdenum-containing hydroxylases. *Arch. Biochem. Biophys.* 433(1):107-116. Doi: 10.1016/j.abb.2004.08.012
- Hoeft SE, Blum JS, Stoltz JF, Tabita FR, Witte B, King GM, Santini JM & Oremland RS (2007) Alkalilimnicola ehrlichii sp. nov., a novel, arsenite-oxidizing haloalkaliphilic gammaproteobacterium capable of chemoautotrophic or heterotrophic growth with nitrate or oxygen as the electron acceptor. *Int. J. Syst. Evol. Microbiol.* 57(3):504-512. Doi: 10.1099/ijss.0.64576-0
- Hugendieck I & Meyer O (1992) The structural genes encoding CO dehydrogenase subunits (coxl, M and S) in *Pseudomonas carboxydovorans* OM5 reside on plasmid pHCG3 and are, with the exception of *Streptomyces thermoautotrophicus*, conserved in carboxydrophic bacteria. *Arch. Microbiol.* 157(3):301-304. Doi: 10.1007/BF00245166
- Ingersoll RB, Inman RE & Fisher WR (1974) Soil's potential as a sink for atmospheric carbon monoxide. *Tellus* 26(1-2):151-159. Doi: 10.1111/j.2153-3490.1974.tb01961.x
- Inman RE & Ingersoll RB (1971a) Uptake of carbon monoxide by soil fungi. *J. Air Pollut. Contr. Ass.* 21(10):646-647. Doi: 10.1080/00022470.1971.10469582
- Inman RE, Ingersoll RB & Levy EA (1971b) Soil: A natural sink for carbon monoxide. *Science* 172(3989):1229-1231.
- Jeoung JH, Fesseler J, Goetzl S & Dobbek H (2014) Carbon monoxide. Toxic gas and fuel for anaerobes and aerobes: Carbon monoxide dehydrogenases. in *Met. Ions Life Sci.* p. 37-69.
- Johnson JL, Rajagopalan KV & Meyer O (1990) Isolation and characterization of a second molybdopterin dinucleotide: molybdopterin cytosine dinucleotide. *Arch. Biochem. Biophys.* 283(2):542-545.
- Kaneko T, Nakamura Y, Sato S, Minamisawa K, Uchiumi T, Sasamoto S, Watanabe A, Idesawa K, Iriguchi M, Kawashima K, Kohara M, Matsumoto M, Shimpo S, Tsuruoka H, Wada T, Yamada M & Tabata S (2002) Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res.* 9(6):189-197. Doi: 10.1093/dnares/9.6.189
- Kang BS & Kim YM (1999) Cloning and molecular characterization of the genes for carbon monoxide dehydrogenase and localization of molybdopterin, flavin adenine dinucleotide, and iron-sulfur centers in the enzyme of *Hydrogenophaga pseudoflava*. *J. Bacteriol.* 181(18):5581-5590.
- Kiessling M & Meyer O (1982) Profitable oxidation of carbon monoxide or hydrogen during heterotrophic growth of *Pseudomonas carboxydoflava*. *FEMS Microbiol. Lett.* 13(3):333-338. Doi: 10.1111/j.1574-6968.1982.tb08283.x

- Kim YJ & Kim YM (1989) Induction of carbon monoxide dehydrogenase during heterotrophic growth of *Acinetobacter* sp. strain JC1 DSM 3803 in the presence of carbon monoxide. *FEMS Microbiol. Lett.* 59(1-2):207-210. Doi: 10.1111/j.1574-6968.1989.tb03111.x
- Kim YM & Hegeman GD (1981) Purification and some properties of carbon monoxide dehydrogenase from *Pseudomonas carboxydohydrogena*. *J. Bacteriol.* 148(3):904-911.
- Kim YM & Hegeman GD (1983) Oxidation of carbon monoxide by bacteria. *Int. Rev. Cytol.* 81:1-32.
- Kim YM & Park SW (2012) Microbiology and genetics of CO utilization in mycobacteria. *Antonie van Leeuwenhoek* 101(4):685-700.
- King CE & King GM (2014) Description of *Thermogemmatispora carboxidivorans* sp. nov., a carbon-monoxide-oxidizing member of the class Ktedonobacteria isolated from a geothermally heated biofilm, and analysis of carbon monoxide oxidation by members of the class Ktedonobacteria. *Int. J. Syst. Evol. Microbiol.* 64(Pt 4):1244-1251.
- Doi: 10.1099/ijss.0.059675-0
- King GM (1999a) Attributes of atmospheric carbon monoxide oxidation by Maine forest soils. *Appl. Environ. Microbiol.* 65(12):5257-5264.
- King GM (1999b) Characteristics and significance of atmospheric carbon monoxide consumption by soils. *Chemosphere - Global Change Sci.* 1(1-3):53-63.
- King GM (2000) Land use impacts on atmospheric carbon monoxide consumption by soils. *Global Biogeochem. Cycles* 14(4):1161-1172.
- King GM (2003a) Molecular and Culture-Based Analyses of Aerobic Carbon Monoxide Oxidizer Diversity. *Appl. Environ. Microbiol.* 69(12):7257-7265.
- King GM (2003b) Uptake of carbon monoxide and hydrogen at environmentally relevant concentrations by mycobacteria. *Appl. Environ. Microbiol.* 69(12):7266-7272.
- King GM & Hungria M (2002) Soil-atmosphere CO exchanges and microbial biogeochemistry of CO transformations in a Brazilian agricultural ecosystem. *Appl. Environ. Microbiol.* 68(9):4480-4485.
- King GM & Weber CF (2007) Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat. Rev. Microbiol.* 5(2):107-118. Doi: 10.1038/nrmicro1595
- King GM & Weber CF (2008a) Interactions between bacterial carbon monoxide and hydrogen consumption and plant development on recent volcanic deposits. *ISME J.* 2(2):195-203. Doi: 10.1038/ismej.2007.101
- King GM, Weber CF, Nanba K, Sato Y & Ohta H (2008b) Atmospheric CO and hydrogen uptake and CO oxidizer phylogeny for miyake-jima, Japan volcanic deposits. *Microbes and Environ.* 23(4):299-305.
- Kopacz M, Jacob DJ, Fisher JA, Logan JA, Zhang L, Megretskaya IA, Yantosca RM, Singh K, Henze DK, Burrows JP, Buchwitz M, Khlystova I, McMillan WW, Gille JC, Edwards DP, Eldering A, Thouret V & Nedelec P (2010) Global estimates of CO sources with high resolution by adjoint inversion of multiple satellite datasets (MOPITT, AIRS, SCIAMACHY, TES). *Atmos. Chem. Phys.* 10(3):855-876. Doi: 10.5194/acp-10-855-2010
- Langfelder P & Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9(1):559.

- Levy H, 2nd (1971) Normal atmosphere: large radical and formaldehyde concentrations predicted. *Science* 173(3992):141-143. Doi: 10.1126/science.173.3992.141
- Logan JA, Prather MJ, Wofsy SC & McElroy MB (1981) Tropospheric chemistry: A global perspective. *J. Geophys. Res. Oceans* 86(C8):7210-7254.
Doi: 10.1029/JC086iC08p07210
- Lorite MJ, Tachil J, Sanjuan J, Meyer O & Bedmar EJ (2000) Carbon monoxide dehydrogenase activity in *Bradyrhizobium japonicum*. *Appl. Environ. Microbiol.* 66(5):1871-1876.
- Lynch RC, King AJ, Farías ME, Sowell P, Vitry C & Schmidt SK (2012) The potential for microbial life in the highest-elevation (>6000 m.a.s.l.) mineral soils of the Atacama region. *J. Geophys. Res. Biogeosci.* 117(G2):G02028. Doi: 10.1029/2012JG001961
- Lynd LR, Weimer PJ, van Zyl WH & Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* 66(3):506-577
- Magoč T & Salzberg SL (2011) FLASH: Fast Length Adjustment of short reads to improve genome assemblies. *Bioinformatics*. Doi: 10.1093/bioinformatics/btr507
- McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R & Hugenholtz P (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 6(3):610-618.
Doi: 10.1038/ismej.2011.139
- Meehl GA, Stocker TF, Collins WD, Friedlingstein P, Gaye AT, Gregory JM, Kitoh A, Knutti R, Murphy JM, Noda A, Raper SCB, Watterson IG, Weaver AJ & Zhao Z-C (2007) Global Climate Projections. *Climate Change 2007: Working Group I: The Physical Science Basis*, Cambridge Press University, Cambridge, UK.
- Meyer O, Gremer L, Ferner R, Ferner M, Dobbek H, Gnida M, Meyer-Klaucke W & Huber R (2000) The role of Se, Mo and Fe in the structure and function of carbon monoxide dehydrogenase. *Biol. Chem.* 381(9-10):865-876.
- Meyer O, Jacobitz S & Krüger B (1986) Biochemistry and physiology of aerobic carbon monoxide-utilizing bacteria. *FEMS Microbiol. Lett.* 39(3):161-179.
- Meyer O & Schlegel HG (1983) Biology of aerobic carbon monoxide-oxidizing bacteria. *Annu. Rev. Microbiol.* 37:277-310. Doi: 10.1146/annurev.mi.37.100183.001425
- Monks PS (2005) Gas-phase radical chemistry in the troposphere. *Chem. Soc. Rev.* 34(5):376-395. Doi: 10.1039/b307982c
- Monson RK & Holland EA (2001) Biospheric trace gas fluxes and their control over tropospheric chemistry. *Annu. Rev. Ecol. Syst.* 32:547-576.
Doi: 10.1146/Annurev.Ecolsyst.32.081501.114136
- Moxley JM & Smith KA (1998a) Carbon monoxide production and emission by some Scottish soils. *Tellus B Chem. Phys. Meteorol.* 50(2):151-162. Doi: 10.1034/j.1600-0889.1998.t01-1-00003.x
- Moxley JM & Smith KA (1998b) Factors affecting utilisation of atmospheric CO by soils. *Soil Biol. Biochem.* 30(1):65-79.
- Novelli PC, Masarie KA, Lang PM, Hall BD, Myers RC & Elkins JW (2003) Reanalysis of tropospheric CO trends: Effects of the 1997–1998 wildfires. *J. Geophys. Res. Atmos.* 108(D15):4464. Doi: 10.1029/2002JD003031

- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH & Wagner H (2015) *vegan: Community Ecology Package*, <http://CRAN.R-project.org/package=vegan>
- Ottesen EA, Hong JW, Quake SR, Leadbetter JR (2006) Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. *Science* 314:1464-1467.
- Park K, Emmons LK, Wang Z & Mak JE (2015) Joint application of concentration and $\delta^{18}\text{O}$ to investigate the global atmospheric CO budget. *Atmosphere* 6(5):547-578.
- Pelzmann A, Ferner M, Gnida M, Meyer-Klaucke W, Maisel T & Meyer O (2009) The CoxD protein of Oligotropha carboxidovorans is a predicted AAA+ ATPase chaperone involved in the biogenesis of the CO dehydrogenase [CuSMoO₂] cluster. *J. Biol. Chem.* 284(14):9578-9586. Doi: 10.1074/jbc.M805354200
- Potter CS, Klooster SA & Chatfield RB (1996) Consumption and production of carbon monoxide in soils: A global model analysis of spatial and seasonal variation. *Chemosphere* 33(6):1175-1193.
- Prather MJ (1996) Time scales in atmospheric chemistry: Theory, GWP_s for CH₄ and CO, and runaway growth. *Geophys. Res. Lett.* 23(19):2597-2600. Doi: Doi 10.1029/96gl02371
- Quiza L, Lalonde I, Guertin C & Constant P (2014) Land-use influences the distribution and activity of high affinity CO-oxidizing bacteria associated to type I-coxL genotype in soil. *Front. Microbiol.* 5:271. Doi: 10.3389/fmicb.2014.00271
- Ramanathan V, Callis L, Cess R, Hansen J, Isaksen I, Kuhn W, Lacis A, Luther F, Mahlman J, Reck R & Schlesinger M (1987) Climate-chemical interactions and effects of changing atmospheric trace gases. *Rev. Geophys.* 25(7):1441-1482.
Doi: 10.1029/RG025i007p01441
- Ramaswamy V, Boucher O, Haigh J, Hauglustaine DA, Haywood J, Myhre G, Nakajima T, Shi GY & Solomon S (2001) Radiative Forcing of Climate Change. *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, UK.
- Regensburger B & Hennecke H (1983) RNA polymerase from Rhizobium japonicum. *Arch. Microbiol.* 135(2):103-109. Doi: 10.1007/BF00408017
- Ro YT & Kim YM (1993) Constitutive expression of carbon monoxide dehydrogenase in Acinetobacter sp. strain JC1 DSM 3803. *Kor. J. Microbiol.* 31(3):214-217.
- Rodkey FL, O'Neal JD, Collison HA & Uddin DE (1974) Relative affinity of hemoglobin S and hemoglobin A for carbon monoxide and oxygen. *Clin. Chem.* 20(1):83-84.
- Rohde M, Mayer F, Jacobitz S & Meyer O (1985) Attachment of CO dehydrogenase to the cytoplasmic membrane is limiting the respiratory rate of *Pseudomonas carboxydovorans*. *FEMS Microbiol. Lett.* 28(2):141-144. Doi: 10.1016/0378-1097(85)90140-5
- Roughton FJW & Darling RC (1944) The effect of carbon monoxide on the oxyhemoglobin dissociation curve. *Am. J. Physiol.* 141(1):17-31
- Sanhueza E, Donoso L, Scharffe D & Crutzen PJ (1994) Carbon monoxide fluxes from natural, managed, or cultivated savannah grasslands. *J. Geophys. Res.* 99(D8):16421-16427.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ & Weber CF

- (2009) Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75(23):7537-7541. Doi: 10.1128/AEM.01541-09
- Schubel U, Kraut M, Morsdorf G & Meyer O (1995) Molecular characterization of the gene cluster coxMSL encoding the molybdenum-containing carbon monoxide dehydrogenase of Oligotropha carboxidovorans. *J. Bacteriol.* 177(8):2197-2203.
- Seiler W (1974) The cycle of atmospheric CO. *Tellus* 26(1-2):116-135. Doi: 10.1111/j.2153-3490.1974.tb01958.x
- Shiba H, Kawasumi T, Igarashi Y, Kodama T & Minoda Y (1985) The CO₂ assimilation via the reductive tricarboxylic acid cycle in an obligately autotrophic, aerobic hydrogen-oxidizing bacterium, Hydrogenobacter thermophilus. *Arch. Microbiol.* 141(3):198-203.
Doi: 10.1007/BF00408058
- Siegbahn PEM & Shestakov AF (2005) Quantum chemical modeling of CO oxidation by the active site of molybdenum CO dehydrogenase. *J. Comput. Chem.* 26(9):888-898.
Doi: 10.1002/jcc.20230
- Spratt HG & Hubbard JS (1981) Carbon monoxide metabolism in roadside soils. *Appl. Environ. Microbiol.* 41(5):1192-1201.
- Stevens CM, Krout L, Walling D, Venters A, Engelkemeir A & Ross LE (1972) The isotopic composition of atmospheric carbon monoxide. *Earth Planet. Sci. Lett.* 16(2):147-165.
Doi: 10.1016/0012-821X(72)90183-5
- Strauss G & Fuchs G (1993) Enzymes of a novel autotrophic CO₂ fixation pathway in the phototrophic bacterium Chloroflexus aurantiacus, the 3-hydroxypropionate cycle. *European J. Biochem.* 215(3):633-643.
- Sze ND (1977) Anthropogenic CO emissions: Implications for the atmospheric CO-OH-CH₄ Cycle. *Science* 195(4279):673-675. Doi: 10.1126/science.195.4279.673
- Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24(8):1596-1599.
Doi: 10.1093/molbev/msm092
- Team RDC (2008) *R: A language and environment for statistical computing*. Vienna, Austria: the R Foundation for Statistical Computing. ISBN: 3-900051-07-0. Available online at <http://www.R-project.org/>
- Thompson AM (1992) The oxidizing capacity of the earth's atmosphere: Probable past and future changes. *Science* 256(5060):1157-1165. Doi: 10.2307/2877254
- Thompson AM, Chappellaz JA, Fung IY & Kucsera TL (1993) The atmospheric CH₄ increase since the last glacial maximum. *Tellus B* 45(3). Doi: 10.3402/tellusb.v45i3.15727
- Tijhuis L, Van Loosdrecht MCM & Heijnen JJ (1993) A thermodynamically based correlation for maintenance gibbs energy requirements in aerobic and anaerobic chemotrophic growth. *Biotechnology and Bioengineering* 42(4):509-519. Doi: 10.1002/bit.260420415
- Vetrovsky T & Baldrian P (2013) The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* 8(2):e57923.
Doi: 10.1371/journal.pone.0057923

- Watson RT, Rodhe H, Oeschger H & Siegenthaler U (1990) Greenhouses gases and aerosols. *Climate Change: The IPCC Scientific Assessment. Working Group I: Scientific Assessment of Climate Change*, Cambridge University Press, Cambridge, UK.
- Weber CF & King GM (2007) Physiological, ecological, and phylogenetic characterization of Stappia, a marine CO-oxidizing bacterial genus. *Appl. Environ. Microbiol.* 73(4):1266-1276.
- Weber CF & King GM (2010a) Distribution and diversity of carbon monoxide-oxidizing bacteria and bulk bacterial communities across a succession gradient on a Hawaiian volcanic deposit. *Environ. Microbiol.* 12(7):1855-1867. Doi: 10.1111/j.1462-2920.2010.02190.x
- Weber CF & King GM (2010b) Quantification of burkholderia coxL genes in Hawaiian volcanic deposits. *Appl. Environ. Microbiol.* 76(7):2212-2217. Doi: 10.1128/AEM.01861-09
- Weber CF & King GM (2012) The phylogenetic distribution and ecological role of carbon monoxide oxidation in the genus Burkholderia. *FEMS Microbiol. Ecol.* 79(1):167-175.
- Weinstock B (1969) Carbon monoxide: Residence time in the atmosphere. *Science* 166(3902):224-225. Doi: 10.1126/science.166.3902.224
- Whalen SC & Reeburgh WS (2001) Carbon monoxide consumption in upland boreal forest soils. *Soil Biol. Biochem.* 33(10):1329-1338.
- Worden HM, Deeter MN, Frankenberg C, George M, Nichitiu F, Worden J, Aben I, Bowman KW, Clerbaux C, Coheur PF, de Laat ATJ, Detweiler R, Drummond JR, Edwards DP, Gille JC, Hurtmans D, Luo M, Martínez-Alonso S, Massie S, Pfister G & Warner JX (2013) Decadal record of satellite carbon monoxide observations. *Atmos. Chem. Phys.* 13(2):837-850. Doi: 10.5194/acp-13-837-2013
- Yang J, Jiang H, Dong H, Wu G, Hou W, Zhao W, Sun Y & Lai Z (2013) Diversity of carbon monoxide-oxidizing bacteria in five lakes on the Qinghai-Tibet Plateau, China. *Geomicrobiol. J.* 30(8):758-767. Doi: 10.1080/01490451.2013.769652
- Yang J, Zhou E, Jiang H, Li W, Wu G, Huang L, Hedlund BP & Dong H (2015) Distribution and diversity of aerobic carbon monoxide-oxidizing bacteria in geothermal springs of China, the Philippines, and the United States. *Geomicrobiol. J.* 32(10) 903-913.
Doi: 10.1080/01490451.2015.1008605
- Yoon J & Pozzer A (2014) Model-simulated trend of surface carbon monoxide for the 2001–2010 decade. *Atmos. Chem. Phys.* 14(19):10465-10482. Doi: 10.5194/acp-14-10465-2014
- Yurganov L, McMillan W, Grechko E & Dzhola A (2010) Analysis of global and regional CO burdens measured from space between 2000 and 2009 and validated by ground-based solar tracking spectrometers. *Atmos. Chem. Phys.* 10(8):3479-3494. Doi: 10.5194/acp-10-3479-2010
- Zavarzin GA & Nozhevnikova AN (1977) Aerobic carboxydobacteria. *Microbial. Ecol.* 3(4):305-326.
- Zhang R, Wang M & Ren L (2001) Long-term trends of carbon monoxide inferred using a two-dimensional model. *Chemosphere - Global Change Sci.* 3(2):123-132.
Doi: [http://dx.doi.org/10.1016/S1465-9972\(01\)00002-2](http://dx.doi.org/10.1016/S1465-9972(01)00002-2)

- Anderson MJ, Ellingsen KE & McArdle BH (2006) Multivariate dispersion as a measure of beta diversity. *Ecol. Lett.* 9(6):683-693.
- Badr O & Probert SD (1995) Sinks and environmental impacts for atmospheric carbon monoxide. *Appl. Energy* 50(4):339-372. Doi: 10.1016/0306-2619(95)98803-A
- Beh M, Strauss G, Huber R, Stetter K-O & Fuchs G (1993) Enzymes of the reductive citric acid cycle in the autotrophic eubacterium *Aquifex pyrophilus* and in the archaeabacterium *Thermoproteus neutrophilus*. *Arch. Microbiol.* 160(4):306-311. Doi: 10.1007/BF00292082
- Berthelet M, Whyte LG & Greer CW (1996) Rapid, direct extraction of DNA from soils for PCR analysis using polyvinylpolypyrrolidone spin columns. *FEMS Microbiol. Lett.* 138(1):17-22. Doi: 10.1111/j.1574-6968.1996.tb08128.x
- Black GW, Lyons CM, Williams E, Colby J, Kehoe M & O'Reilly C (1990) Cloning and expression of the carbon monoxide dehydrogenase genes from *Pseudomonas thermocarboxydovorans* strain C2. *FEMS Microbiol. Lett.* 58(3):249-254.
- Burnett RT, Cakmak S, Raizenne ME, Stieb D, Vincent R, Krewski D, Brook JR, Philips O & Ozkaynak H (1998) The association between ambient carbon monoxide levels and daily mortality in Toronto, Canada. *J. Air Waste Manag. Assoc.* 48(8):689-700.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK & Gordon JI (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7(5):335-336.
- Chan ASK & Steudler PA (2006) Carbon monoxide uptake kinetics in unamended and long-term nitrogen-amended temperate forest soils. *FEMS Microbiol. Ecol.* 57(3):343-354.
- Chevalier A, Gheusi F, Attié JL, Delmas R, Zbinden R, Athier G & Cousin JM (2008) Carbon monoxide observations from ground stations in France and Europe and long trends in the free troposphere. *Atmos. Chem. Phys. Discuss.* 8(1):3313-3356. Doi: 10.5194/acpd-8-3313-2008
- Cho JW, Yim HS & Kim YM (1985) Acinetobacter isolates growing with carbon monoxide. *Kor. J. Microbiol.* 23:1-8.
- Conrad R (1994) Compensation concentration as critical variable for regulating the flux of trace gases between soil and atmosphere. *Biogeochemistry* 27(3):155-170.
- Conrad R (1996) Soil microorganisms as controllers of atmospheric trace gases (H_2 , CO, CH_4 , OCS, N_2O , and NO). *Microbiol. Rev.* 60(4):609-640.
- Conrad R (1999) Soil microorganisms oxidizing atmospheric trace gases (CH_4 , CO, H_2 , NO). *Indian J. Microbiol.* 39(4):193-204.
- Conrad R, Meyer O & Seiler W (1981) Role of carboxydotobacteria in consumption of atmospheric carbon monoxide by soil. *Appl. Environ. Microbiol.* 42(2):211-215.
- Conrad R & Seiler W (1980) Role of microorganisms in the consumption and production of atmospheric carbon monoxide by soil. *Appl. Environ. Microbiol.* 40(3):437-445.
- Conrad R & Seiler W (1985) Characteristics of abiological carbon monoxide formation from soil organic matter, humic acids, and phenolic compounds. *Environ. Sci. Technol.* 19(12):1165-1169. Doi: 10.1021/es00142a004

- Constant P, Poissant L & Villemur R (2008) Isolation of Streptomyces sp. PCB7, the first microorganism demonstrating high-affinity uptake of tropospheric H₂. *ISME J.* 2(10):1066-1076.
- Crutzen PJ & Gidel LT (1983) A two-dimensional photochemical model of the atmosphere. 2: The tropospheric budgets of the anthropogenic chlorocarbons CO, CH₄, CH₃Cl and the effect of various NO_x sources on tropospheric ozone. *J. Geophys. Res. Oc. Atm.* 88(Nc11):6641-6661. Doi: 10.1029/Jc088ic11p06641
- Cunliffe M (2011) Correlating carbon monoxide oxidation with cox genes in the abundant Marine Roseobacter Clade. *ISME J.* 5(4):685-691. Doi: 10.1038/ismej.2010.170
- Cunliffe M (2013) Physiological and metabolic effects of carbon monoxide oxidation in the model marine bacterioplankton Ruegeria pomeroyi DSS-3. *Appl. Environ. Microbiol.* 79(2):738-740. Doi: 10.1128/AEM.02466-12
- Cunliffe M, Schafer H, Harrison E, Cleave S, Upstill-Goddard R & Murrell JC (2008) Phylogenetic and functional gene analysis of the bacterial and archaeal communities associated with the surface microlayer of an estuary. *ISME J.* 2(7):776-789.
- Cypionka H & Meyer O (1982) Influence of carbon monoxide on growth and respiration of carboxydobacteria and other aerobic organisms. *FEMS Microbiol. Lett.* 15(3):209-214. Doi: 10.1111/j.1574-6968.1982.tb00069.x
- Cypionka H & Meyer O (1983) Carbon monoxide-insensitive respiratory chain of *Pseudomonas carboxydovorans*. *J Bacteriol.* 156(3):1178-1187.
- Cypionka H, Meyer O & Schlegel HG (1980) Physiological-Characteristics of Various Species of Strains of Carboxydobacteria. *Arch. Microbiol.* 127(3):301-307. Doi: 10.1007/Bf00427208
- Daniel JS & Solomon S (1998) On the climate forcing of carbon monoxide. *J Geophys. Res.* 103(D11):13249. Doi: 10.1029/98jd00822
- Davenport HE (1949) The haemoglobins of *Ascaris lumbricoides*. *Proc. R. Soc. Lond., B, Biol. Sci.* 136(883):255-270.
- Dobbek H, Gremer L, Kiefersauer R, Huber R & Meyer O (2002) Catalysis at a dinuclear [CuSMo(=O)OH] cluster in a CO dehydrogenase resolved at 1.1-A resolution. *Proc. Natl. Acad. Sci. U.S.A.* 99(25):15971-15976. Doi: 10.1073/pnas.212640899
- Dobbek H, Gremer L, Meyer O & Huber R (1999) Crystal structure and mechanism of CO dehydrogenase, a molybdo iron-sulfur flavoprotein containing S-selanyl cysteine. *Proc. Natl. Acad. Sci. U.S.A.* 96(16):8884-8889.
- Dobbek H, Svetlitchnyi V, Gremer L, Huber R & Meyer O (2001) Crystal structure of a carbon monoxide dehydrogenase reveals a [Ni-4Fe-5S] cluster. *Science* 293(5533):1281-1285. Doi: 10.1126/science.1061500
- Dunfield KE & King GM (2004) Molecular analysis of carbon monoxide-oxidizing bacteria associated with recent Hawaiian volcanic deposits. *Appl. Environ. Microbiol.* 70(7):4242-4248. Doi: 10.1128/aem.70.7.4242-4248.2004
- Edgar RC, Haas BJ, Clemente JC, Quince C & Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27(16):2194-2200. Doi: 10.1093/bioinformatics/btr381
- Ehhalt DH, Prather MJ, Dentener F, Derwent R, Dlugokencky E, Holland EA, Isaksen I, Katima J, Kirchhoff VWJH, Matson P, Midgley P & Wang M (2001) Atmospheric chemistry and

greenhouse gases contents. *Climate Change 2001: Working Group I: The Scientific Basis*, Cambridge University Press, Cambridge, UK.

- Elghamry W & Elashkar M (1962) Simplified textural lassification triangles. *Soil Sci. Soc. Am. J.* 26(6):612-613. Doi: 10.2136/sssaj1962.03615995002600060028x
- Evans MC, Buchanan BB & Arnon DI (1966) A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *Proc. Natl. Acad. Sci. U.S.A.* 55(4):928-934.
- Evans WFJ & Puckrin E (1995) An observation of the greenhouse radiation associated with carbon monoxide. *Geophys. Res. Lett.* 22(8):925-928. Doi: 10.1029/95GL00606
- Forster P, Ramaswamy V, Artaxo P, Berntsen T, Betts R, Fahey DW, Haywood J, Lean J, Lowe DC, Myhre G, Nganga J, Prinn R, Raga G, Schulz M & Van Dorland R (2007) Changes in atmospheric constituents and in radiative forcing. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, UK.
- Fortems-Cheiney A, Chevallier F, Pison I, Bousquet P, Szopa S, Deeter MN & Clerbaux C (2011) Ten years of CO emissions as seen from Measurements of Pollution in the Troposphere (MOPITT). *J. Geophys. Res. Atmos.* 116(D5):n/a-n/a. Doi: 10.1029/2010JD014416
- Fuglestvedt JS, Isaksen ISA & Wang WC (1996) Estimates of indirect global warming potentials for CH₄, CO and NO_x. *Clim. Change* 34(3-4):405-437. Doi: 10.1007/Bf00139300
- Fuhrmann S, Ferner M, Jeffke T, Henne A, Gottschalk G & Meyer O (2003) Complete nucleotide sequence of the circular megaplasmid pHCG3 of Oligotropha carboxidovorans: Function in the chemolithoautotrophic utilization of CO, H₂ and CO₂. *Gene* 322(1-2):67-75.
- Gnida M, Ferner R, Gremer L, Meyer O & Meyer-Klaucke W (2003) A novel binuclear [CuSMo] cluster at the active site of carbon monoxide dehydrogenase: Characterization by x-ray absorption spectroscopy. *Biochemistry* 42(1):222-230.
- Gödde M, Meuser K & Conrad R (2000) Hydrogen consumption and carbon monoxide production in soils with different properties. *Biol. Fertil. Soils* 32(2):129-134.
- Gomez-Alvarez V, King GM & Nusslein K (2007) Comparative bacterial diversity in recent Hawaiian volcanic deposits of different ages. *FEMS Microbiol. Ecol.* 60(1):60-73. Doi: 10.1111/j.1574-6941.2006.00253.x
- Gremer L, Kellner S, Dobbek H, Huber R & Meyer O (2000) Binding of flavin adenine dinucleotide to molybdenum-containing carbon monoxide dehydrogenase from Oligotropha carboxidovorans. *J. Biol. Chem.* 275(3):1864-1872.
- Hänelmann P, Dobbek H, Gremer L, Huber R & Meyer O (2000) The effect of intracellular molybdenum in Hydrogenophaga pseudoflava on the crystallographic structure of the seleno-molybdo-iron-sulfur flavoenzyme carbon monoxide dehydrogenase. *J. Mol. Biol.* 301(5):1221-1235. Doi: 10.1006/jmbi.2000.4023
- Hardy KR & King GM (2001) Enrichment of high-affinity CO oxidizers in Maine forest soil. *Appl. Environ. Microbiol.* 67(8):3671-3676.
- Harrell FEJ (2015) *Hmisc: Harrell Miscellaneous*, <http://CRAN.R-project.org/package=Hmisc>
- Heichel GH (1973) Removal of carbon monoxide by field and forest soils. *J. Environ. Qual.* 2(4):419-423. Doi: 10.2134/jeq1973.00472425000200040001x

- Heidt LE, Krasnec JP, Lueb RA, Pollock WH, Henry BE & Crutzen PJ (1980) Latitudinal distributions of CO and CH₄ over the Pacific. *J. Geophys. Res.* 85(C12):7329-7336. Doi: 10.1029/JC085iC12p07329
- Hille R (2005) Molybdenum-containing hydroxylases. *Arch. Biochem. Biophys.* 433(1):107-116. Doi: 10.1016/j.abb.2004.08.012
- Hoeft SE, Blum JS, Stoltz JF, Tabita FR, Witte B, King GM, Santini JM & Oremland RS (2007) Alkalilimnicola ehrlichii sp. nov., a novel, arsenite-oxidizing haloalkaliphilic gammaproteobacterium capable of chemoautotrophic or heterotrophic growth with nitrate or oxygen as the electron acceptor. *Int. J. Syst. Evol. Microbiol.* 57(3):504-512. Doi: 10.1099/ijss.0.64576-0
- Hugendieck I & Meyer O (1992) The structural genes encoding CO dehydrogenase subunits (coxL, M and S) in *Pseudomonas carboxydovorans* OM5 reside on plasmid pHCG3 and are, with the exception of *Streptomyces thermoautotrophicus*, conserved in carboxydrophic bacteria. *Arch. Microbiol.* 157(3):301-304. Doi: 10.1007/BF00245166
- Ingersoll RB, Inman RE & Fisher WR (1974) Soil's potential as a sink for atmospheric carbon monoxide. *Tellus* 26(1-2):151-159. Doi: 10.1111/j.2153-3490.1974.tb01961.x
- Inman RE & Ingersoll RB (1971a) Uptake of carbon monoxide by soil fungi. *J. Air Pollut. Contr. Ass.* 21(10):646-647. Doi: 10.1080/00022470.1971.10469582
- Inman RE, Ingersoll RB & Levy EA (1971b) Soil: A natural sink for carbon monoxide. *Science* 172(3989):1229-1231.
- Jeoung JH, Fesseler J, Goetzl S & Dobbek H (2014) Carbon monoxide. Toxic gas and fuel for anaerobes and aerobes: Carbon monoxide dehydrogenases. in *Met. Ions Life Sci.* p. 37-69.
- Johnson JL, Rajagopalan KV & Meyer O (1990) Isolation and characterization of a second molybdopterin dinucleotide: molybdopterin cytosine dinucleotide. *Arch. Biochem. Biophys.* 283(2):542-545.
- Kaneko T, Nakamura Y, Sato S, Minamisawa K, Uchiumi T, Sasamoto S, Watanabe A, Idesawa K, Iriuchi M, Kawashima K, Kohara M, Matsumoto M, Shimpou S, Tsuruoka H, Wada T, Yamada M & Tabata S (2002) Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res.* 9(6):189-197. Doi: 10.1093/dnares/9.6.189
- Kang BS & Kim YM (1999) Cloning and molecular characterization of the genes for carbon monoxide dehydrogenase and localization of molybdopterin, flavin adenine dinucleotide, and iron-sulfur centers in the enzyme of *Hydrogenophaga pseudoflava*. *J. Bacteriol.* 181(18):5581-5590.
- Kiessling M & Meyer O (1982) Profitable oxidation of carbon monoxide or hydrogen during heterotrophic growth of *Pseudomonas carboxydoflava*. *FEMS Microbiol. Lett.* 13(3):333-338. Doi: 10.1111/j.1574-6968.1982.tb08283.x
- Kim YJ & Kim YM (1989) Induction of carbon monoxide dehydrogenase during heterotrophic growth of *Acinetobacter* sp. strain JC1 DSM 3803 in the presence of carbon monoxide. *FEMS Microbiol. Lett.* 59(1-2):207-210. Doi: 10.1111/j.1574-6968.1989.tb03111.x
- Kim YM & Hegeman GD (1981) Purification and some properties of carbon monoxide dehydrogenase from *Pseudomonas carboxydohydrogena*. *J. Bacteriol.* 148(3):904-911.

- Kim YM & Hegeman GD (1983) Oxidation of carbon monoxide by bacteria. *Int. Rev. Cytol.* 81:1-32.
- Kim YM & Park SW (2012) Microbiology and genetics of CO utilization in mycobacteria. *Antonie van Leeuwenhoek* 101(4):685-700.
- King CE & King GM (2014) Description of Thermogemmatispora carboxidivorans sp. nov., a carbon-monoxide-oxidizing member of the class Ktedonobacteria isolated from a geothermally heated biofilm, and analysis of carbon monoxide oxidation by members of the class Ktedonobacteria. *Int. J. Syst. Evol. Microbiol.* 64(Pt 4):1244-1251.
Doi: 10.1099/ijss.0.059675-0
- King GM (1999a) Attributes of atmospheric carbon monoxide oxidation by Maine forest soils. *Appl. Environ. Microbiol.* 65(12):5257-5264.
- King GM (1999b) Characteristics and significance of atmospheric carbon monoxide consumption by soils. *Chemosphere - Global Change Sci.* 1(1-3):53-63.
- King GM (2000) Land use impacts on atmospheric carbon monoxide consumption by soils. *Global Biogeochem. Cycles* 14(4):1161-1172.
- King GM (2003a) Molecular and Culture-Based Analyses of Aerobic Carbon Monoxide Oxidizer Diversity. *Appl. Environ. Microbiol.* 69(12):7257-7265.
- King GM (2003b) Uptake of carbon monoxide and hydrogen at environmentally relevant concentrations by mycobacteria. *Appl. Environ. Microbiol.* 69(12):7266-7272.
- King GM & Hungria M (2002) Soil-atmosphere CO exchanges and microbial biogeochemistry of CO transformations in a Brazilian agricultural ecosystem. *Appl. Environ. Microbiol.* 68(9):4480-4485.
- King GM & Weber CF (2007) Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat. Rev. Microbiol.* 5(2):107-118. Doi: 10.1038/nrmicro1595
- King GM & Weber CF (2008a) Interactions between bacterial carbon monoxide and hydrogen consumption and plant development on recent volcanic deposits. *ISME J.* 2(2):195-203.
Doi: 10.1038/ismej.2007.101
- King GM, Weber CF, Nanba K, Sato Y & Ohta H (2008b) Atmospheric CO and hydrogen uptake and CO oxidizer phylogeny for miyake-jima, Japan volcanic deposits. *Microbes and Environ.* 23(4):299-305.
- Kopacz M, Jacob DJ, Fisher JA, Logan JA, Zhang L, Megretskaya IA, Yantosca RM, Singh K, Henze DK, Burrows JP, Buchwitz M, Khlystova I, McMillan WW, Gille JC, Edwards DP, Eldering A, Thouret V & Nedelev P (2010) Global estimates of CO sources with high resolution by adjoint inversion of multiple satellite datasets (MOPITT, AIRS, SCIAMACHY, TES). *Atmos. Chem. Phys.* 10(3):855-876. Doi: 10.5194/acp-10-855-2010
- Langfelder P & Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9(1):559.
- Levy H, 2nd (1971) Normal atmosphere: large radical and formaldehyde concentrations predicted. *Science* 173(3992):141-143. Doi: 10.1126/science.173.3992.141
- Logan JA, Prather MJ, Wofsy SC & McElroy MB (1981) Tropospheric chemistry: A global perspective. *J. Geophys. Res. Oceans* 86(C8):7210-7254.
Doi: 10.1029/JC086iC08p07210

- Lorite MJ, Tachil J, Sanjuan J, Meyer O & Bedmar EJ (2000) Carbon monoxide dehydrogenase activity in *Bradyrhizobium japonicum*. *Appl. Environ. Microbiol.* 66(5):1871-1876.
- Lynch RC, King AJ, Farías ME, Sowell P, Vitry C & Schmidt SK (2012) The potential for microbial life in the highest-elevation (>6000 m.a.s.l.) mineral soils of the Atacama region. *J. Geophys. Res. Biogeosci.* 117(G2):G02028. Doi: 10.1029/2012JG001961
- Magoč T & Salzberg SL (2011) FLASH: Fast Length Adjustment of short reads to improve genome assemblies. *Bioinformatics*. Doi: 10.1093/bioinformatics/btr507
- McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R & Hugenholtz P (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 6(3):610-618. Doi: 10.1038/ismej.2011.139
- Meehl GA, Stocker TF, Collins WD, Friedlingstein P, Gaye AT, Gregory JM, Kitoh A, Knutti R, Murphy JM, Noda A, Raper SCB, Watterson IG, Weaver AJ & Zhao Z-C (2007) Global Climate Projections. *Climate Change 2007: Working Group I: The Physical Science Basis*, Cambridge Press University, Cambridge, UK.
- Meyer O, Gremer L, Ferner R, Ferner M, Dobbek H, Gnida M, Meyer-Klaucke W & Huber R (2000) The role of Se, Mo and Fe in the structure and function of carbon monoxide dehydrogenase. *Biol. Chem.* 381(9-10):865-876.
- Meyer O, Jacobitz S & Krüger B (1986) Biochemistry and physiology of aerobic carbon monoxide-utilizing bacteria. *FEMS Microbiol. Lett.* 39(3):161-179.
- Meyer O & Schlegel HG (1983) Biology of aerobic carbon monoxide-oxidizing bacteria. *Annu. Rev. Microbiol.* 37:277-310. Doi: 10.1146/annurev.mi.37.100183.001425
- Monks PS (2005) Gas-phase radical chemistry in the troposphere. *Chem. Soc. Rev.* 34(5):376-395. Doi: 10.1039/b307982c
- Monson RK & Holland EA (2001) Biospheric trace gas fluxes and their control over tropospheric chemistry. *Annu. Rev. Ecol. Syst.* 32:547-576.
Doi: 10.1146/Annurev.Ecols.32.081501.114136
- Moxley JM & Smith KA (1998a) Carbon monoxide production and emission by some Scottish soils. *Tellus B Chem. Phys. Meteorol.* 50(2):151-162. Doi: 10.1034/j.1600-0889.1998.t01-1-00003.x
- Moxley JM & Smith KA (1998b) Factors affecting utilisation of atmospheric CO by soils. *Soil Biol. Biochem.* 30(1):65-79.
- Novelli PC, Masarie KA, Lang PM, Hall BD, Myers RC & Elkins JW (2003) Reanalysis of tropospheric CO trends: Effects of the 1997–1998 wildfires. *J. Geophys. Res. Atmos.* 108(D15):4464. Doi: 10.1029/2002JD003031
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH & Wagner H (2015) *vegan: Community Ecology Package*, <http://CRAN.R-project.org/package=vegan>
- Pelzmann A, Ferner M, Gnida M, Meyer-Klaucke W, Maisel T & Meyer O (2009) The CoxD protein of *Oligotropha carboxidovorans* is a predicted AAA+ ATPase chaperone involved in the biogenesis of the CO dehydrogenase [CuSMoO₂] cluster. *J. Biol. Chem.* 284(14):9578-9586. Doi: 10.1074/jbc.M805354200

- Potter CS, Klooster SA & Chatfield RB (1996) Consumption and production of carbon monoxide in soils: A global model analysis of spatial and seasonal variation. *Chemosphere* 33(6):1175-1193.
- Prather MJ (1996) Time scales in atmospheric chemistry: Theory, GWPs for CH₄ and CO, and runaway growth. *Geophys. Res. Lett.* 23(19):2597-2600.
Doi: Doi 10.1029/96gl02371
- Quiza L, Lalonde I, Guertin C & Constant P (2014) Land-use influences the distribution and activity of high affinity CO-oxidizing bacteria associated to type I-coxL genotype in soil. *Front. Microbiol.* 5:271. Doi: 10.3389/fmicb.2014.00271
- Ramanathan V, Callis L, Cess R, Hansen J, Isaksen I, Kuhn W, Lacis A, Luther F, Mahlman J, Reck R & Schlesinger M (1987) Climate-chemical interactions and effects of changing atmospheric trace gases. *Rev. Geophys.* 25(7):1441-1482.
Doi: 10.1029/RG025i007p01441
- Ramaswamy V, Boucher O, Haigh J, Hauglustaine DA, Haywood J, Myhre G, Nakajima T, Shi GY & Solomon S (2001) Radiative Forcing of Climate Change. *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, UK.
- Regensburger B & Hennecke H (1983) RNA polymerase from Rhizobium japonicum. *Arch. Microbiol.* 135(2):103-109. Doi: 10.1007/BF00408017
- Ro YT & Kim YM (1993) Constitutive expression of carbon monoxide dehydrogenase in Acinetobacter sp. strain JC1 DSM 3803. *Kor. J. Microbiol.* 31(3):214-217.
- Rodkey FL, O'Neal JD, Collison HA & Uddin DE (1974) Relative affinity of hemoglobin S and hemoglobin A for carbon monoxide and oxygen. *Clin. Chem.* 20(1):83-84.
- Rohde M, Mayer F, Jacobitz S & Meyer O (1985) Attachment of CO dehydrogenase to the cytoplasmic membrane is limiting the respiratory rate of *Pseudomonas carboxydovorans*. *FEMS Microbiol. Lett.* 28(2):141-144. Doi: 10.1016/0378-1097(85)90140-5
- Roughton FJW & Darling RC (1944) The effect of carbon monoxide on the oxyhemoglobin dissociation curve. *Am. J. Physiol.* 141(1):17-31
- Sanhueza E, Donoso L, Scharffe D & Crutzen PJ (1994) Carbon monoxide fluxes from natural, managed, or cultivated savannah grasslands. *J. Geophys. Res.* 99(D8):16421-16427.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ & Weber CF (2009) Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75(23):7537-7541. Doi: 10.1128/AEM.01541-09
- Schubel U, Kraut M, Morsdorf G & Meyer O (1995) Molecular characterization of the gene cluster coxMSL encoding the molybdenum-containing carbon monoxide dehydrogenase of Oligotropha carboxidovorans. *J. Bacteriol.* 177(8):2197-2203.
- Seiler W (1974) The cycle of atmospheric CO. *Tellus* 26(1-2):116-135. Doi: 10.1111/j.2153-3490.1974.tb01958.x
- Shiba H, Kawasumi T, Igarashi Y, Kodama T & Minoda Y (1985) The CO₂ assimilation via the reductive tricarboxylic acid cycle in an obligately autotrophic, aerobic hydrogen-oxidizing bacterium, *Hydrogenobacter thermophilus*. *Arch. Microbiol.* 141(3):198-203.

Doi: 10.1007/BF00408058

Siegbahn PEM & Shestakov AF (2005) Quantum chemical modeling of CO oxidation by the active site of molybdenum CO dehydrogenase. *J. Comput. Chem.* 26(9):888-898.
Doi: 10.1002/jcc.20230

Spratt HG & Hubbard JS (1981) Carbon monoxide metabolism in roadside soils. *Appl. Environ. Microbiol.* 41(5):1192-1201.

Stevens CM, Krout L, Walling D, Venters A, Engelkemeir A & Ross LE (1972) The isotopic composition of atmospheric carbon monoxide. *Earth Planet. Sci. Lett.* 16(2):147-165.
Doi: 10.1016/0012-821X(72)90183-5

Strauss G & Fuchs G (1993) Enzymes of a novel autotrophic CO₂ fixation pathway in the phototrophic bacterium *Chloroflexus aurantiacus*, the 3-hydroxypropionate cycle. *European J. Biochem.* 215(3):633-643.

Sze ND (1977) Anthropogenic CO emissions: Implications for the atmospheric CO-OH-CH₄ Cycle. *Science* 195(4279):673-675. Doi: 10.1126/science.195.4279.673

Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24(8):1596-1599.
Doi: 10.1093/molbev/msm092

Team RDC (2008) *R: A language and environment for statistical computing*. Vienna, Austria : the R Foundation for Statistical Computing. ISBN: 3-900051-07-0. Available online at <http://www.R-project.org/>

Thompson AM (1992) The oxidizing capacity of the earth's atmosphere: Probable past and future changes. *Science* 256(5060):1157-1165. Doi: 10.2307/2877254

Thompson AM, Chappellaz JA, Fung IY & Kucsera TL (1993) The atmospheric CH₄ increase since the last glacial maximum. *Tellus B* 45(3). Doi: 10.3402/tellusb.v45i3.15727

Tijhuis L, Van Loosdrecht MCM & Heijnen JJ (1993) A thermodynamically based correlation for maintenance gibbs energy requirements in aerobic and anaerobic chemotrophic growth. *Biotechnology and Bioengineering* 42(4):509-519. Doi: 10.1002/bit.260420415

Watson RT, Rodhe H, Oeschger H & Siegenthaler U (1990) Greenhouses gases and aerosols. *Climate Change: The IPCC Scientific Assessment. Working Group I: Scientific Assessment of Climate Change*, Cambridge University Press, Cambridge, UK.

Weber CF & King GM (2007) Physiological, ecological, and phylogenetic characterization of Stappia, a marine CO-oxidizing bacterial genus. *Appl. Environ. Microbiol.* 73(4):1266-1276.

Weber CF & King GM (2010a) Distribution and diversity of carbon monoxide-oxidizing bacteria and bulk bacterial communities across a succession gradient on a Hawaiian volcanic deposit. *Environ. Microbiol.* 12(7):1855-1867. Doi: 10.1111/j.1462-2920.2010.02190.x

Weber CF & King GM (2010b) Quantification of burkholderia coxL genes in Hawaiian volcanic deposits. *Appl. Environ. Microbiol.* 76(7):2212-2217. Doi: 10.1128/AEM.01861-09

Weber CF & King GM (2012) The phylogenetic distribution and ecological role of carbon monoxide oxidation in the genus Burkholderia. *FEMS Microbiol. Ecol.* 79(1):167-175.

Weinstock B (1969) Carbon monoxide: Residence time in the atmosphere. *Science* 166(3902):224-225. Doi: 10.1126/science.166.3902.224

- Whalen SC & Reeburgh WS (2001) Carbon monoxide consumption in upland boreal forest soils. *Soil Biol. Biochem.* 33(10):1329-1338.
- Worden HM, Deeter MN, Frankenberg C, George M, Nichitiu F, Worden J, Aben I, Bowman KW, Clerbaux C, Coheur PF, de Laat ATJ, Detweiler R, Drummond JR, Edwards DP, Gille JC, Hurtmans D, Luo M, Martínez-Alonso S, Massie S, Pfister G & Warner JX (2013) Decadal record of satellite carbon monoxide observations. *Atmos. Chem. Phys.* 13(2):837-850. Doi: 10.5194/acp-13-837-2013
- Yang J, Jiang H, Dong H, Wu G, Hou W, Zhao W, Sun Y & Lai Z (2013) Diversity of Carbon Monoxide-Oxidizing Bacteria in Five Lakes on the Qinghai-Tibet Plateau, China. *Geomicrobiol. J.* 30(8):758-767. Doi: 10.1080/01490451.2013.769652
- Yang J, Zhou E, Jiang H, Li W, Wu G, Huang L, Hedlund BP & Dong H (2015) Distribution and diversity of aerobic carbon monoxide-oxidizing bacteria in geothermal springs of China, the Philippines, and the United States. *Geomicrobiol. J.* 32(10) 903-913. Doi: 10.1080/01490451.2015.1008605
- Yoon J & Pozzer A (2014) Model-simulated trend of surface carbon monoxide for the 2001–2010 decade. *Atmos. Chem. Phys.* 14(19):10465-10482. Doi: 10.5194/acp-14-10465-2014
- Yurganov L, McMillan W, Grechko E & Dzhola A (2010) Analysis of global and regional CO burdens measured from space between 2000 and 2009 and validated by ground-based solar tracking spectrometers. *Atmos. Chem. Phys.* 10(8):3479-3494. Doi: 10.5194/acp-10-3479-2010
- Zavarzin GA & Nozhevnikova AN (1977) Aerobic carboxydobacteria. *Microbial. Ecol.* 3(4):305-326.
- Zhang R, Wang M & Ren L (2001) Long-term trends of carbon monoxide inferred using a two-dimensional model. *Chemosphere - Global Change Sci.* 3(2):123-132. Doi: [http://dx.doi.org/10.1016/S1465-9972\(01\)00002-2](http://dx.doi.org/10.1016/S1465-9972(01)00002-2)

ANNEXE

Annexe 1: Matériel supplémentaire (ARTICLE)

Succession of bacterial communities, *coxL* genotypes and carbon monoxide oxidation activity in soil microcosms to identify unknown carboxydovore bacteria dominating deciduous forest soil

Isabelle Lalonde and Philippe Constant*

INRS-Institut Armand-Frappier, 531 boulevard des Prairies, Laval (Québec), H7V 1B7, Canada

* Correspondence: Philippe.Constant@iaf.inrs.ca

Figure S1. Abundance of total *coxL*, *coxL* (group x) and 16S rRNA genes during the incubation, as determined by qPCR.

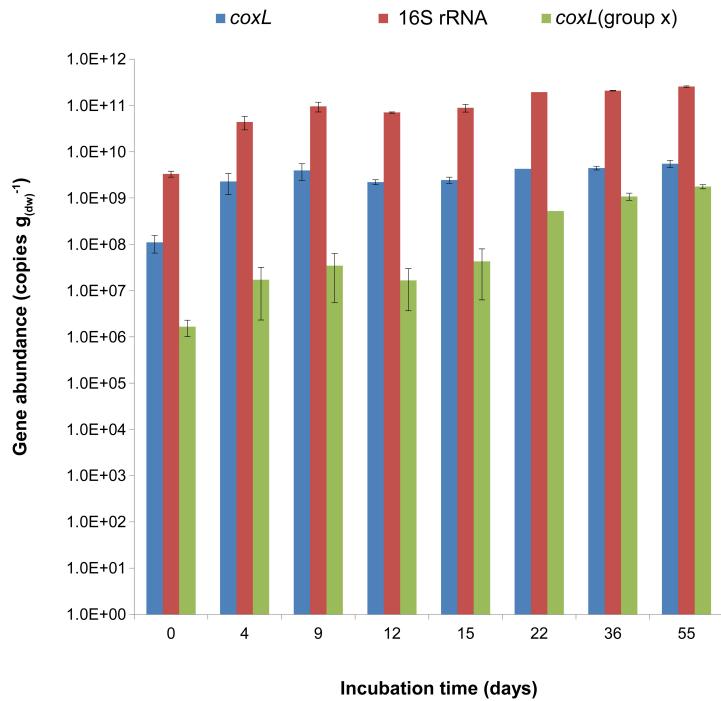


Figure S2. Homogeneity multivariate dispersion of (A) soil microbial community based of 16S rRNA gene profiles and (B) carboxydovore bacteria based on *coxL* gene profiles.

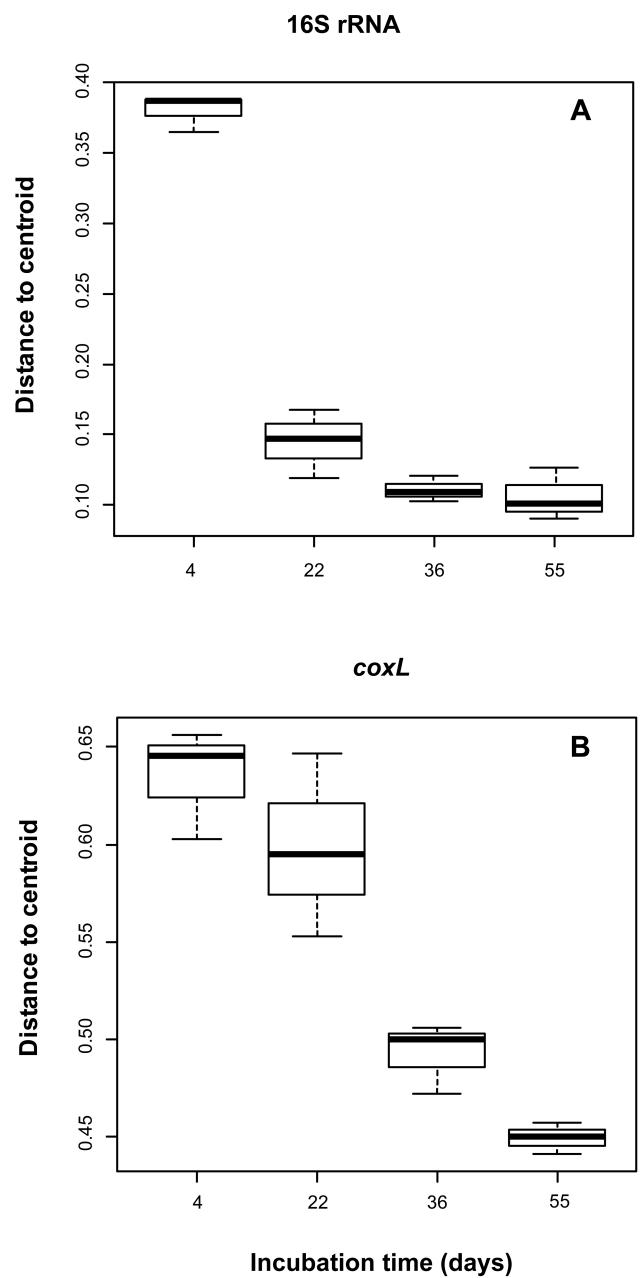
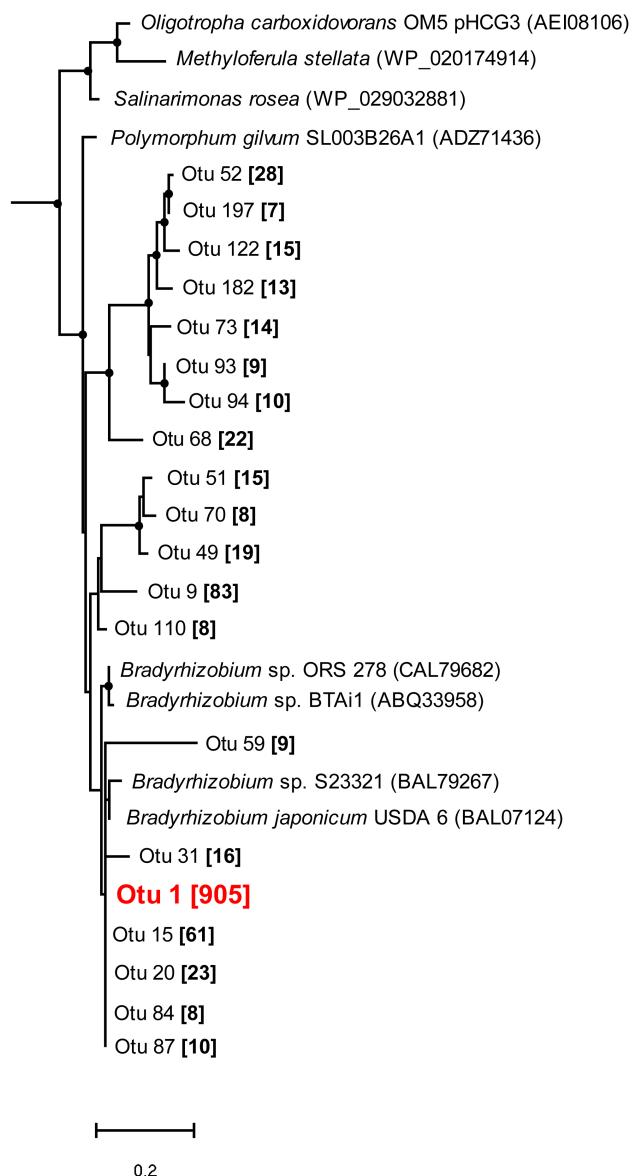
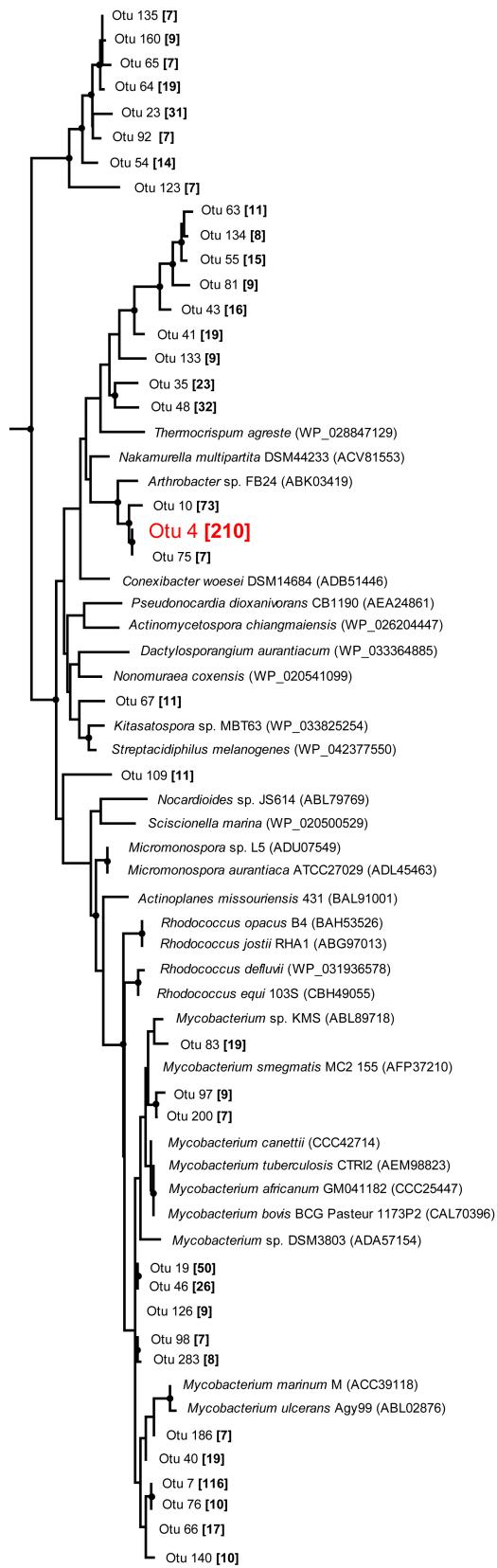


Figure S3. Detailed view of the clusters (A) *coxL*(α -Proteobacteria), (B) *coxL*(Actinobacteria), and (C) *coxL* (group x) in the *coxL* phylogenetic tree presented in the article (Fig. 3). The numbers in brackets indicate the number of sequences representing each OTU.

A



B

0.2

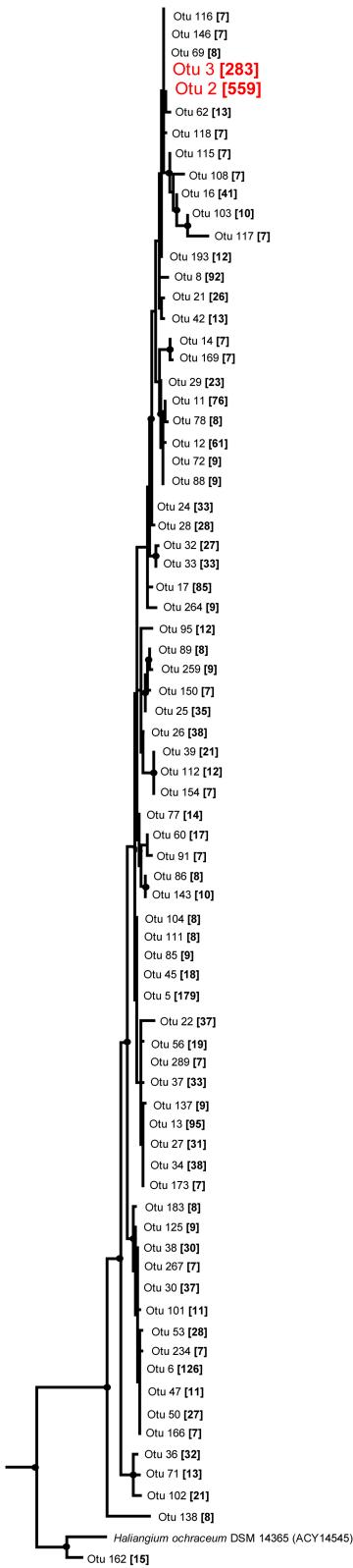
C

Table S1. Two case studies supporting the taxonomic inference of *coxL* sequences for which the colonization profile was correlated to the maturation of CO uptake activity in soil microcosms. Number of *coxL* OTUs encompassing *coxL*(group x), *coxL*(*α-Proteobacteria*) and *coxL*(*Actinobacteria*) and their number of copies per gram of soil_(dw) (see text for details).

The taxonomic inference of these OTUs assessed in the *coxL* phylogenetic analysis was validated by finding the corresponding 16S rRNA OTU present in module 1 related to CO uptake rate.

	<i>coxL</i>			16S rRNA		
	Taxonomic assignation	Number of OTUs	Absolute abundance (copies per g of soil _(dw))	Taxonomic assignation	Number of OTUs	Absolute abundance (copies per g of soil _(dw))
Case 1	<i>coxL(α-Proteobacteria)</i> Assignation based on <i>coxL</i> reference sequence: <i>Bradyrhizobium japonicum</i> USDA 6 (<i>Bradyrhizobiaceae</i>)	5	1.9 ± 1.7 x10 ⁷	<i>Bradyrhizobiaceae</i>	7	2.4 ± 0.3 x10 ¹⁰
Case 2	<i>coxL(Actinobacteria)</i> Assignation based on <i>coxL</i> reference sequence: <i>Thermocrispum agreste</i> (<i>Pseudonocardiaceae</i>)	1	3.1 ± 1.2 x10 ⁷	<i>Pseudonocardiaceae</i>	2	2.6 ± 0.8 x10 ⁸
Hypothetical	<i>coxL</i> (group x) Assignation based on <i>coxL</i> reference sequence: no taxonomic assignation	27	2.4 ± 0.3 x10 ⁹	<i>Actinospicaceae</i>	67	9.2 ± 1.2 x10 ¹⁰
				<i>Microbacteriaceae</i>	38	5.7 ± 0.6 x10 ¹⁰
				<i>Acetobacteraceae</i>	28	7.8 ± 2.8 x10 ⁹

The comparative analysis presented in table S4 shows that taxonomic affiliation of *coxL* sequences belonging to *coxL(α-Proteobacteria)* and *coxL(Actinobacteria)* inferred in the *coxL* phylogenetic analysis comprising reference *coxL* sequence was supported by the occurrence of 16S rRNA OTUs classified in the families *Bradyrhizobiaceae* and *Pseudonocardiaceae*. This linkage between *coxL* and 16S rRNA gene was supported by the species richness and the absolute abundance of *coxL* and 16S rRNA OTUs. Higher number of *Bradyrhizobiaceae* and *Pseudonocardiaceae* 16S rRNA OTUs and their higher abundance than observed in the case of *Bradyrhizobiaceae* and *Pseudonocardiaceae* *coxL* OTUs suggest that a fraction of these two *α-Proteobacteria* and *Actinobacteria* were carboxydovore bacteria (it is assumed that each 16S rRNA and *coxL* OTU belong to unique species). Comparative analysis between the species richness and the absolute abundance of *coxL*(group x) led to the proposition that unknown carboxydovores bacteria encompass the families *Actinospicaceae* and *Microbacteriaceae*. This taxonomic inference of *coxL*(group x) sequences is hypothetical and will need confirmation by the isolation of presumptive carboxydovores bacteria possessing this atypical *coxL* genotype.

Annexe 2: Commandes mothur

Ce document énumère les commandes utilisées lors de l'analyse de séquences *coxL* issues du pyroséquençage 454. La partie du nom des fichiers contenant « solA » ou « solcoxL » est présente à titre indicatif seulement. Il doit être remplacé par le nom de vos fichiers tels qu'ils s'y retrouvent dans l'ordinateur. Ce qui se retrouve après le point du nom de ces fichiers, indique le type de fichier, selon les fichiers générés après chaque commande.

A) Contrôle de qualité

1. Aperçu des séquences. À faire pour chaque échantillon.

```
summary.seqs(fasta=solA.fna)
```

2. Élimination des séquences de mauvaises qualités. À faire pour chaque échantillon.

```
trim.seqs(fasta=solA.fna, qfile=solA.qual, qaverage=25, maxambig=0, minlength=300,  
maxlength=840, maxhomop=8, oligos=F_oligosBACT.txt, pdiffs=2, keepforward=false)
```

qaverage :	score minimum de qualité des séquences conservées
maxambig :	nombre de bases ambiguës (N) maximum accepté
minlength et maxlength :	longueur minimum et maximum des séquences conservées
maxhomop :	longueur maximale d'homopolymère acceptée (suite répétitive d'un nucléotide)
oligos :	fichier comprenant l'amorce ou les amorces (dégénérations) utilisées pour le séquençage ex : forward GGGGGGGGTTTGCG forward GGGGGGGGTTTCGGC forward GGGGGGGGGCTTGGC *(TAB entre forward et l'amorce)
pdiffs :	nombre maximal d'erreurs acceptées dans la séquence d'amorce
keepfoward :	conserver la séquence d'amorce (true or false)

Génère le fichier suivant :

- solA.trim.fasta

3. Aperçu des séquences. À faire pour chaque échantillon.

Observer si les séquences ont vraiment été éliminées

```
summary.seqs(fasta=solA.trim.fasta)
```

B) Analyse des séquences

1. Regroupement des échantillons en un seul fichier

```
merge.files(input=solA.trim.fasta-solB.trim.fasta-solC.trim.fasta, output=solcoxL.fasta)
```

2. Génération du fichier group

```
make.group(fasta=solA.trim.fasta-solB.trim.fasta-solC.trim.fasta, groups=solA-solB-solC)
```

Group file : Fichier compilant les fichiers initiaux dans le groupe

Génère le fichier suivant :

- solcoxL.group

3. Conservation des séquences uniques

Les séquences identitaires sont représentées par une seule séquence pour la suite des analyses

```
unique.seqs(fasta=solcoxL.fasta)
```

Génère les fichiers suivants :

- solcoxL.names
- solcoxL.unique.fasta

Name file : Énumération des séquences que représentent une séquence unique conservée

4. Aperçu des séquences

```
summary.seqs(fasta=solcoxL.unique.fasta)
```

C) Alignement des séquences (sur LINUX absolument)

1. Alignement des séquences

reference : Fichier de séquences de référence

Génère les fichiers suivants :

- solcoxL.unique.align
- solcoxL.unique.align.report
- solcoxL.unique.flip.accnos

Align file : Ficher où les séquences sont alignées

2. Aperçu des séquences. Nécessaire pour choisir certains critères pour la prochaine commande.

```
summary.seqs(fasta=solcoxL.unique.align, name=solcoxL.names)
```

3. Élimination des séquences mal alignées

```
screen.seqs(fasta=solcoxL.unique.align, name=solcoxL.names, group=solcoxL.groups,  
end=1615, optimize=start, criteria=95, processors=2)
```

- end : Éliminer les séquences qui se terminent après la position 1615
- optimize : Éliminer les séquences qui commencent après la position où X%
des séquences se terminent
- criteria : Le pourcentage choisis pour la commande optimize
- processors : Le nombre de processors de l'ordinateur utilisé pour effectuer la
commande (L'utilisation de 2 processors ou + peu seulement être possible
si l'ordinateur est assez performant. 1 processors est utilisé par défaut)

Génère les fichiers suivants :

- solcoxL.unique.good.align
- solcoxL.unique.bad.accnos
- solcoxL.good.names
- solcoxL.good.groups

4. Aperçu des séquences restantes

```
summary.seqs(fasta=solcoxL.unique.good.align, name=solcoxL.good.names)
```

5. Élimination des caractères (« . » ou « - ») ajoutés durant l'alignement de séquences.

```
filter.seqs(fasta=solcoxL.unique.good.align, vertical=T, trump=., processors=1)
```

- vertical : T=les colonnes contenant seulement le caractère choisi sont ignorées
trump : Caractère à éliminer

Génère les fichiers suivants :

- solcoxL.filter
- solcoxL.unique.good.filter.fasta

6. Conservation des séquences uniques

```
unique.seqs(fasta=solcoxL.unique.good.filter.fasta, names= solcoxL.good.names)
```

Génère les fichiers suivants :

- solcoxL.unique.good.filter.names
- solcoxL.unique.good.filter.unique.fasta

Le fichier names indique toutes les séquences qui représentent chaque séquence unique

7. Aperçu des séquences

```
summary.seqs(fasta=solcoxL.unique.good.filter.unique.fasta,  
name=solcoxL.unique.good.filter.names)
```

8. Regroupement de séquences en ordre d'abondance

```
pre.cluster(fasta=solcoxL.unique.good.filter.unique.fasta, name=solcoxL.unique.good.filter.names,  
group=solcoxL.good.groups, diffs=5)
```

diffs : Nombre de mésappariement accepté pour le regroupement

Génère les fichiers suivants :

- solcoxL.unique.good.filter.unique.precluster.fasta
- solcoxL.unique.good.filter.precluster.names

9. Aperçu des séquences

```
summary.seqs(fasta=solcoxL.unique.good.filter.unique.precluster.fasta,  
name=solcoxL.unique.good.filter.precluster.names)
```

C) Détection et élimination des chimères

1. Détection des chimères

```
chimera.uchime(fasta=solcoxL.unique.good.filter.unique.precluster.fasta,  
name=solcoxL.unique.good.filter.unique.precluster.names, group=solcoxL.good.groups,  
processors=1)
```

Génère les fichiers suivants :

- solcoxL.unique.good.filter.unique.precluster.uchime.accnos
- solcoxL.unique.good.filter.unique.precluster.uchime.chimeras

2. Élimination des chimères

```
remove.seqs(accnos=solcoxL.unique.good.filter.unique.precluster.uchime.accnos,  
fasta=solcoxL.unique.good.filter.unique.precluster.fasta, name=solcoxL.unique.good.filter.unique.p  
recluster.names, group=solcoxL.good.groups, dups=T)
```

dups : T= si une séquence de name file est dans le accnos file, tout les séquences de cette ligne seront éliminées

Génère les fichiers suivants :

- solcoxL.unique.good.filter.unique.precluster.pick.names
- solcoxL.unique.good.filter.unique.precluster.pick.fasta
- solcoxL.good.pick.groups

3. Aperçu des séquences

```
summary.seqs(fasta=solcoxL.unique.good.filter.unique.precluster.pick.fasta,  
name=solcoxL.unique.good.filter.unique.precluster.pick.names)
```

4. Classification des séquences selon leur taxonomie

```
classify.seqs(fasta=solcoxL.unique.good.filter.unique.precluster.pick.fasta,  
name=solcoxL.unique.good.filter.unique.precluster.pick.names, group=solcoxL.good.pick.groups,  
template=refcoxL.fasta, taxonomy=refcoxL.tax, cutoff=10, processors=2)
```

Génère les fichiers suivants :

- solcoxL.unique.good.filter.unique.precluster.pick.refcoxL.wang.taxonomy
- solcoxL.unique.good.filter.unique.precluster.pick.refcoxL.wang.tax.summary

5. Élimination de séquences d'un certain taxon

* Tentative d'élimination des séquences *coxL*-forme II (BMS) qui n'a pas fonctionné. Ils ont du être éliminer manuellement à la fin des analyses.

```
remove.lineage(fasta=solcoxL.unique.good.filter.unique.precluster.pick.fasta,  
names=solcoxL.unique.good.filter.unique.precluster.pick.names, group=solcoxL.good.pick.groups,  
taxonomy=solcoxL.unique.good.filter.unique.precluster.pick.refcoxL.wang.taxonomy, taxon=BMS)
```

Génère les fichiers suivants :

- solcoxL.unique.good.filter.unique.precluster.pick.refcoxL.wang.pick.taxonomy
→renommé(final.taxonomy)
- solcoxL.unique.good.filter.unique.precluster.pick.pick.names
→renommé final.names
- solcoxL.unique.good.filter.unique.precluster.pick.pick.fasta
→renommé final.fasta
- solcoxL.good.pick.pick.groups
→renommé final.groups

5. Aperçu des séquences

```
summary.seqs(fasta=solcoxL.unique.good.filter.unique.precluster.pick.pick.fasta,  
name=solcoxL.unique.good.filter.unique.precluster.pick.pick.names)
```

D) Classification des séquences en OTUs

1. Calcul de distance

```
dist.seqs(fasta=final.fasta, cutoff=0.04, processors=2)
```

Génère le fichier suivant:

- final.dist

2. Assignment des séquences en OTUs

```
cluster(column=final.dist, name=final.names)
```

Génère les fichiers suivants:

- final.an.sabund
- final.an.rabund
- final.an.list

3. Génération du tableau d'OTUs

```
make.shared(list=final.an.list, group=final.groups, label=0.04)
```

shared file : Fichier qui indique le nombre de fois qu'un OTU se retrouve dans chacun des échantillons (Tableau d'OTUs)

Génère les fichiers suivants:

- final.an.shared
- final.an.(nom échantillon).rabund (chaque échantillon)

4. Dénombrement des séquences dans chaque échantillon

Affiche le nombre de séquences restantes dans chaque échantillon après les analyses.

```
count.groups(shared=final.an.shared)
```

5. Égalisation des séquences

```
sub.sample(shared=final.an.shared, size=641)
```

size : Nombre de séquences du plus petit échantillon pour égaliser tous les échantillons au même nombre de séquences

Génère les fichiers suivants:

- final.an.0.04.subsample.shared

6. Dénombrement des séquences dans chaque échantillon

S'assurer que tous les échantillons ont le même nombre de séquences

```
count.groups(shared=final.an.0.04.subsample.shared)
```

7. Assignment taxonomique des OTUs

```
classify.otu(list=final.an.list, name=final.names, taxonomy=final.taxonomy, label=2)
```

Génère les fichiers suivants:

- final.an.0.04.cons.taxonomy

Annexe 3: Commandes R - WGCNA

```
#####
##### THIS SCRIPT IS USED TO PERFORM MOLECULAR NETWORK ANALYSIS.#####
#####

# For new users, please install the package WGCNA, otherwise, start with Step 1.
# Automatic installation of WGCNA
source("http://bioconductor.org/biocLite.R")
biocLite("impute")
install.packages("WGCNA")

#####
# STEP 1 - Preparation of molecular and environmental datasets
# See the first chapter of Langfelder & Horvath, BMC Bioinformatics, 2008
#####

# Specification of the location of the data files
setwd("C:\\RBook\\")

# Load the package
library(WGCNA)
library(flashClust)
# The following setting is important, do not omit.
options(stringsAsFactors = FALSE)

# Read in the OTU distribution data set (we use the Order as taxonomic rank to delineate OTUs)
# We use one file per incubation condition
# The samples must be randomly sampled in order to get the same effective (number of reads) from
# one sample to the other and then avoid sampling bias.
# The dataset was prepared by keeping only OTUs found in 50% or more of the samples.
datExpr <- read.csv("16S_OTU_copiesgsol.csv", row.names=1)

# Look at the distribution of the samples to identify potential outliers
# This command will generate a dendrogram (UPGMA)
# We can thus remove samples if they are identified as outliers.
# This approach will be rarely applied in the lab because we will mostly have
# biologically independent triplicates.
sampletree <- flashClust(dist(datExpr), method = "average")
sizeGrWindow(12,9)
par(cex = 0.6);
par(mar = c(0,4,2,0))
plot(sampletree, main = "Sample clustering to detect outliers", sub="", xlab="", cex.lab = 1.5,
cex.axis = 1.5, cex.main = 2)

# Look at the environmental dataset.
# We explore the composition of our file and make sure the arrangement of variables and
# samples is correct.
datTraits = read.csv("coxL_OTU_copiesgsol.csv", row.names=1);
dim(datTraits)
names(datTraits)

# Merge and save both molecular and environmental datasets in one file.
```

```

# The software R will be able to distinguish both datasets in the next commands.
save(datExpr, datTraits, file = "coxL_16S_data.RData")

#####
# STEP 2 - Step-by-step network construction
#####

# Allow multi-threading within WGCNA. This helps speed up certain calculations.
# At present this call is necessary for the code to work.
# Any error here may be ignored but you may want to update WGCNA if you see one.
# Caution: skip this line if you run RStudio or other third-party R environments.
# See note above.
enableWGCNAThreads()

# Load the data saved in the first part
Inames = load(file = "coxL_16S_data.RData");

# The variable Inames contains the names of loaded variables.
# We can check that both datasets are distinguished by the following command:
Inames

#####
# STEP 2a - We set the threshold defining significant relationships between the different OTUs.
#####

# The structure of biological network, namely the distribution of the probabilities of
# node connectivities in a network, neither follows a normal distribution nor a Poisson
# distribution. We call it a "scale free topology" (réseaux invariants d'échelle). This
# concept is very well presented in an article (Barabasi & Albert, Science, 1999, 286:509-512).
# Briefly, the probability of a node to be connected with another one is not random and cannot
# be explained by classical statistical distribution. The authors used model datasets to show that
# "network growth" and "attachment preference" determines network structure. The distribution of
# probability
# that a node possess k links follows a power law distribution:  $P(k) = k(\text{EXP-Bêta})$ . In the next section, we
# need
# find the optimal Bêta value parameterizing our data according to a power law distribution. In principle,
# this
# Bêta value will fall between 2 and 4, but we encourage you to record the value that work best for your
# dataset!

# We will use a soft-thresholding approach. This is different from a "hard thresholding approach" where
# significance of correlation coefficients are used to delineate significant from insignificant
# pairwise interactions. The latter approach may lead to a loss of information (e.g. pairwise correlation of
# 0.80 used as hard threshold results in a loss of all pairwise correlations of 0.79 - even if this may be
# biologically meaningful). What we are doing here is basically a series of calculations into which the
# parameter Bêta
# is manipulated. After each run of calculation (distribution of threshold for each node connectivities as
# a function of a defined Bêta value), the software compares the theoretical distributions with the
# distribution
# of our real dataset. From a peculiar Bêta value, we should see convergence of the theoretical
# distribution with
# our own dataset.
#
# Choose a set of soft-thresholding powers
powers = c(c(1:10), seq(from = 12, to=20, by=2))

```

```

# Call the network topology analysis function
sft = pickSoftThreshold(datExpr, powerVector = powers, verbose = 5)

# Plot the results:
# Scale-free topology fit index as a function of the soft-thresholding power
# We are looking for the optimal Bêta value (displayed in the graph).
sizeGrWindow(9, 5)
par(mfrow = c(1,2))
cex1 = 0.9
plot(sft$fitIndices[,1], -sign(sft$fitIndices[,3])*sft$fitIndices[,2],
xlab="Soft Threshold (power)",ylab="Scale Free Topology Model Fit,signed R^2",type="n",
main = paste("Scale independence"));
text(sft$fitIndices[,1], -sign(sft$fitIndices[,3])*sft$fitIndices[,2],
labels=powers,cex=cex1,col="red");

# this line corresponds to using an R^2 cut-off of h
abline(h=0.90,col="red")

# It is possible to explore network topology as a function of the
# Bêta value utilized to parameterize the distribution of our data according
# to the scale-free distribution.
# Mean connectivity as a function of the soft-thresholding power.
plot(sft$fitIndices[,1], sft$fitIndices[,5],
xlab="Soft Threshold (power)",ylab="Mean Connectivity", type="n",
main = paste("Mean connectivity"))
text(sft$fitIndices[,1], sft$fitIndices[,5], labels=powers, cex=cex1,col="red")

#####
# STEP 2b - Calculation of adjacent matrix
#####

# Here, we will calculate the adjacent matrix (pairwise correlation in the relative abundance of each OTU).
# This matrix will be used for network construction.
# The parameter "softpower" to be entered in the function corresponds to the Bêta value calculated
# in the previous step. In contrast to hard thresholding where correlations values below a critical
# threshold are converted to adjacency of "zero" in the matrix, the soft thresholding will calculate a
# value between 0 and 1 (the spearman correlation coefficient if elevated to the power Bêta). Although
# this is a "relaxed thresholding approach", we keep track the strength of the interaction and we can
# represent this information in Cytoscape by varying the thickness of the links between the node as
# a function of the adjacency coefficient.
softPower <- 12
adjacency <- adjacency(datExpr, type = "signed", corFnc = "cor", corOptions = "use = 'p', method =
'spearman'", power = softPower)

# We should mention here that Pearson correlation (the default argument of the last command)
# is not working. This might be due to the fact that we have some zero values in our dataset,
# resulting to non normal distribution of our dataset (if this is true, then it means that R can
# avoid misuse of pairwise correlation methodology. I would prefer this non parametric correlation
# anyway because we don't have a normal distribution (don't forget that we are working in a
# scale free environment and thus no rational for supposing or adjusting our data for a normal
# distribution). We should verify that in the future, but I feel this is a correct approach.

# We will use the function "scalefreeplot" to verify if the adjacency matrix we calculated
# resulted to a scale free topology. We will get a regression coefficient of approximated

```

```

# scale free topology. We should notice that scale free topology might not be observed in
# the presence of were different samples or in the presence of outliers. It is important
# to check are good is the fit and then see if we can pursue the analysis described below.
hist(adjacency)
scaleFreePlot(adjacency, nBreaks = 10, truncated = FALSE, removeFirst = FALSE,
main = "Check of scale free topology")

#####
# STEP 2c - Module detection
#####

# We will now detect subsets of nodes that are tightly connected to each other (i.e. module
# detection. We will measure the
# Turn adjacency into topological overlap
TOM = TOMsimilarity(adjacency)

# We calculate the distance in order to get topological overlap based dissimilarity measure
dissTOM = 1-TOM

# Call the hierarchical clustering function
# Here, we group the nodes according to their similarities in term of adjacency coefficient. If
# nodes are displaying similar adjacency profiles, then they are connected together.
geneTree = flashClust(as.dist(dissTOM), method = "average")

# Plot the resulting clustering tree (dendrogram)
sizeGrWindow(12,9)
plot(geneTree, xlab="", sub="", main = "OTU clustering on TOM-based dissimilarity",
labels = FALSE, hang = 0.04)

# We would like large modules, so we set the minimum module size relatively high. We
# could also keep smaller modules but their interpretation might be difficult.
minModuleSize = 10

# Module identification using dynamic tree cut:
dynamicMods = cutreeDynamic(dendro = geneTree, distM = dissTOM,
deepSplit = 2, pamRespectsDendro = FALSE,
minClusterSize = minModuleSize);
table(dynamicMods)

# After this command, we get the number of detected modules as well as the number
# of node into each module.

# Convert numeric lables into colors (each module is assigned to a color)
dynamicColors = labels2colors(dynamicMods)
table(dynamicColors)

# Plot the dendrogram and colors underneath. We can visualize the composition of
# the network we selected with the cutoff value (10 in this script) as well as all the others
# comprising less nodes.
sizeGrWindow(8,6)
plotDendroAndColors(geneTree, dynamicColors, "Dynamic Tree Cut",
dendroLabels = FALSE, hang = 0.03,
addGuide = TRUE, guideHang = 0.05,
main = "Gene dendrogram and module colors")

```

```

#####
# STEP 2d - Polishing module detection
#####

# We can then calculate a vector represening each module (i.e. eigengenes)
# In the end of this command, we will vusualize the modules in independant clusters.
# In the case of clusters encompassing more than one module, we may merge them as
# their nodes are highly correlated.
# Calculate eigengenes
MEList = moduleEigengenes(datExpr, colors = dynamicColors)
MEs = MEList$eigengenes

# Calculate dissimilarity of module eigengenes
MEDiss = 1-cor(MEs);

# Cluster module eigengenes
METree = flashClust(as.dist(MEDiss), method = "average");

# Plot the result
sizeGrWindow(7, 6)
plot(METree, main = "Clustering of module eigengenes",
xlab = "", sub = "")

# If we want to merge clusters, we must specify a cutoff value (this value corresponds to
# 1-correlation coefficient). Let's fix a cutoff of 0.25 (corresponding to a correlation coefficient
# of 0.75).
MEDissThres = 0.45

# Plot the cut line into the dendrogram
abline(h=MEDissThres, col = "red")

# Call an automatic merging function
merge = mergeCloseModules(datExpr, dynamicColors, cutHeight = MEDissThres, verbose = 3)

# The merged module colors
mergedColors = merge$colors

# Eigengenes of the new merged modules:
mergedMEs = merge$newMEs

# We can visualize the new modules with the following function:
sizeGrWindow(12, 9)

#pdf(file = "Plots/geneDendro-3.pdf", wi = 9, he = 6)
plotDendroAndColors(geneTree, cbind(dynamicColors, mergedColors),
c("Dynamic Tree Cut", "Merged dynamic"),
dendroLabels = FALSE, hang = 0.03,
addGuide = TRUE, guideHang = 0.05)
#dev.off()

#####
# STEP 2e - Save data for further analyses
#####

# In the subsequent analysis, we will use the merged module colors in mergedColors.

```

```

# Rename to moduleColors
moduleColors = mergedColors

# Construct numerical labels corresponding to the colors
colorOrder = c("grey", standardColors(50));
moduleLabels = match(moduleColors, colorOrder)-1;
MEs = mergedMEs;

# Save module colors and labels for use in subsequent parts
# In this file, MEs = MEList$eigengenes, moduleLabels = module color, geneTree = modules
# calculated from adjacency matrix. The file thus includes ALL MODULES, so we don't lose
# all information regarding smaller potential modules
# ***** WE CONFIRM THAT LATER WHEN WE WILL USE CYTOSCAPE *****).
save(MEs, moduleLabels, moduleColors, geneTree, file = "coxL_16S-networkConstruction-
stepByStep.RData")

#####
# STEP 3 - Association of module eigengene with environmental variables
#####

# Load the expression and trait data saved in the first part
Inames = load(file = "coxL_16S_data.RData")

#The variable Inames contains the names of loaded variables.
Inames

# Load network data saved in the second part.
Inames = load(file = "coxL_16S-networkConstruction-stepByStep.RData")
Inames

#In this analysis we would like to identify modules that are significantly associated
# with the measured environmental variables. Since we already have a summary profile (eigengene)
# for each module, we simply correlate eigengenes with external variables and look
# for the most significant associations.
# Define numbers of genes and samples
nGenes = ncol(datExpr);
nSamples = nrow(datExpr);
# Recalculate MEs with color labels
MEs0 = moduleEigengenes(datExpr, moduleColors)$eigengenes
MEs = orderMEs(MEs0)
# Previous command where I used Pearson correlation. In this work, Spearman correlation
# is more suitable because the distribution of certain OTU does not follow a normal
# distribution. Previous command: moduleTraitCor = cor(MEs, datTraits, use = "p")
moduleTraitCor = cor(MEs, datTraits, method = "spearman")
moduleTraitPvalue = corPvalueStudent(moduleTraitCor, nSamples)

# Since we have a moderately large number of modules and traits, a suitable
# graphical representation will help in reading the table. We color code
# each association by the correlation value
sizeGrWindow(10,6)

# Will display correlations and their p-values
textMatrix = paste(signif(moduleTraitCor, 2), "\n",
signif(moduleTraitPvalue, 1), ") ", sep = "")
dim(textMatrix) = dim(moduleTraitCor)

```

```

par(mar = c(6, 8.5, 3, 3))

# Display the correlation values within a heatmap plot
labeledHeatmap(Matrix = moduleTraitCor,
xLabels = names(datTraits),
yLabels = names(MEs),
ySymbols = names(MEs),
colorLabels = FALSE,
colors = greenWhiteRed(50),
invertColors = TRUE,
textMatrix = NULL,
setStdMargins = FALSE,
cex.text = 0.1,
cex.lab = 0.5,
xLabelsAngle = 90,
zlim = c(-1,1),
main = paste("Module-trait relationships"))

# The analysis identifies the several significant module {environmental variable associations.
# We will concentrate on coxL OTU 2 (i.e. X2) as the trait of interest.
# We quantify associations of individual OTUs with our variable of interest (X2)
# by defining OTU Significance GS as (the absolute value of) the correlation between
# the OTU and the variable X2. For each module, we also define a quantitative measure of
# module membership "MM" as the correlation of the module eigengene and the OTU
# distribution profile. This allows us to quantify the similarity of all detected OTUs
# to every module.
## Define variable weight containing the X2 column of datTrait
weight = as.data.frame(datTraits$X2)
names(weight) = "X2"
# names (colors) of the modules
modNames = substring(names(MEs), 3)
geneModuleMembership = as.data.frame(cor(datExpr, MEs, use = "p"))
MMPvalue = as.data.frame(corPvalueStudent(as.matrix(geneModuleMembership), nSamples))

names(geneModuleMembership) = paste("MM", modNames, sep="")
names(MMPvalue) = paste("p.MM", modNames, sep="")

geneTraitSignificance = as.data.frame(cor(datExpr, weight, use = "p"))
GSPvalue = as.data.frame(corPvalueStudent(as.matrix(geneTraitSignificance), nSamples))

#Using the GS and MM measures, we can identify genes that have a high significance
# for X2 as well as high module membership in interesting modules. As an
# example, we look at the XYZ (brown in the script) module that has the highest association with
# X2. We plot a scatterplot of Gene Significance vs. Module Membership in the
# brown module:
module = "blue"
column = match(module, modNames);
moduleGenes = moduleColors==module;
sizeGrWindow(7, 7);
par(mfrow = c(1,1));
verboseScatterplot(abs(geneModuleMembership[moduleGenes, column]),
abs(geneTraitSignificance[moduleGenes, 1]),
xlab = paste("Module Membership in", module, "module"),
ylab = "OTU significance for X2",
main = paste("Module membership vs. gene significance\n"),

```

```

cex.main = 1.2, cex.lab = 1.2, cex.axis = 1.2, col = module)

#####
# STEP 4 - Data exportation for analysis using Excel
#####

# We have found modules with high association with our trait of interest, and have
# identified their central players by the Module Membership measure. We now merge
# this statistical information with gene annotation and write out a file that summarizes
# the most important results and can be inspected in standard spreadsheet software such as
# Excel. Our OTU data are only annotated by Taxonomic ID names: the command
# If we want to see the names we have in our dataset, we type:
names(datExpr)
# Alternatively, a list of the name belonging to the blue module can be obtained:
names(datExpr)[moduleColors=="turquoise"]

# To facilitate interpretation of the results, we use a OTU annotation file
# generated during the taxonomic identification of representative OTUs during sequence
# classification using QIIME.
annot = read.csv(file = "16S_taxonomy_Isabelle.csv", row.names=1)
dim(annot)
names(annot)
probes = names(datExpr)
probes2annot = match(probes, annot$"Phyla")
# The following is the number of probes without annotation:
sum(is.na(probes2annot))

# We now create a data frame holding the following information for all OTUs: Phyla,
# Order, module color, gene significance for weight, and module membership and
# p-values in all modules. The modules will be ordered by their significance for
# weight, with the most significant ones to the left.
# Create the starting data frame
geneInfo0 = data.frame(OTU = names(datExpr),
moduleColor = moduleColors,
geneTraitSignificance,
GSPvalue)
# Order modules by their significance for weight
modOrder = order(-abs(cor(MEs, weight, use = "p")));
# Add module membership information in the chosen order
for (mod in 1:ncol(geneModuleMembership))
{
oldNames = names(geneInfo0)
# Order the genes in the geneInfo variable first by module color, then by geneTraitSignificance
geneOrder = order(geneInfo0$moduleColor, -abs(geneInfo0$GS.weight))
geneInfo = geneInfo0[geneOrder, ]

# This data frame can be written into a text-format spreadsheet, for example by:
write.csv(geneInfo0, file = "OTUInfo2.csv")

#####
# STEP 5 - Network visualization
#####
# In the next steps, we will get dendrogram showing how each node is specific

```

```

# We use the dissTOM calculated before, i.e. right after the calculation of the adjacency matrix.
# Transform dissTOM with a power to make moderately strong connections more visible in the heatmap
plotTOM = dissTOM^7
# Set diagonal to NA for a nicer plot
diag(plotTOM) = NA
# Call the plot function
sizeGrWindow(9,9)
TOMplot(plotTOM, geneTree, moduleColors, main = "Network heatmap plot, all OTUs")

nSelect = 50
# For reproducibility, we set the random seed
set.seed(10);
select = sample(nGenes, size = nSelect);
selectTOM = dissTOM[select, select];
# There's no simple way of restricting a clustering tree to a subset of genes, so we must re-cluster.
selectTree = flashClust(as.dist(selectTOM), method = "average")
selectColors = moduleColors[select];
# Open a graphical window
sizeGrWindow(9,9)
# Taking the dissimilarity to a power, say 10, makes the plot more informative by effectively changing
# the color palette; setting the diagonal to NA also improves the clarity of the plot
plotDiss = selectTOM^7;
diag(plotDiss) = NA;
TOMplot(plotDiss, selectTree, selectColors, main = "Network heatmap plot, selected genes")

```

Annexe 3: Commandes R - Homogeneity Multivariate Dispersion

```
library(vegan)

# Specification of file location
setwd("C:\\RBook\\")

# Import the data from CSV files
env <- read.csv("coxL.csv", row.names=1)

#####
##### Step 1, Emphasis on the distribution of the proportion of each species within replicated experimental plots and
##### between treatments
#####

# Hellinger transformation of dataset (owing to several "zero" values)
# In this case, we will compare the proportion of OTUs in the samples.
spe.dbln <- decostand(env, "hel")

# We calculate Euclidean distance between samples on hellinger-transformed dataset.
dis <- dist(spe.dbln, "euc")

# Definition of the groups
groups <- factor(c(rep(1,3), rep(2,3), rep(3,3), rep(4,3)), labels = c("4","22","36","55"))

# Calculate multivariate dispersions
mod <- betadisper(dis, groups, type = "centroid")
mod

## S3 method for class 'betadisper':
boxplot(mod, ylab = "Distance to centroid")

## S3 method for class 'betadisper':
scores(mod, display = c("sites", "centroids"), choices = c(1,2))

## Perform test
anova(mod)

## Permutation test for F
permute(mod, pairwise = TRUE)

## Tukey's Honest Significant Differences
mod.HSD <- TukeyHSD(mod)
plot(mod.HSD)

## S3 method for class 'betadisper':
plot(mod, axes = c(1,2), cex = 0.7, hull = TRUE)
```

