UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

UMR 7156 Génétique Moléculaire Génomique, Microbiologie

THÈSE présentée par :

Pauline CHAIGNAUD

soutenue le : **29 Juin 2016**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline : Sciences du vivant

Spécialité : Aspects moléculaires et cellulaires de la biologie

Le rôle des bactéries dans le filtrage du chlorométhane, un gaz destructeur de la couche d'ozone – des souches modèles aux communautés microbiennes de sols forestiers

Bacteria as chloromethane sinks – from model strains to forest soil communities

THÈSE dirigée par :

Mme. BRINGEL Françoise	Directrice de recherche, Université de Strasbourg, France
M. KOLB Steffen	Docteur, Leibnizzentrum für Agrarlandschaftsforschung (ZALF) und Universität Bayreuth, Allemagne

RAPPORTEURS :

Mme. LAUGA Béatrice	Professeure, Université de Pau, France
M. HORN Marcus	Docteur, Universität Bayreuth, Allemagne

AUTRES MEMBRES DU JURY :

M. POTIER Serge	Professeur, Université de Strasbourg, France
Mme. KNIEF Claudia	Professeure, Universität Bonn, Allemagne

Ce travail de thèse est une cotutelle entre l'Université de Strasbourg et l'Université de Bayreuth. Il a été réalisé respectivement dans l'équipe AIME dirigée par le Professeur Stéphane Vuilleumier au sein du laboratoire de Génétique Moléculaire, Génomique et Microbiologie, ainsi qu'au sein du département d'écologie microbienne (EMIC) dirigé par le Professeur Harold Drake à Bayreuth.

Je tiens à remercier les membres de mon jury de thèse ; le Pr. Serge Potier, Dr. Marcus Horn, Pr. Béatrice Lauga et le Pr. Claudia Knief, d'avoir accepté d'évaluer mon travail.

Je remercie également mes 2 directeurs de thèse pour leur encadrement, le Dr. Françoise Bringel et le Dr. Steffen Kolb. Françoise, merci pour tes nombreuses remarques et critiques qui m'ont permis d'avancer et d'aller au bout de mon cursus universitaire. Steffen, merci pour ton accueil à Bayreuth et pour ton aide dans la réalisation de ce projet.

Merci aux membres de l'équipe AIME pour son soutien tout au long de ma thèse et d'autant plus ces derniers mois. Stéphane, merci de m'avoir fait confiance lors de mon stage de master, et montré que je pouvais me dépasser. Sans ce stage, je n'en serai pas là aujourd'hui. Cette expérience n'aurait pas été possible sans ton aide, ton écoute, et ton soutien à toute épreuve.
Un grand merci à Thierry pour m'avoir proposé de faire mon stage de master au sein de l'équipe. Merci pour ta disponibilité, tes conseils et les nombreuses corrections de mon mémoire. Tes blagues sur mon choix de partir en thèse vont me manquer !
Christelle, je ne te remercierai jamais assez pour tout surtout pour tes blagues et clichés sur Lille et le Nord… Je pourrai faire un chapitre complet de remerciements !
Merci à Yousra, ton grain de folie, tes chansons chantées en yaourt et tes prises de kung-fu au milieu de couloir me feront toujours rire !

Merci aux post-docs, et thésards rencontrés pendant toutes ces années dans l'équipe. Sabrina, ma compatriote du Nord, merci pour ton aide, ton écoute et ta bonne humeur. Merci de m'avoir permis de me remémorer des monuments de la chanson française injustement oubliés ; Larusso, Ophélie Winter….
Merci à Bruno, pour ta présence au quotidien (et le soir à table). Qui va te saouler, maintenant, avec des débats qui se terminent toujours en monologue ?! Merci pour ta disponibilité et ton aide. Ludo, merci pour ton aide et surtout ton sourire qui illuminait notre bureau tous les jours…
Merci également aux « anciens » thésards de l'équipe, Fahran et Omniae pour leurs conseils et Louis, Panyu, Elova (et ses Elovades…) ainsi qu'à Emmanuel.
Merci également à Jackson, Jean-Seb, Christian, David et Julien pour leur aide, soutien et les bières après le labo.

Merci à toute l'équipe EMIC pour son accueil et sa bienveillance. Un merci tout particulier à Mareen, pour son aide, ses conseils et sa disponibilité lors des expériences de SIP. Merci aussi à Adam et Sindy pour avoir pris le temps de me faire visiter la ville de Bayreuth.

Enfin, merci à mes proches pour leur présence ces dernières années, mes parents, Charlotte, Antoine, Caro, Brigitte, Lucile et Sophie sans vous je n'en serai pas là aujourd'hui.

Diese Diplomarbeit entstand mit der Unterstützung der Universität Straßburg und der Universität Bayreuth .
Sie wurde realisiert durch das AIME Team von dem Professor des Labors für molekulare Genetik, Genomik und Mikrobiologie Stephane Vuilleumier, sowie Professor Harold Drake, Leiter der Abteilung für mikrobielle Ökologie (EMIC) in Bayreuth.

Ich möchte den Mitgliedern meines Bewertungskomitees danken; Professor Serge Potier, Doktor Marcus Horn, Professor Béatrice Lauga und Professor Claudia Knief, die sich bereiterklärt haben meine Arbeit zu bewerten.

Ich danke auch meinen beiden Betreuern für ihre Führung, Doktor Françoise Bringel und Doktor Steffen Kolb.
Françoise, danke für die vielen Kommentare und die Kritik, die mich weitergebracht haben und es mir ermöglicht haben meinen Studiengang an der Universität zu beenden.
Steffen, vielen Dank mich in Bayreuth willkommen geheißen zu haben und für deine Hilfe an diesem Projekt.

Vielen Dank an die Mitglieder des AIME Teams für ihre Unterstützung während meiner Diplomarbeit und vor allem in den letzten Monaten.
Stéphane, danke für dein Vertrauen in mich während meiner Masterarbeit, das mir zeigte mich selbst zu übertreffen. Ohne dieses Referendariat wäre ich heute nicht hier. Diese Erfahrung wäre nicht möglich gewesen ohne deine Hilfe, dein aufmerksames Ohr und deine unermüdliche Unterstützung.
Ein großes Dankeschön an Thierry für die Einladung meinen Master innerhalb des Teams abzuleisten.
Vielen Dank für deine Verfügbarkeit, deine Beratung und zahlreiches „auf die Sprünge helfen“ meines Gedächtnisses. Deine Witze über meine Wahl eine Diplomarbeit zu schreiben werden mir fehlen.

Christelle, ich könnte dir niemals genug danken vor allem für deine Witze und Klischees über Lille und den Norden ... Ich könnte ein ganzes Kapitel mit Danksagungen vollschreiben!

Vielen Dank an Yousra, deine Verrücktheit, deine in Joghurt gesungen Lieder und deine Kung-Fu-Vorstellungen inmitten des Flurs, die mich immer zum Lachen bringen werden!

Vielen Dank an die „Post-Docs“ und Doktoranden, die sich all die Jahre im Team trafen.
Sabrina, meine Kollegin des Nordens, vielen Dank für deine Hilfe, dein Ohr und deine gute Laune. Vielen Dank, mich an die zu Unrecht vergessenen, französischen Lieder erinnert zu haben ; Larusso, Ophélie Winter ....
Vielen Dank an Bruno für deine tägliche Anwesenheit (und Abends am Tisch). Wer wird dich jetzt mit Diskussionen nerven, die immer im Monolog enden? Vielen Dank für deine Verfügbarkeit und deine Hilfe.
Ludo, ich danke dir für deine Hilfe und vor allem dein Lächeln, das unser Büro jeden Tag erleuchtete.
Vielen Dank an das "alte" Team von Doktoranden, Farhan und Omniae für ihre Beratung und Louis, Panyu, EloVA (und seine Elovades ...), sowie Emmanuel.
Danke auch an Jackson, Jean-Seb, Christian, David und Julian für ihre Hilfe, Unterstützung und das Bier nach der Arbeit im Labor.

Vielen Dank an das EMIC Team für ihre Gastfreundschaft und Freundlichkeit.
Ein ganz besonderer Dank an Mareen, für ihre Hilfe, Beratung und Verfügbarkeit bei SIP-Experimenten.
Vielen Dank ebenfalls an Adam und Sindy, die sich die Zeit genommen haben, mir die Stadt Bayreuth zu zeigen.
Abschließend danke Ich meiner Familie für ihre Anwesenheit in den letzten Jahren, meinen Eltern, Charlotte, Anthony, Caro, Brigitte, Lucile und Sophie. Ohne euch würde ich heute nicht hier stehen.

Table of contents

List of abbreviations

aa	Amino acid
ADH:	Alcohol dehydrogenase
ADN/DNA	*Acide desoxyribonculéique* /deoxyribonucleic acid
ADP	Adenosine diphosphate
ANAH	Adenine Nucleotide Alpha Hydrolase
ARN/RNA	*Acide ribonucléique* /ribonucleic acid
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
bp	Base pair
BSA	Bovine serum albumin
cDNA	Complementary deoxyribonucleic acid
CDS	Coding DNA sequence
CM	*Chlorométhane* /chloromethane
Cmu	Chloromethane utilization
CoA	Coenzyme A
COV/VOC	*Composé organique volatile* /volatile organic compound
CsCl	Cesium chloride
CTAB	Cetyl trimethylammonium bromide
DCM	*Dichlorométhane* /dichloromethane
ddNTP	Dideoxynucleotide triphosphate
DEPC	Diethyl pyrocarbonate
DESeq	Differential expression analysis for sequence count data
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleotide triphosphate
DMB	Dimethylbenzimidazole
ECD	Electron capture detector
EDTA	Ethylenediaminetetraacetic acid
FDH	Formate dehydrogenase
FDR	False discovery rate
GC	Gas chromatography
GC-MS	Gas chromatography-mass spectometry
GEI	Genomic islet
GST	Glutathion S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOL	Harmless to ozone layer
IARC	International Agency for Research on Cancer
IS	Insertion sequence
kb	Kilo base pair
KEGG	Kyoto Encyclopedia of Genes and Genomes
log2fc	log2 fold-change
M3	Mineral media for *Methylobacterium*
MADH	Methylamine dehydrogenase
MaGe	Magnifying Genomes
Mb	Mega base pair
Mchl	*Methylobacterium chloromethanicum*
MDH	Methanol dehydrogenase
MEGA	Molecular Evolutionary Genetics Analysis
METDI	*Methylobacterium extorquens* dichloromethane
MNNG	N-methyl-N'-nitrosoguanidine
mRNA	Messenger Ribonucleic acid
MS	Methanesulfonate d'éthyle
NAD(H)	Nicotinamide adenine dinucleotide
NADP (H)	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
nd	Not determined
NGS	Next generation sequencing
NJ	Neighbour-joining
NMDS	Non-metric multidimensional scaling
nt	Nucleotide
OD	Optical density
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PHB	Poly-beta-hydroxybutyrate
PICRUST	Phylogenetic investigation of communities by reconstruction of unobserved states
PkGDB	Prokaryotic Genome DataBase

pMMO	Particulate methanol monooxygenase
ppb	Part per billion
PPi	Pyrophosphate inorganique
ppt	Part per trillion
Pptv	Part per trillion by volume
PQQ	Pyrroloquinoline quinone
qPCR	Quantitative polymerase chain reaction
RefSeq	Reference Sequence collection
RIN	RNA integrity number
RNase	Ribonuclease
RNAseq	Ribonucleic acid sequencing
RPM	Rotation per minute
rRNA	Ribosomal ribonucleic acid
RT-qPCR	Reverse transcription-quantitative polymerase chain reaction
RuBP	Ribulose biphosphate
RuMP	Ribulose monophosphate
SAH	S-adenosyl L-homocystein
SAM	S-adenosyl L-methionine
SBS	Sequencing by synthesis
SIP	Stable isotope probing
sMMO	Soluble methanol monooxygenase
SMRT	Single molecule real time
ssDNA	Single-stranded deoxyribonucleic acid
STAMP	Statistical analysis of metagenomic profiles
TAMARA	Transcriptome analyses based on massive sequencing of RNAs
TCA	Tricarboxylic acid cycle
TCD	Thermal conductivity detector
TCM	Tetrachloromethane
TE	Tris-EDTA buffer (10 mM Tris-HCl, 1 mM disodium EDTA, pH 8.0)
UV	Ultraviolet
WGS	Whole genome sequencing

List of figures

Chapter 1 (in English) Introduction

Chapitre 1 (en français) Introduction

Chapter 5 Conclusions and perspectives

List of tables

Chapter 1

Introduction

Chapter 1. Introduction (in English)

This chapter begins with the general information on chlorinated methanes and chloromethane in particular. Chlorinated methane chemical properties and industrial use as well as their impact on human health and the environment will be described. The following paragraphs are dedicated to chloromethane, including its production (sources) and consumption (sinks), with a focus on the role of microorganisms. Knowledge on microbial degradation of chloromethane will be presented in the context of one carbon compound metabolism and the genetics of methylotrophic model strains. The last paragraph will explain the aims and objectives of my PhD thesis.

1. Chlorinated methanes

Chlorinated methanes are one-carbon organic compounds with one or more chlorine atoms. A carbon atom has four covalent bonds. Four chlorinated methanes are found: chloromethane (CH_3Cl also called methyl chloride), dichloromethane (CH_2Cl_2), chloroform ($CHCl_3$) and tetrachloromethane (CCl_4). Of those, only chloromethane is in a gaseous state at room temperature and one atmosphere pressure. Other chlorinated methanes are liquid and colorless (Huang *et al*., 2014). These compounds have different physical and chemical properties and are, since the 20^{th} century, produced and used in large quantities by the industry (Table 1.1). Chlorinated methanes are also produced naturally. Natural production of chloroform and dichloromethane are estimated at 0.56 and 0.25 $Tg.year^{-1}$, respectively (Cox *et al*., 2003; Xiao, 2008). However, chloromethane is the most abundant chlorinated methane found in the atmosphere with 2.8 $Tg.year^{-1}$ (Ruecker *et al*., 2014) (Table 1.2).

Table 1.1. Chlorinated methane utilizations and characteristics

Compound	State	Water solubility ($g.L^{-1}$)	Effect on health[a]	Pollutant[b]	Industrial utilization[c]
chloromethane	colorless gas	5.3	group 3	no	silicone production, construction products
dichloromethane	colorless liquid	13.0	group 2B	yes	paint stripping, pharmaceuticals products
chloroform	colorless liquid	8.0	group 2B	yes	solvents, chemicals synthesis pharmaceuticals
tetrachloromethane	colorless liquid	0.8	group 2B	yes	products, cleaning products, agrochemicals

[a] Data from the International Agency Research against Cancer (IARC). Group 2B contains compounds probably carcinogenic to humans. Group 3 consists of compounds not classified such as carcinogenic for humans (Huang *et al*., 2012).
[b] Compounds recognized as pollutants by the United States, China. Dichloromethane and chloroform are also recognized as pollutant by the European Union (Huang *et al*., 2014).
[c] Data from Frascari *et al*. , 2015

Chlorinated methanes are present in soil, water and air (http://www.epa.gov). With the exception of chloromethane, they have been classified by the International Agency for Research on Cancer (IARC) as possibly carcinogenic, and are considered as pollutants in many countries (Table 1.1). The impact of chlorinated compounds on the environment has been extensively documented (Harper, 2000).

Table 1.2. Chlorinated methanes in the troposphere

	Troposphere		Flow (Gg. year^{-1})	
	Conc. (pptv)	Half-life (year)		Conc. (pptv)
chloromethane	550 ± 30[a]	1.3	chloromethane	550 ± 30[a]
dichloromethane	25 ± 5[d]	0.38	dichloromethane	25 ± 5[d]
chloroform	15 ± 5[c]	0.41	chloroform	15 ± 5[c]
tetrachloromethane	95 ± 5[f]	26	tetrachloromethane	95 ± 5[f]

From S. Roselli PhD thesis, 2009
[a] Montzka and Fraser, 2003
[b] McCulloch *et al.*, 2003
[c] O'Doherty *et al*., 2001
[d] Schauffler, 2003
[e] Kurylo and Rodriguez, 1998
[f] Prinn *et al*., 2000
[g] UNEP, 2005

In the stratosphere, chlorinated methanes are activated by ultraviolet radiation. This process results in the formation of radical $Cl^{\cdot}$, which will contribute to the depletion of the ozone layer by catalyzing the formation of O_2 (Figure 1.1). It is estimated that one chlorine atom can destroy 100,000 ozone molecules (http://www.epa.gov/ozone/science/process.html). An example of this impact is the formation of a hole in the ozone layer over the Antarctic every spring (Clerbaux *et al*., 2007).

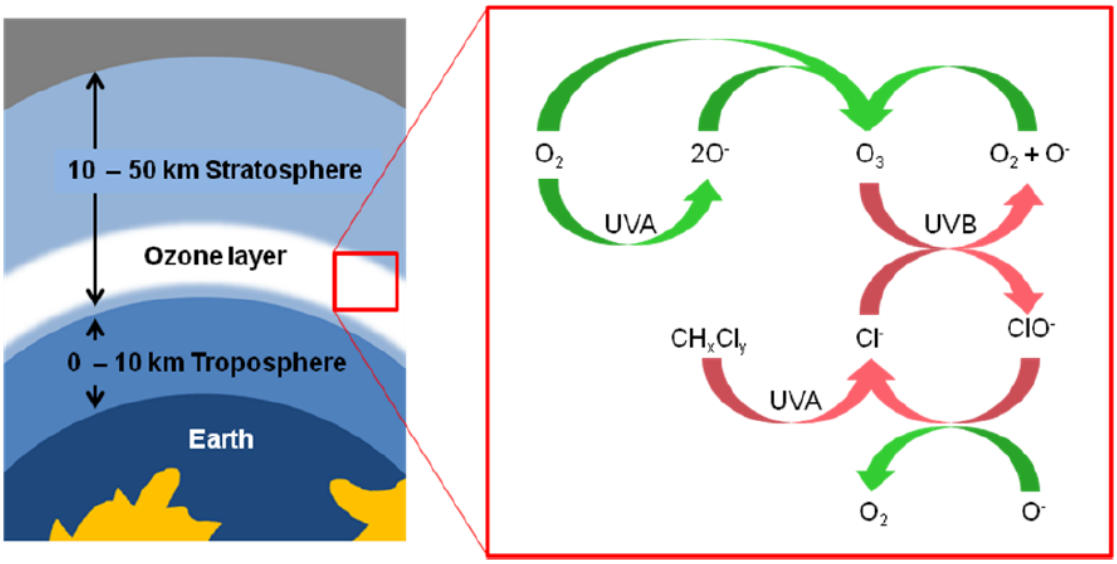


Figure 1.1. Ozone production and degradation reactions in the atmosphere
In the presence of chlorinated methanes (CHxClx), natural reactions of ozone (O_3) are disrupted. UVA between 315 nm and 400 nm and UVB radiation from 280 nm to 315 nm (Figure modified from http://www.cec.org).

Because of their impact on the environment, the use of halogenated compounds has been regulated by the Montreal Protocol in 1987, thus significantly reducing their use in the industry, up to a total cessation of production of compounds such as carbon tetrachloride in 1995. This resulted in the augmentation of the share of chloromethane, the chlorinated C_1, compound the most produced naturally in the environmental impact of chlorinated methanes. Indeed, from chlorinated methanes found in the troposphere, chloromethane is the most abundant (Table 1.2). Today it is estimated that 16 % of the degradation of the ozone layer is due to chlorinated compounds (Montzka *et al.*, 2011).

2. Chloromethane

The overall budget of chloromethane, or its concentration in the atmosphere, is determined by the sum of its production sources, subtracted from all sinks dissipation. The diversity of abiotic and biotic sources and sinks of chloromethane is detailed below. This budget remains uncertain mainly because of the unknown quantitative impact of bacteria on the biotic filtering of chloromethane emitted from terrestrials and other environments.

2.1. Sources of chloromethane

Total annual production is estimated at 2.8 Tg (Ruecker *et al.*, 2014; Sailaukhanuly *et al.*, 2014). As more than 5,000 halogenated substances, chloromethane is produced predominantly

naturally. The natural production of chloromethane based on abiotic and biotic sources (Table 1.3). Its low production of anthropogenic origin is due to the burning of fossil fuels such as coal, the combustion of municipal wastes, and other industrial activities (Table 1.3, McCulloch and Aucott, 1999).

Table 1.3. Estimation of chloromethane sources in the atmosphere

Type	Estimation (Gg. year^{-1})[a]	Low/high values (Gg. year^{-1})[b]
Plant senescence	1,800[d]	30/2500
Forest fires	911	655/ 1,125
Tropical plants	910	820/ 8,200
Oceans	600	325/ 1,300
Salt marshes	170	65/ 440
Fungi	160	43/ 470
Combustion of fossil organic matters	105	5/ 205
Incineration	45	15/ 75
Wetlands	40	6-270[c]
Industrial	10	10
Rice fields	5	2,4/ 4,9
Total sources	*4,746*	*2,019/ 1,4378*

[a] Data from Keppler *et al.*, 2005
[b] Data from Clerbaux *et al*., 2007
[c] Data from Montzka and Fraser, 2003
[d] Data from Montzka *et al*., 2011

2.1.1. Abiotic production

Forest fires are the largest source of abiotic chloromethane (Sailaukhanuly *et al*., 2014). The produced chloromethane is released into the atmosphere, but is also found in the soil surface, where abiotic processes of the production of halogenated methanes have also been characterized (Keppler *et al*., 2005). Abiotic production of chloromethane by redox reactions (Figure 1.2) and substitution reactions were demonstrated (Hamilton, 2003; Keppler *et al*., 2005; Wishkerman *et al*., 2008) although their share in budget chloromethane was not

assessed. The oceans represent the second source of abiotic chloromethane (Xiao, 2008; Hu *et al*., 2010).

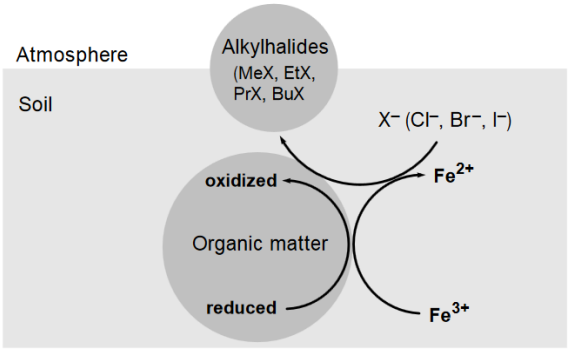


Figure 1.2. Abiotic chloromethane production in soils
Halogenated alkyl formation by the reduction of Fe (III) with the organic material in the presence of chloride ions in soils (from Keppler *et al*., 2003).

2.1.2. Biotic production

The main production of chloromethane is the result of biotic processes in plants (Hamilton, 2003; Saito *et al.*, 2008), fungi (Moore *et al.*, 2005), seaweed (Traunecker *et al.*, 1991) or salt marshes (Rhew *et al*., 2003) (Table 1.3). Terrestrial ecosystems contribute significantly to this production, with more than 50% of emissions of chloromethane derived from them (Redeker and Kalin, 2012). In the environment, the presence of chloride ions is highly variable and may influence the emission rates of chloromethane. The first results of a possible correlation are mentioned in the literature and are described below, as well as the current knowledge of chloromethane production by plants and in soils. Also explained below is the emission rate of chloromethane as compared to other Volatile Organic Compound (VOC) and the development of a bio-reporter for its quantification.

– Production by plants

The genetic mechanism of production of chloromethane was characterized in detail in the model plant *Arabidopsis thaliana* (Rhew *et al*., 2003). This study shows the production of chloromethane but also bromomethane (CH_3Br) and iodomethane (CH_3I), by the action of a S-Adenosyl Methionine (SAM)-dependent methyltransferase, named HOL for "Harmless to Ozone Layer". The HOL gene is conserved in plants, suggesting that the production of chloromethane is a ubiquitous process of plants (Rhew *et al*., 2003; Nagatoshi and Nakamura,

2007). Chloromethane is produced by the transfer of the methyl group of the SAM to a chloride ion (Cl^-) (Figure 1.3). The SAM is involved in numerous synthesis reactions of lignin, a component of the plant cell wall (Campbell and Sederoff, 1996). Rhew *et al*., 2003 has shown that the methyltransferase coded by the HOL gene catalyzes the methylation of chloride, and is dependent on SAM.

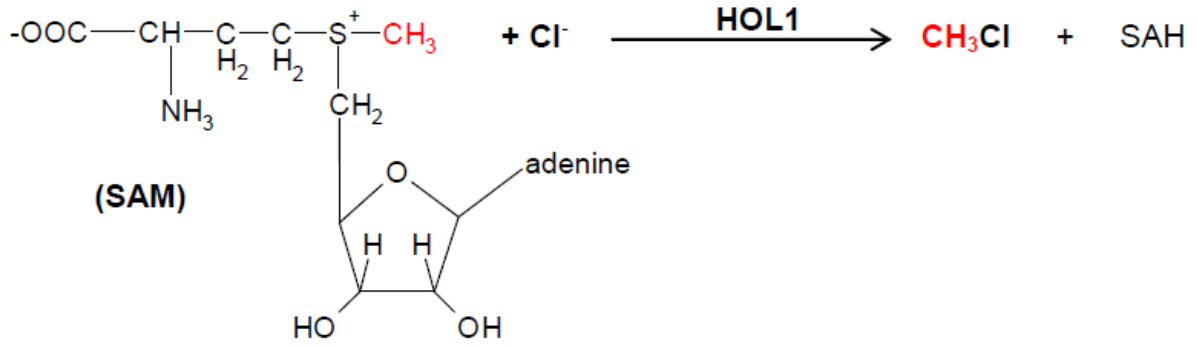


Figure 1.3. Chloromethane production in *Arabidopsis thaliana*
(Figure obtained and modified from the thesis manuscript of Muhammad Farhan Ul Haque, May 2013).

In fact, different studies have demonstrated the existence of plants that emit chloromethane. For example, of 187 plants studied in a subtropical region, 33 species belonging to different families emit chloromethane (Yokouchi *et al*., 2007). Similarly, the study of chloromethane fluxes in a tropical forest showed that 25 out of 117 plants studied emit chloromethane (Saito *et al*., 2008). The most important emissions were observed in fern species (*Osmunda banksiifolia*, *Cibotium balometz*, *Angiopteris palmiformis*), and in halophile plants (*Vitex rotundifolia*, *Vitex trifolia* and *Excoecaria agalloch*) with emission rates in the order of 1 µg of chloromethane.g^{-1} of dry matter. h^{-1}. Nevertheless, the capacity of chloromethane production depends more on the plant species than on the family (Yokouchi *et al*., 2007).

– Production of chloromethane by the soil and decomposing vegetable matter

Decomposing plants are the biggest source of chloromethane (Hamilton, 2003; Montzka *et al*., 2011). Production of chloromethane is mostly found in the O horizon of soils and seems to be insignificant in lower horizons (Figure 1.4) (Redeker and Kalin, 2012).

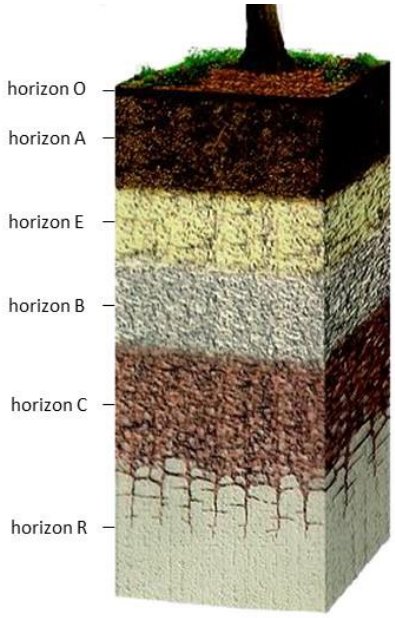


Figure 1.4. Cross-section of forest soil top revealing horizons
The forest soil is divided into different horizons: O, A, E, B, C, and R. Horizon O is mostly composed of decomposing organic matter; A is equally composed of organic and mineral matter; E is a leached horizon, where soluble nutrients are no longer present due to precipitation or irrigation; B is a horizon rich in minerals from the leaching of horizons A and E; C is a non-consolidated underground horizon; finally, E is a rocky horizon, composed of rocks such as granite or limestone. Horizons O and A are biologically active and highly oxygenated, while the others are poor in organic matter and have lower microbial activity. Within the framework of my thesis project, an experiment on the forest soil was performed at the level of the O horizon.
(Figure modified from http://www.ctahr.hawaii.edu/mauisoil/a_profile.aspx)

Soil is an environment rich in chloride ions, and the bioavailability of these is essential to the formation of chloromethane. However, the chloride ion concentration in the soil is difficult to estimate due to the intervention of several processes and mechanisms such as the flow between the vegetation and soil, alterations of the rock, or the decomposition of organic matter (Bastviken *et al*., 2007; Öberg, 2003). It has been observed that a part of the chloride ions is considered non-bioavailable because of its absorption by vegetation or microorganisms (Bastviken *et al*., 2007). Chloride ions also occur in tree trunks and fungi that colonize the wood utilize them for the chloromethane synthesis (Watling and Harper, 1998). The amount of Cl^- in the wood of temperate forest species is estimated to vary between 2.4 and 123 $mg.kg^{-1}$ of dry matter (Watling and Harper, 1998).

Studies of fungi have demonstrated that members of the *Hymenochaetaceae* are able of re-emitting a proportion of the chloride ions found in their substrates in the form of

chloromethane (Harper and Kennedy, 1986). In fact, decomposing wood is colonized by fungi that, by means of a group of peroxidases, degrade lignin, a major component of wood, into SAM. The SAM and the Cl^- ions present in the medium forms chloromethane, a reaction mediated by a methyltransferase (Harper, 2000; Anke and Weber, 2006). Such chloromethane production mechanism has been studied and described in the fungus *Phellinus tuberculosus* (White, 1982; Wuosmaa and Hager, 1990); however, the most common example is that of *Phellinus pomaceus* (Figure 1.5), which is capable of re-emitting as chloromethane 90% of the chloride of its degraded substrates (Harper *et al*., 1988). This chloromethane production occurs when a Cl^- ion is transferred to the methyl group of the SAM (Figure 1.5).

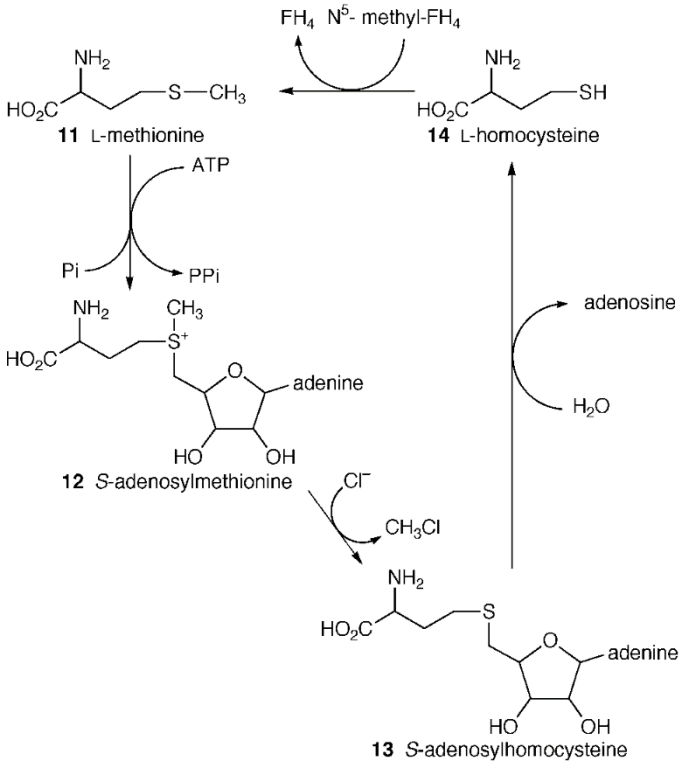


Figure 1.5. Chloromethane production by the *Phellinus pomaceus* fungus
(Figure modified from Harper, 2000)

Chloromethane production by fungi associated to the root systems of forest soils exists, as well. Roots represent an important source of Volatile Organic Compounds (VOCs) (Lin *et al*., 2007). In the study carried out by Redeker *et al*., 2004, 9 ectomycorrhizal fungi isolates, among them *Cenococcum geophilum,* one of the most widespread and common species, (Trappe, 1977; Massicotte *et al*., 1992), demonstrated their ability to produce halogenated methanes, chloromethane being one of them. They produce between 0.003 µg and 65 µg of

chloromethane per gram of dry matter (Redeker *et al*., 2004). However, it is difficult to separately evaluate the role of the root system, since it is closely tied to the microbial system, which itself influences the flow of VOCs in the soil (Peñuelas *et al*., 2014). In fact, 95% of plants' short roots are colonized by fungi which are located close to communities of mycorrhiza helper bacteria that aid fungi in the rhizosphere, notably during processes of mycorrhizal development (Frey-Klett *et al*., 2007; Bonfante and Anca, 2009; Rigamonte *et al.*, 2010).

– Chloromethane is emitted at lower rates than other VOCs

Plants re-emit a proportion of fixed as VOCs, methanol being among them (Vorholt, 2012; Bringel and Couée, 2015). These compounds are emitted into the atmosphere or into the soil, where they can be directly assimilated by microorganisms (Schade and Goldstein, 2001). Thus, the largest part of atmospheric methanol is linked to the production by plants (Jacob, 2005; Galbally and Kirstine, 2002). Other VOCs are produced by plants, among which there are aromatic compounds and halogenated compounds (Guenther *et al*., 2012; Forczek *et al*., 2015). The plant production of 4 chlorinated methanes has been described, even though chloromethane is the most produced, with a production of 0.8 to 8 $Tg.year^{-1}$ (Clerbaux *et al*., 2007). This production is lower than that of methanol (50 - 132 $Tg.\ year^{-1}$) (Jacob, 2005; Forczek *et al*., 2015).

In the case of forest soils, chloromethane is far from being the only VOC produced (Keppler *et al*., 2005; Redeker and Kalin, 2012; Peñuelas *et al*., 2014). Soil is a heterogeneous complex ecosystem, an assembly of complex microenvironments (Peñuelas *et al*., 2014) with an huge microbial diversity. Bacterial and fungal diversity can be evaluated by the number of different species, respectively, 10^5 and 10^6 species per gram of soil (Peñuelas *et al*., 2014). Numerous VOCs are emitted by these microorganisms (Figure 1.6). This production is directly related to the availability of substrates and of growth conditions (Kai *et al*., 2010; Blom *et al*., 2011). A large part of emitted VOCs comes from microbial degradation of plant litter (dead plant matter), of the decomposition of organic matter (Kögel-Knabner, 2002; Peñuelas *et al*., 2014), as well as from root exudates.

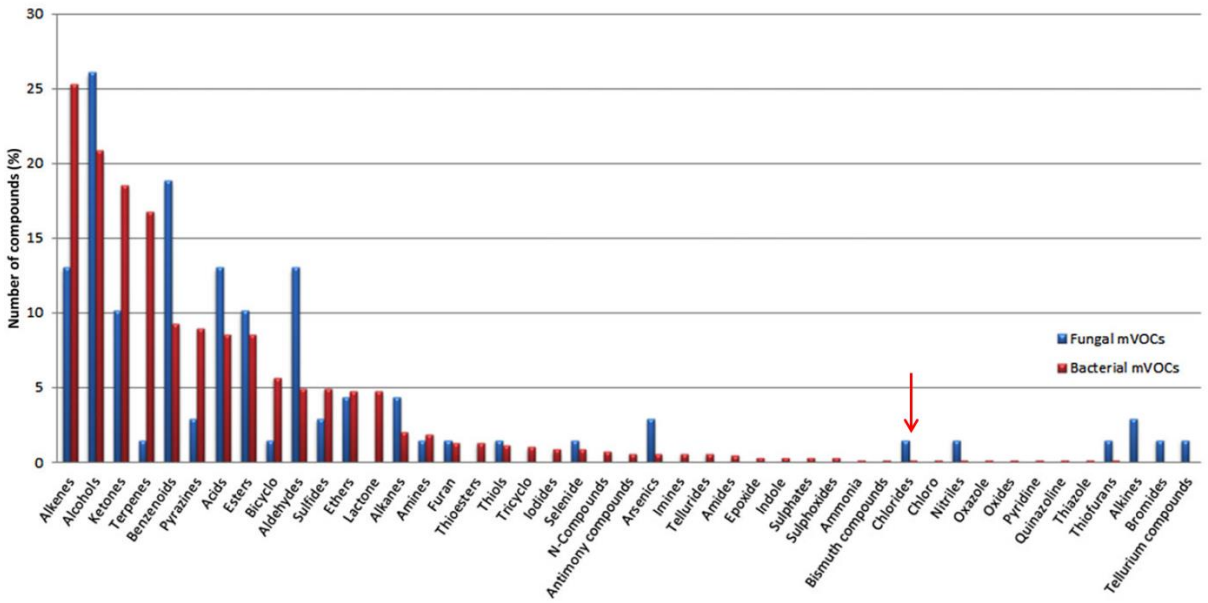


Figure 1.6. Volatile organic compounds emitted by soils
Bacterial emissions are marked in red, and fungal emissions are marked in blue (modified from Peñuelas *et al.,* 2014). The complete inventory of these VOCs is listed in a database (http://bioinformatics.charite.de/mvoc/). The red arrow marks the category of VOCs in which chloromethane is found.

In general, bacteria produce more alkenes, alcohols, ketones, and terpenes, while fungi produce more alcohols, benzenoids, acids, and aldehydes (Figure 1.6). Chlorinated compounds are produced less, and this production in the soil is more important in fungi than in bacteria (Figure 1.6). The “chloride” group includes chloromethane (indicated with a red arrow).

– Chloromethane dosage using a bioreporter strain

The quantification of the low amounts of chloromethane produced by plants and soils (in the order of ppt) is difficult using classical approaches of gas chromatography (GC) dosage that is not sensitive enough (Figure 1.7). In the laboratory in Strasbourg, with the goal of evaluating chloromethane emissions by plants, *Methylobacterium extorquens* CM4 was transformed in order to be able to emit fluorescence in proportion to the weak concentration of chloromethane (Farhan Ul Haque *et al*., 2013). This bioreporter strain is based on the presence of a plasmid that carries the *yfp* gene, which then encodes a fluorescent protein under the control of the promoter of gene *cmuA* (Figure 1.7). The bioreporter strain allows a sensitive and proportional detection of chloromethane concentrations, with a detection threshold of 2

pM. This bioreporter has been tested for the quantification of chloromethane emitted by the plants *A. thaliana* and *V. rotundifolia*, which were demonstrated to have emission rates of respectively 13 and 2800 ng.g of dry matter^{-1}. h^{-1} (Farhan Ul Haque *et al.*, 2013).

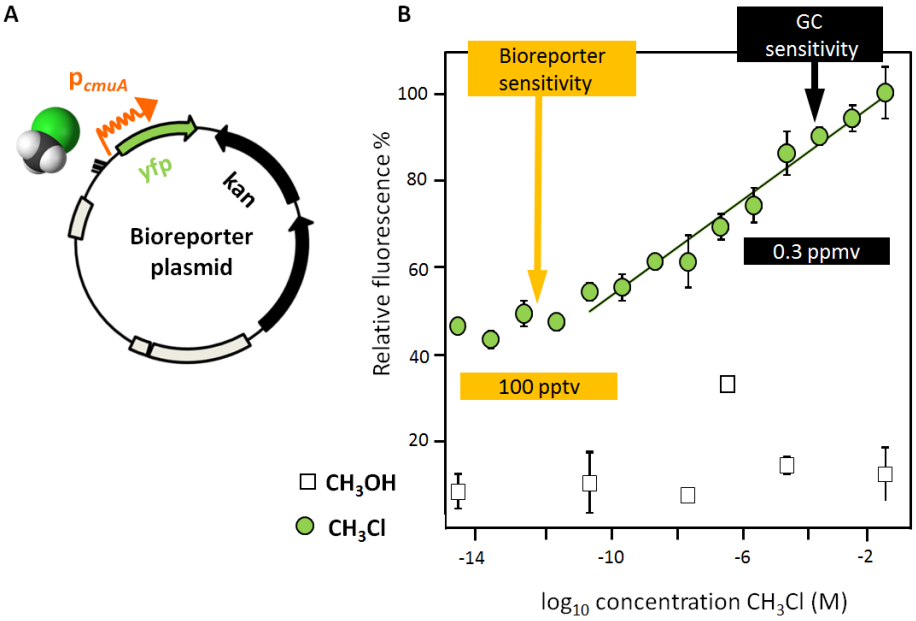


Figure 1.7. Chloromethane dosage using a bioreporter strain
(A) In the bioreporter strain, a plasmid that carries the *yfp* gene under the control of the promoter of the *cmuA* gene (encoding a protein with a methyltransferase domain and a corrinoid-binding domain, essential for chloromethane utilization in *M. extorquens* CM4 (Studer *et al.*, 2001, 2002)), which is induced in the presence of chloromethane, is present in *M. extorquens* CM4 (Farhan Ul Haque *et al.*, 2013). The plasmid carries a gene resistant to kanamycin in order to maintain its presence during the propagation of the bioreporter strain. (B) The emission of fluorescence of the Yfp protein is specific to the presence of chloromethane. The threshold of detection with the bioreporter strain is marked in yellow, and in black with the GC dosage.

2.2. Chloromethane sinks

2.2.1. Abiotic degradation of chloromethane

Different abiotic sinks of chloromethane, which vary in their ability to eliminate chloromethane, have been identified (Tableau 1.4). Even though an abiotic degradation of VOCs has been observed in soils, abiotic degradation of chloromethane has not been described (Miller *et al.*, 2004; Insam and Seewald, 2010).

Table 1.4. Estimation of atmospheric chloromethane sinks

Sink type	Estimation (Gg. year^{-1})[a]	Low/high values (Gg. year^{-1})[b]
Tropospheric reaction with OH	-3,180	-3800/ -4,100
Stratospheric loss	-200	-100/ -300
Reaction with Cl^- in marine boundary layer	-370	-18/ -550
Microbial degradation in soil	< -1,000	-100/ -1,600
Cold ocean water loss	-75	-93/-145
Total sinks	< -4,875	-4273/ -6,695

[a] Data from Keppler *et al.*, 2005
[b] Data from Clerbaux *et al*., 2007

2.2.2. Biotic degradation of chloromethane

Biotic degradation of chloromethane is mainly due to microorganisms activity, although its estimation is variable (from 180 to 1,600 Gg.an^{-1}) (Keppler *et al*., 2005).

– Degradation by bacteria in marine environments

Marine environments play a role in the chloromethane degradation. Bacterial strains able to utilize chloromethane have been isolated from marine environments, such as *Leisingera methylohalidovorans* MB2, the first class of *Rhodobacteracea* shown to be able to degrade chloromethane (Goodwin *et al*., 1997; Schaefer; 2002, Tableau 1.5). Three other chloromethane-degrading strains of this family have been isolated from marine environments: *Roseovarius* sp. 179, *Roseovarius* sp. 217 and *Ruegeria* sp. 198 (Schäfer *et al*., 2005; Tableau 1.5).

Table 1.5. Chloromethane-utilizing bacterial isolated from contrasting environments

Bacterial strain	Origin	Gram type	Metabolism	Trophic type	Presence of *cmuA*[a]	Reference
Acetobacterium dehalogenans MC	Activated sludge	Positive	Anaerobic	Homoacetogenic	nd	Traunecker *et al*., 1991
Aminobacter ciceronei IMB1	Fumigated strawberry field	Negative	Aerobic	Facultative methylotroph	Yes	Hancock *et al*., 1998
Aminobacter lissarensis CC495	Beech woodland soil	Negative	Aerobic	Facultative methylotroph	Yes	Coulter *et al*., 1999
Hyphomicrobium sp. AT1	Phyllosphere	Negative	Aerobic	Facultative methylotroph	Yes	Nadalig *et al*., 2011
Hyphomicrobium sp. AT2	Phyllosphere	Negative	Aerobic	Facultative methylotroph	Yes	Nadalig *et al*., 2011
Hyphomicrobium sp. AT3	Phyllosphere	Negative	Aerobic	Facultative methylotroph	Yes	Nadalig *et al*., 2011
Hyphomicrobium sp. AT4	Phyllosphere	Negative	Aerobic	Facultative methylotroph	Yes	Nadalig *et al*., 2011
Hyphomicrobium sp. MC1	Industrial sewage plant	Negative	Aerobic	Facultative methylotroph	Yes	Hartmans *et al*., 1986
Hyphomicrobium sp. MC2	Soil from a petrochemical factory	Negative	Aerobic	Facultative methylotroph	Yes	Doronina *et al*., 1996
Leisingera methylohalidovorans MB2	Marine tide pool	Negative	Aerobic	Facultative methylotroph	No	Schaefer, 2002
Methylomicrobium album BG8	Fresh water	Negative	Aerobic	Obligatory methylotroph	nd	Han and Semrau, 2000
Methylobacterium extorquens CM4	Soil from a petrochemical factory	Negative	Aerobic	Facultative methylotroph	Yes	Doronina *et al*., 1996
Pseudomonas aeruginosa NB1	Activated sludge	Negative	Anaerobic	Facultative methylotroph	nd	Freedman *et al*., 2004
Roseovarius sp. strain 179	Coastal seawater	Negative	Aerobic	Facultative methylotroph	Yes	Schäfer *et al*., 2005
Roseovarius sp. strain 198	Coastal seawater	Negative	Aerobic	Facultative methylotroph	Yes	Schäfer *et al*., 2005
Roseovarius sp. strain 217	Seawater	Negative	Aerobic	Facultative methylotroph	Yes	Schäfer *et al*., 2005

[a] nd; not determined

– Bacterial degradation in the phyllosphere

The aerial parts of plants constitute the phyllosphere, which is home for many microorganisms. The total leaf surface (bottom and upper side) corresponds to twice the surface of Earth, with a bacterial density of 10^7.cm^{-2} (Vorholt, 2012). Few studies have studied so far the role of phyllospheric microorganisms in the degradation of plant-emitted chloromethane. Some bacteria able to utilize chloromethane have been isolated from the surface of leaves of *A. thaliana* (Nadalig *et al*., 2011). Nevertheless, the potential role of these epiphytic bacteria as biotic filters for chloromethane has so far not been taken into account.

– Bacterial degradation in soils

Soil contains a large variety of compounds, such as halogenated compounds. Soil has been shown to be major sink for some of halogenated compounds, and for instance 70% of bromomethane (CH_3Br) is consumed by soils (Shorter *et al*., 1995). Although chloromethane degradation in soils is misevaluated, it is estimated to exceed 1,000 Gg. $year^{-1}$ (Tableau 1.4). Chloromethane consumption in soil has been correlated to bacterial and fungal degradation mainly in the O horizon with was found to be very weak in lower horizons (Redeker and Kalin, 2012).

The exchange of chloromethane between the terrestrial ecosystems and the atmosphere are modulated by bacteria living in soil (Miller *et al*., 2004; Borodina *et al*., 2005; Keppler *et al*., 2005; Clerbaux *et al*., 2007; Schäfer *et al*., 2007; Rhew *et al*., 2010). A variety of bacteria affiliated to *Alpha*- or *Beta-proteobacteria* that are able to utilize chloromethane have isolated from forest soils, but their contribution to the chloromethane budget remains to be evaluated (Miller *et al*., 2004; Borodina *et al*., 2005).

Chloromethane consumption has also been demonstrated in fungi via reactions with produced secondary metabolites including more than 200 halogenated metabolites (Gribble, 2003) such as brominated, fluorinated, iodinated, and chlorinated compounds (Anke et Weber, 2006). Among chlorinated metabolites, chloromethane can be released during methylation reactions in fungi (Figure 1.8). Methylation is a common reaction of fungal metabolism (Anke and Weber, 2006).

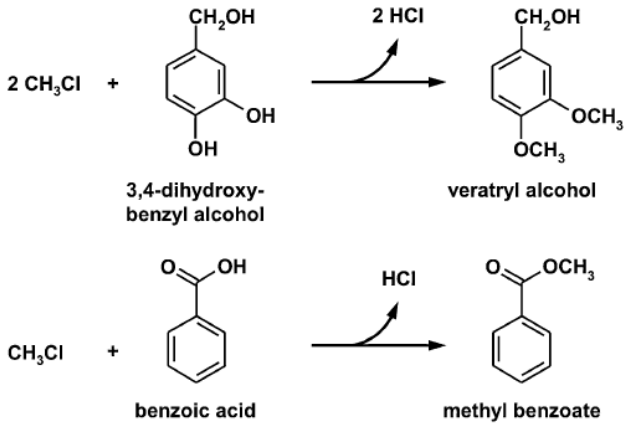


Figure 1.8. Chloromethane degradation reactions by fungi
(Modified from Anke and Weber, 2006)

Methylation in fungi involves SAM, as a donor of the methyl group, as found for chloromethane synthesis in plants. In fungi, the utilization of chloromethane is preferred over SAM, since the reaction needs less energy (Harper, 2000). When chloromethane serves as a methyl group donor to an hydroxyl (-OH) or carboxyl (-COOH) group, products such as veratryl alcohol or methyl benzoate are produced (Figure 1.8). Veratryl alcohol is a secondary metabolite involved in the synthesis of enzymes that are necessary for lignin degradation (Dekker *et al*., 2001). Methyl benzoate is an olfactive compound synthesized to attract pollinators, and acts in inhibition of *A. thaliana* root growth (Horiuchi *et al*., 2007). Chloromethane balance in fungi is difficult to determine since they are able to simultaneously produced and degraded chloromethane (Anke and Weber, 2006).

Although the role of soil as a chloromethane sink does not need to be further evidenced, uncertainties persist about global chloromethane fluxes in soils (see section below).

2.3. Global chloromethane balance in soil

Soil is a complex environment for which chloromethane sources and sinks are difficult to estimate. Besides its intrinsic heterogeneity (Figure 1.4), contrasting variations of physico-chemical, climatic and biological parameters occur. For instance, a seasonal effect on chloromethane production has been described (Peñuelas *et al*., 2014). Increased chloromethane emissions have been observed in summer (Llusia *et al*., 2013; Oderbolz *et al*., 2013), compared to in the falls where the leaf cover reduce emission of COVs such as chloromethane (Figure 1.9). Temperature also has an impact on chloromethane production.

The release of chloromethane by decomposing plants increases with higher temperatures (Hamilton, 2003).

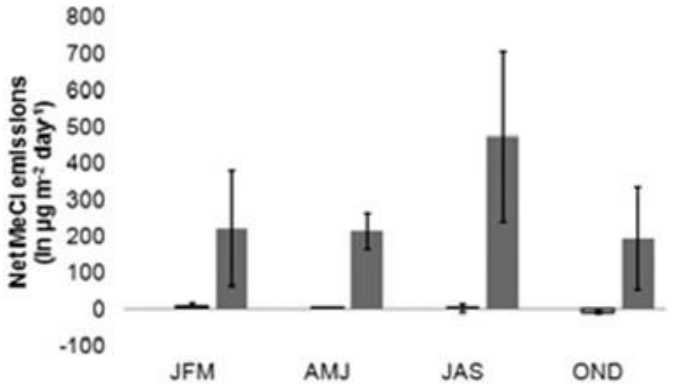


Figure 1.9. Seasonal chloromethane fluxes in forest soil
Chloromethane emissions from a forest soil have been monitored during one year. JFM, January, February and March; AMJ, April, May and June, JAS, July, August and September; OND, October, November and December (from Redeker *et al.,* 2012). Error bars show standard errors.

Chloromethane production can oscillate between 30 and 60 % (Saito and Yokouchi, 2008). Estimation of chloromethane production is difficult since chloromethane fluxes are bidirectional (from the soil to the atmosphere and from the atmosphere to the soil), and forest soil microorganisms modulate these exchanges (Schäfer *et al*., 2005). The estimated chloromethane sinks in soils varies from 256 Gg.year^{-1} (Clerbaux *et al*., 2007) to 1,000 Gg.year^{-1} (Keppler *et al*., 2005), which represents 6 up to 20% of the total tropospheric chloromethane sinks. Further studies are needed to better understand the global cycle of chloromethane in forest soils, which are both source and sink of chloromethane.

Quantification uncertainties of different chloromethane sources and sinks prevent accurate global balance calculations. The fact that some environments are sources and sinks, even makes it more complex to estimate. The estimation of chloromethane production and consumption can be improved using analytical methods basing on carbon and hydrogen isotope fractionation (Keppler *et al.,* 2005; Greule *et al*., 2012; Nadalig *et al*., 2013) and the development of three-dimensional model of terrestrial chloromethane fluxes (Xiao *et al.,* 2010). Studies of chloromethane degradation in forest soils have shown the presence of bacteria active on chloromethane degradation (Borodina *et al*., 2005; Miller *et al*., 2004). An unpublished study of a forest soil in Steigerwald in Germany (samples from this site were used

for stable isotope probing experiments described in chapter 4) demonstrated chloromethane degradation by bacteria in oxic conditions (Rüffer, 2013).

3. Methylotrophy

Microorganisms able to grow in chloromethane as sole carbon and energy source are methylotrophic microorganisms (Table 1.5). Methylotrophy is defined as the ability of an organism to utilize a compound with no C-C bonds as the sole source of carbon and energy. Substrates for methylotrohic growth include one-carbon compounds such as methane, (CH_4), methanol (CH_3OH), formaldehyde (HCHO), and methylamine (CH_3NH_2), as well as halogenated compounds such as chloromethane and dichloromethane, or multi-carbon compounds such as dimethylamine (C_2H_7N) (Chistoserdova, 2011; Muller *et al*., 2011a).

Methylotrophic organisms are found ubiquitous in the air (DeLeon-Rodriguez *et al*., 2013), in fresh water and marine environments (Neufeld *et al*., 2008), in plant roots leaves (Nadalig *et al*., 2011; Knief *et al*., 2012), and in soils and marine sediments (Kolb, 2009a). These organisms belong to the three domains of living organisms: bacteria, archea (methanotrophic archea) (Antony *et al*., 2012), and eukaryotes (yeasts) (Yurimoto, 2009). Yeasts of the genus *Pichia*, *Candida* and *Hansenula* are able to metabolize methanol via an alcohol oxidase present in the peroxisome cellular compartment (van der Klei *et al*., 2006).

The methylotrophic metabolism has been the best studied in bacteria (Kolb, 2009; Chistoserdova, 2011; Peyraud *et al*., 2011; Peyraud *et al*., 2012). Since the discovery of bacteria able to utilize methane and methanol as the only source of carbon and energy (Loew, 1892; Käserer, 1906; Söhngen, 1906), a large variety of bacteria from different genus have been identified, including facultative methylotrophs (able to grow with C_1 and multi-carbon substrates such as succinate) (Whittenbury *et al*., 1970; Lidstrom, 2006; Boden *et al*., 2008; Hung *et al*., 2011). Methylotrophic bacteria can be found among *Alpha*, *Beta* and *Gamma proteobacteria*, Gram negative and Gram positive bacteria, using aerobic or anaerobic metabolisms.

Methylotrophic bacteria are classified in two groups. The first group of methanotrophs is able to utilize methane as the sole source of carbon and energy, are often strict methylotrophs (*Methylomonas*, *Methylobacter*, *Methylosinus*, and *Methylocystis*). The second group of non-methanotrophic methylotrophs (unable to oxidize methane), are often facultative methylotrophs, like bacteria of the genus *Hyphomicrobium* or *Methylobacterium* (Chistoserdova and Lidstrom, 2013).

3.1. Methylotrophic pathways

Methylotrophy involved 3 major metabolic steps:

- Initial transformation of the methylotrophic compound into formaldehyde (HCHO) (not the case for chloromethane);
- Complete oxidation of the methylotrophic compound to CO_2 in order to produce energy;
- Carbon assimilation for biomass production via 3 alternative pathways: the ribulose biphosphate (RuBP) pathway that utilizes CO_2, the ribulose monophosphate (RuMP) pathway from formaldehyde, or the serine cycle from methylenetetrahydrofolate (CH_2=H_4F) and CO_2 (Chistoserdova, 2011).

For most methylotrophs, the first metabolic step is the oxidation or the hydrolysis of the carbon compound into formaldehyde. The enzymes involved in this step are specific to their substrates: the soluble methane monooxygenase (sMMO) or the particular methane monooxygenase (pMMO) for methane, the methylamine deshydrogenase (MADH, periplasmic) for methylamine, the methanol deshydrogenase (Mxa, periplasmic) for methanol, or a dichloromethane deshydrogenase for dichloromethane (DcmA, cytoplasmic). The *mxaF* gene has been used as a molecular marker of methanol oxidation (Neufeld *et al.,* 2008). More widespread in the environment than the *mxa* genes, alternative methanol degradation systems exist such as the *xox* genes or the *mdh* gene (Kalyuzhnaya *et al*., 2008). Among these genes, *xoxF* codes for a subunit of a methanol dehydrogenase that shares more than 50% of amino acid identity with MxaF (Schmidt *et al.,* 2010; Taubert *et al*., 2015). Several methanol oxidation pathways can be simultaneously present in an organism as found in *M. extorquens* (Figure 1.10, Skovran *et al*., 2011). Several methylamine oxidation pathways have also been found in *M. extorquens* strains (Vuilleumier *et al.,* 2009; Gruffaz *et al*., 2014; Nayak et Marx, 2014). Halogenated compounds (chloromethane, dichloromethane) can be utilized by methylotrophic bacteria such as *M. extorquens* and *Hyphomicrobium* spp. (Gälli and Leisinger, 1985; La Roche and Leisinger, 1990; Vannelli *et al*., 1998; McAnulla *et al*., 2001). The dehalogenation of dichloromethane is catalyzed by a dichloromethane dehalogenase encoded by gene *dcmA*, and formaldehyde is produced.

Formaldehyde is transformed into formate by either a tetrahydrofolate-dependent (H_4F) pathway or a tetrahydromethanopterin (H_4MPT)-dependent pathway in *Methylobacterium*

strains. The first pathway is not the main one regarding the production of energy and involves inter-conversions between methylene-H_4F, methenyl-H_4F, formyl-H_4F and formate (Marison and Attwood, 1982). This metabolic pathway might play a major role by keeping adequate concentrations of the different H_4F-containing metabolites, which are essential intermediates for carbon assimilation via the serine cycle pathway (Marx *et al*., 2005; Crowther *et al*., 2008). In the second pathway, discovered in 1998, H_4MPT is condensed to formaldehyde, and successively transformed into methylene-H_4F, methenyl-H_4F, formyl-H_4F, and finally into formate (Chistoserdova *et al*., 1998). This H_4MPT-dependent pathway has been first described in the *Archaea*, and is more efficient in formaldehyde oxidation into CO_2 than the H_4F-dependent pathway, because its corresponding enzymes display high activities (Vorholt, 2002). Despite the fact that most methylotrophic substrates are oxidized into formaldehyde, this is the case for chloromethane. Chloromethane dehalogenation produces methylene-tetrahydrofolate (CH_2=H_4F), which either enters the serine cycle pathway for biomass production or is oxidized into formate for energy production (Figure 1.10). As a matter of facts, *M. extorquens* CM4 converts directly chloromethane into methylene-tetrahydrofolate (CH_2=H_4F) without formaldehyde production thanks to the activity of CmuA (methyltransferase and corrinoid-binding domain-containing protein), CmuB (methylcobalamin: H_4F methyltransferase) and MetF2 (methylene-H_4F reductase) (Studer *et al*., 1999; Studer *et al*., 2001; Studer *et al*., 2002) (see section 1.3.4.2.).

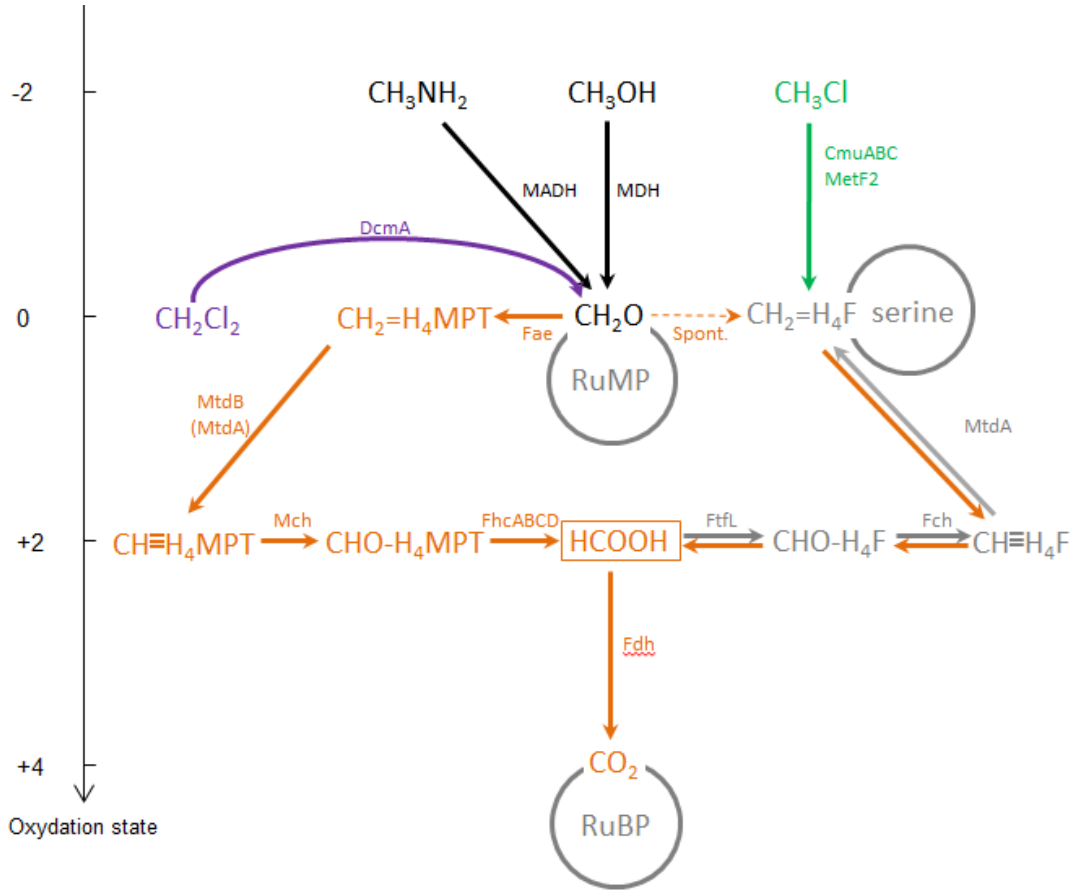


Figure 1.10. Scheme of different methylotrophic pathways in *M. extorquens*
Oxidation state is shown on the Y axis. The reactions specific for chloromethane (CH_3Cl) and dichloromethane (CH_2Cl_2) utilization as sole source of carbon and energy are shown in green and purple, respectively: CmuABC, chloromethane utilizing genes; MetF2, methylene-H_4F reductase; DcmA, dichloromethane dehalogenase. Enzymes for initial oxidation are in black: MDH, methanol dehydrogenase; MADH, methylamine dehydrogenase. Reactions leading to complete oxidation of carbon to energy are in orange: Fae, formaldehyde activation enzyme; MtdAB, methylene-tetrahydromethanopterin (H_4MPT) dehydrogenase; Mch, methenyl-H_4MPT cyclohydrolase; Fch, methylene-tetrahydrofolate (H_4F) cyclohydrolase; FtfL, formyl- H_4F ligase; FhcABCD, formyltransferase-hydrolase complex; Fdh, formate dehydrogenase. Reactions for carbon assimilation to biomass are in grey: FtfL, formyl-H_4F ligase; Fch, methylene- H_4F cyclohydrolase; MtdA, methylene-H_4MPT dehydrogenase; RuMP, ribulose monophosphate cycle; RuBP, ribulose biphosphate cycle.

More than 86 genes involved in *M. extorquens* C_1 compound utilization have have been identified (Marx *et al.*, 2003). These genes are involved in the methanol oxidation methylamine, formaldehyde, and formate or in the serine-cycle pathway (Figure 1.11).

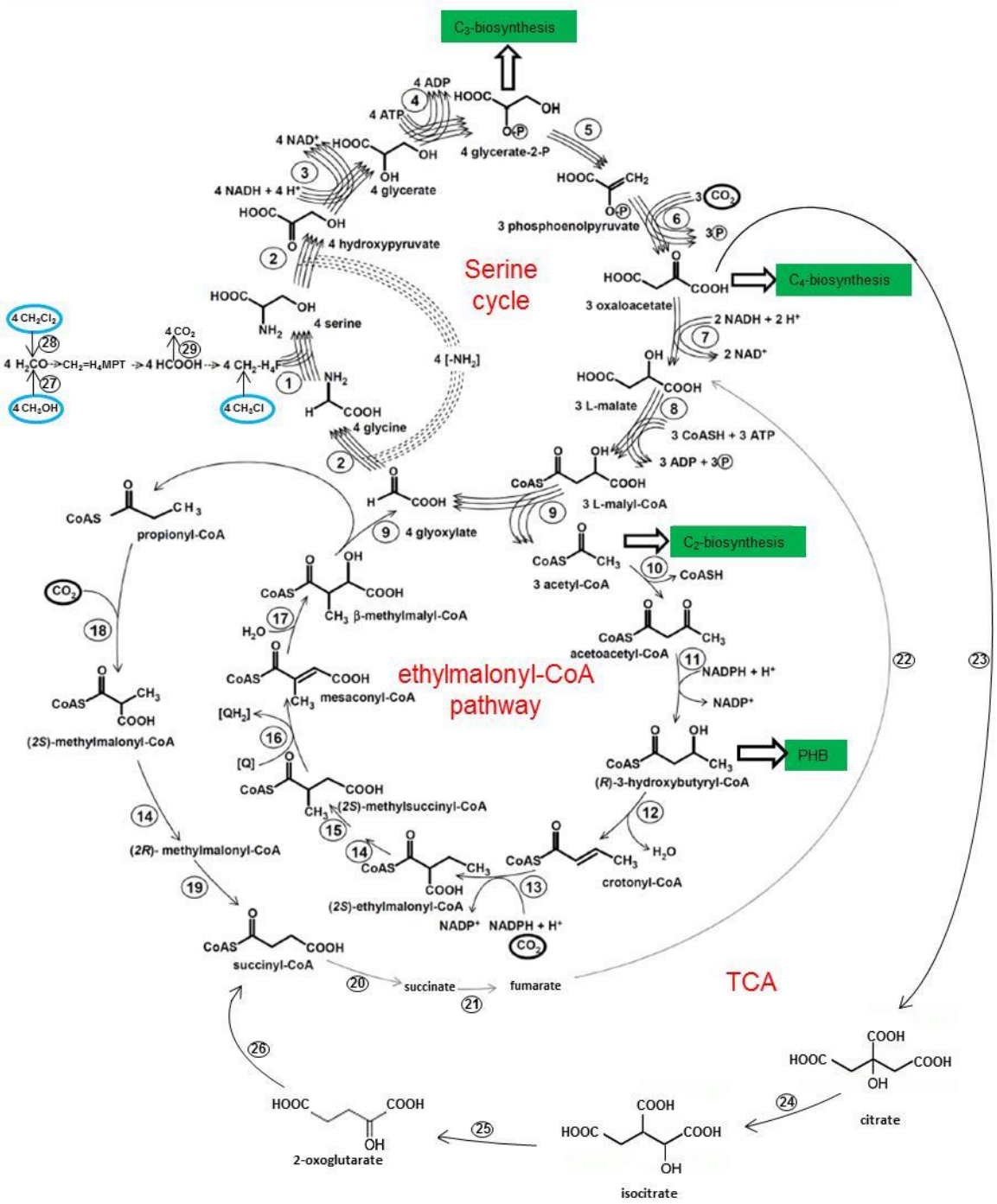


Figure 1.11. Carbon assimilation in *M. extorquens*

Entering C_1 compounds are circled in blue and main biosynthesis pathway outputs are indicated in green. Bacteria oxidize methanol (CH_3OH) or dichloromethane (CH_2Cl_2) to formaldehyde that is condensed with a tetrahydromethanopterin (H_4MPT), and further oxidized to formate (HCOOH). Formate is further converted to methylene-tetrahydrofolate (H_4F) via the serine cycle or transformed in CO_2. Chloromethane (CH_3Cl) is assimilated via the transfer of the methyl group to H_4F which is then assimilated via the serine cycle to produce biomass or oxidized to formate (HCOOH). Acetyl-CoA assimilation and glyoxylate regeneration proceed via the ethylmalonyl-CoA pathway. The tricarboxylic acid (TCA) cycle is mainly used for assimilation of multi-carbon compounds (Šmejkalová *et al.*, 2010). Circled numbers identify distinct reactions: 1, serine hydroxymethyl transferase; 2, serine-glyoxylate aminotransferase; 3, hydroxypyruvate reductase; 4, glycerate kinase; 5, enolase; 6,

phosphoenolpyruvate carboxylase; 7, malate dehydrogenase; 8, malate-CoA ligase (malate thiokinase); 9, L-malyl-CoA/β-methylmalyl-CoA lyase; 10, β-cetothiolase; 11, acetoacetyl-CoA reductase; 12, crotonase (R- specific); 13, crotonyl-CoA carboxylase reductase; 14, ethylmalonyl-CoA/methylmalonyl-CoA epimerase; 15, ethylmalonyl-CoA mutase; 16, methylsuccinyl-CoA deshydrogenase; 17, mesaconyl-CoA hydratase; 18, propionyl-CoA carboxylase; 19, methylmalonyl-CoA mutase; 20, succinyl-CoA synthase; 21, succinate deshydrogenase; 22, fumarase; 23, citryl-CoA lyase; 24, aconitate hydratase; 25, NADP-dependent isocitrate dehydrogenase; 26, 2-oxoglutarate dehydrogenase complex; 27, methanol dehydrogenase; 28, dichloromethane dehalogenase; 29, formate dehydrogenase. For a given reaction, the number of parallel arrows correspond to the number of molecules involved in the reaction. PHB, polyhydroxybutyrate; Q, quinone (adapted from Šmejkalová *et al*., 2010).

In *Methylobacterium,* for biomass production, formaldehyde incorporates the serine cycle pathway via formate and methylene-H_4F. The GlyA enzyme (serine hydroxymethyltransferase) catalyzes the reaction between methylene-H_4F and glycine to produce serine (Figure 1.11). Then, serine (C_3 compound) is successively transformed into hydroxypyruvate, glycerate, glycerate-2-phosphate and phosphoenolpyruvate. Phosphoenolpyruvate is subsequently converted into oxaloacetate, then malate by incorporation of CO_2. Malate (C_4 compound) coupled to co-enzymeA (CoA) is then cleaved in 2 compounds: glyoxylate allowing to close the cycle serine, and acetyl-CoA which is integrated in the ethylmalonyl-CoA pathway where a CO_2 molecule is incorporated (Erb *et al*., 2008; Peyraud *et al*., 2009; Šmejkalová *et al*., 2010). This cycle generates hydroxybutyryl-CoA, a precursor of the polyhydroxybutyrate (PHB) synthesis and granule production for carbon and phosphate storage in the cell.

3.2. Bacterial degradation of dichloromethane (CH_2Cl_2)

Bacterial dichloromethane is better characterized in oxic than in anoxic conditions.

3.2.1. Dichloromethane utilization in anoxic conditions

Dichloromethane can be fermented by *Dehalobacterium formicoaceticum* (Yurimoto, 2009). This anaerobic bacterium has a cobalamin- and H_4F-dependent methyltransferase. Dichloromethane and H_4F are converted into methylene-tetrahydrofolate (CH_2=H_4F) and into inorganic chloride. Part of CH_2=H_4F is converted into formate, whereas another part of the CH_2=H_4F incorporates CO_2 to produce acetate (Mägli *et al*., 1998).

3.2.2. Dichloromethane utilization in oxic conditions

The first characterized isolate able to utilize dichloromethane for growth, belongs to the genus *Methylopila* (Brunner *et al*., 1980). Three aerobic dichloromethane degradation pathways are known. Dichloromethane utilization as carbon and energy sources is possible only in one of these pathways and involved the expression of a dichloromethane dehalogenase belonging to the glutathion-S-transferase (GST) family that is encoded by gene *dcmA* (Figure 1.12). The majority of strains able to utilize dichloromethane for their growth, harbor gene *dcmA* (Vuilleumier *et al*., 2001; Muller *et al*., 2011b). When detected, gene *dcmA* is highly conserved with up to 98-100% identity at the nucleotide level. The DcmA protein catalyzes the dechlorination of one mole of dichloromethane into two moles of protons (H^+), two moles of chloride (Cl^-) and one mole of formate. This reaction involved two genotoxic compounds, S-chloromethylglutathion and formaldehyde (Figure 1.12).

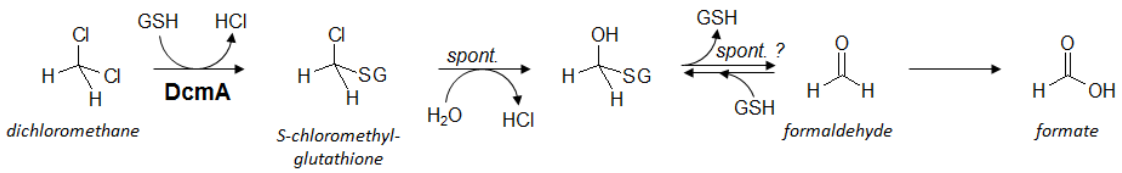


Figure 1.12. Dichloromethane degradation by the hydrolytic pathway of glutathione S-transferase

(adapted from Muller *et al.,* 2011). spont.: spontaneous, abiotic reaction; GSH, glutathione; SG, glutathionyl.

The aerobic bacterial degradation of dichloromethane has been best described in *M. extorquens* DM4, a strain able to utilize dichloromethane as unique source of carbon and

energy (Roche and Leisinger, 1991; Kayser *et al*., 2002; Muller *et al*., 2011a). Utilization genes are located on a catabolic transposon containing four *dcm* genes, framed by insertion sequences. Among *dcm* genes, *dcmA* encoded a dichloromethane dehalogenase, *dcmR* encodes a transcriptional factor, and genes *dcmB* and *dcmC* encode for proteins of unknown function. The *dcm* transposon is localized on a larger genomic island of 126 Kb (Vuilleumier *et al*., 2009). Dichloromethane utilization in *M. extorquens* DM4 generates formaldehyde production, which is readily oxidized into formate via the H_4MPT-dependent pathway. Formate is reduced into CH_2=H_4F, which enters the serine cycle for biomass production, or is substrate to formate dehydrogenase (FDH) to produce CO_2 and energy (Figure 1.10).

3.3. Bacterial degradation of chloromethane

Bacteria affiliated to various genera and isolated from contrasting environments were described as able to degrade chloromethane (Table 1.5). Chloromethane can serve as carbon and energy source but in other cases, chloromethane is degraded and not assimilated as a carbon source in both oxic and anoxic conditions (Traunecker *et al*., 1991; Han and Semrau, 2000; Freedman *et al*., 2004).

3.3.1. Degradation under anoxic conditions

Chloromethane dehalogenation in anaerobic condition is catalyzed by corrinoid-containing enzyme which transfers the methyl group (CH_3) to H_4F. Products of this reaction are chloride ions and methyl-H_4F, which is then oxidized into formate and CO_2 or assimilated in the biomass via the acetyl-CoA pathway (Harper, 2000). *Acetobacterium dehalogenans* strain MC is able to utilize chloromethane in anaerobiosis (Table 1.5) (Traunecker *et al*., 1991; Harper, 2000). This dehalogenation requires a corrinoid-dependent methyltransferase, which in unlike other corrinoid-dependent methyltransferases, maintains cobalt int the corrinoid cofactor using an ATPase dependent system (Studer *et al*., 1999). *Pseudomonas aeruginosa* NB1 is also able to utilize chloromethane as unique source of carbon and energy in nitrate respiration conditions (Freedman *et al*., 2004), by a dechlorination mechanism that remains unknown.

3.3.2. Degradation in oxic conditions

The first described strain able to utilize chloromethane as sole source of carbon and energy was *Hyphomicrobium* sp. MC1 (Hartmans *et al*., 1986). Like this strain, the majority of the

chloromethane-degrading strains isolated so far utilize chloromethane under oxic conditions. They were isolated from various environments such as petrochemical factory soil, wastewater treatment plans, forest soils, freshwater and marine environments (Tableau 1.5). They are affiliated to the class of *Alphaproteobacteria* of the genera *Methylobacterium*, *Hyphomicrobium* and *Aminobacter*. As of this date, the *cmu* pathway is the best-characterized bacterial chloromethane-degradation pathway with a majority of the studies performed in *M. extorquens* CM4, but uncharacterized *cmuA*-independent pathway exist (see below).

– Chloromethane degradation pathway in *Methylomicrobium album* BG8

The methanotroph *Methylomicrobium album* BG8 degrades chloromethane in oxic conditions, independently of the *cmu* pathway, only in presence of another carbon source such as methanol (Figure 1.13) (Han et Semrau, 2000).

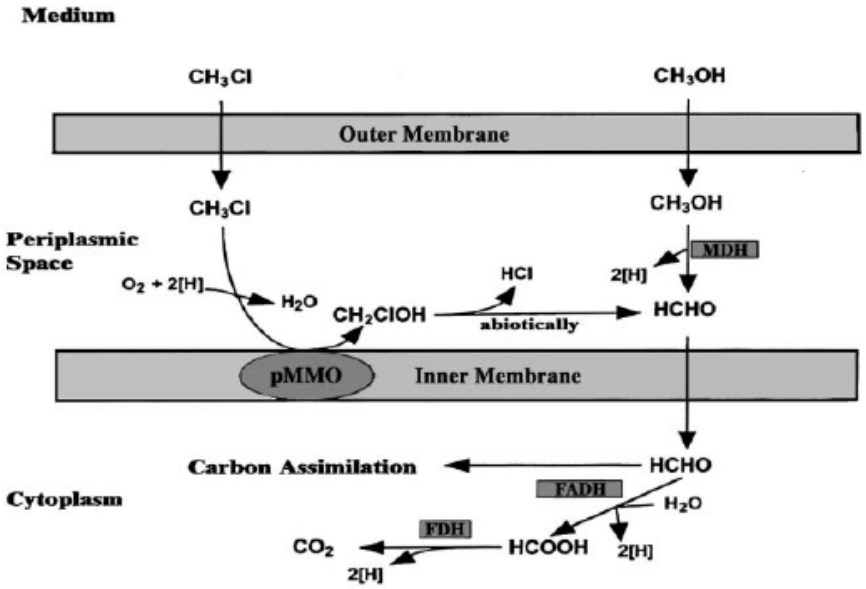


Figure 1.13. Proposed mechanism of chloromethane oxidation in *Methylomicrobium album* BG8

Chloromethane can be oxidized by the particulate methane monooxygenase (pMMO) since *M. album* BG8 stops degrading chloromethane in presence of acetylene, a pMMO inhibitor. Methanol is oxidizes into formaldehyde (HCHO) by the methanol dehydrogenase (MDH). Part of the produced formaldehyde enters the C_1 metabolism to produce biomass and the rest is oxidized into CO_2 and reducing equivalents (2[H]) for energy production via the formaldehyde deshydrogenase (FADH) and the formate dehydrogenase (FDH). Compared to methanol, chloromethane degradation generates 2 reducing equivalents less (From Han and Semrau, 2000).

Chloromethane stimulates *M. album* BG8 growth with methanol, and when using [^{14}C]-labeled chloromethane, it was demonstrated that the labeled carbon was assimilated into the biomass (38%) and released as CO_2 (50%). This strongly suggested that chloromethane can be used as carbon and energy source when methanol was present (Han and Semrau, 2000). When provided alone, chloromethane is unable to sustain growth and serve as the sole source of carbon. The authors suggested that *M. album* BG8 would not be able to generate enough reducing equivalents from chloromethane degradation to sustain growth with chloromethane. In presence of another carbon source, in larger quantities, the chloromethane utilization was observed (5 mM of methanol and 2.6 mM of chloromethane) (Han and Semrau, 2000).

– The *cmu* pathway

Combined biochemical and genetic studies in *M. extorquens* CM4 were used to identify the *cmu* (chloromethane utilization) pathway under oxic conditions (Vannelli *et al*., 1998, 1999). *M. extorquens* CM4 metabolizes one mole of CO_2 and one mole of chloride (Vannelli *et al*., 1998). Using random mutagenesis with the minitransposon *Tn*5, 9 mutants were detected with wild-type growth with methanol or methylamine, but impaired growth with chloromethane (Vannelli *et al*., 1998). Analysis of the *Tn*5 insertion sites in mutants that lost the ability to grow with chloromethane, demonstrated the essential role of *cmuABC*, *purU* and *metF2* (Studer *et al*., 2001, 2002). In addition, *cmuAB* genes were demonstrated to be essential for chloromethane dechlorination (Studer *et al*., 2001b, 2002). CmuA and CmuB catalyze the methyl group transfer from chloromethane to H_4F to form methyl-H_4F (CH_3-H_4F) (Figure 1.14) (Studer *et al*., 2001). CH_3-H_4F is oxidized either into formate and then CO_2 for energy production, or into methylene-H_4F for biomass production via the serine cycle (Vannelli *et al*., 1999).

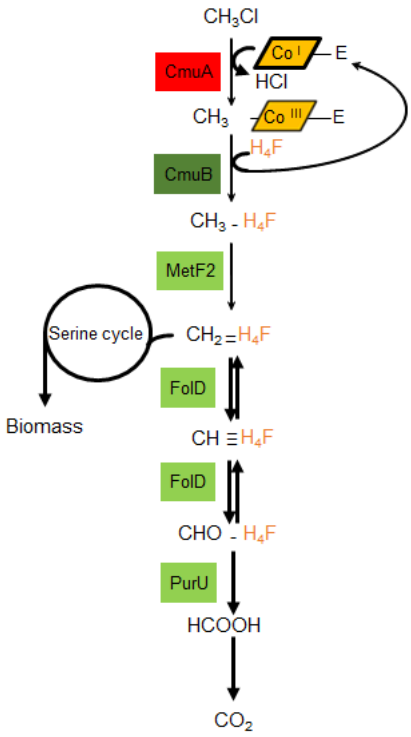


Figure 1.14 The *cmu* pathway of *Methylobacterium extorquens* CM4
The methyl group of chloromethane is transferred to tetrahydrofolate (H_4F) by *cmuA* and *cmuB*. The methylene-H_4F produced by the oxidation of methyl-H_4F enters the serine cycle to produce biomass and is also oxidized to CO_2 to produce energy. CmuA, protein with a methyltransferase domain and a corrinoid-binding domain (Studer *et al.,* 2001), CmuB, methyltransferase methylcobalamin: H_4F (Studer *et al.,* 1999); MetF2, methylene reductase; FolD, bifunctional enzyme methylene-H_4F dehydrogenase/ methenyl-H_4F cyclohydrolase; PurU, 10-formyl-H_4F hydrolase.

All the genes of *cmu* pathway in *M. extorquens* CM4 are plasmid pCMU01-borne in a region of 180 kb (Figure 1.15). The *cmu* genes are clustered in two groups 30 kb apart that displays an atypical GC content and is devoid of detectable mobile elements. Upstream of *purU* and *folD*, gene *folC2* encoded a bi-functional folylpolyglutamate synthase / dihydrofolate synthase protein. This region contains also *paaE-like*, *fmdB* and *hutI*, encoding respectively a putative oxidoreductase, a putative transcriptional regulator and a putative imidazole hydrolase.

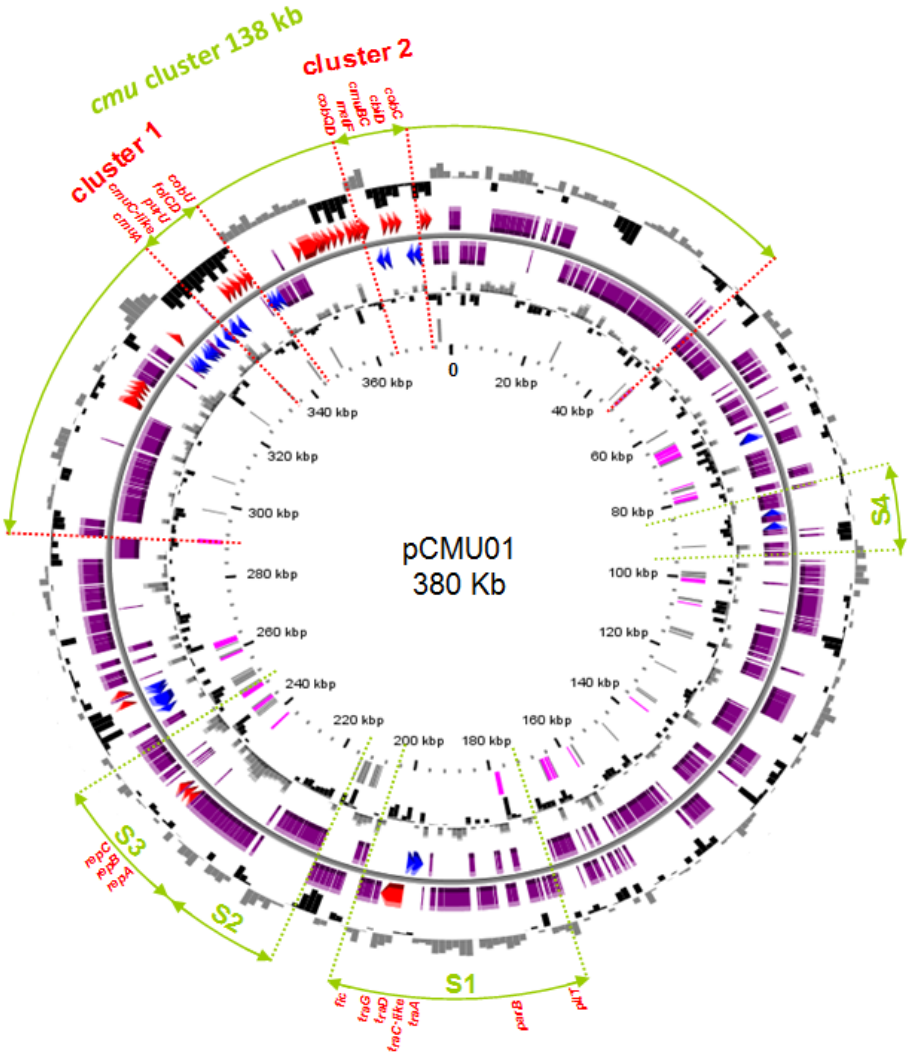


Figure 1.15 The chloromethane utilization pCMU01 plasmid in *M. extorquens* CM4
The 21 kb distant clusters 1 and 2 contain the essential genes for chloromethane utilization (*cmu*) known before the genome has been sequenced (Vannelli *et al*., 1999; Studer *et al*., 2002). The *cmu* region covers 40 kb (cluster 1 and 2). The S1, S2, S3 regions correspond to a region of 69 kb conserved between pCMU01 and the p1METDI plasmid of *M. extorquens* DM4. The S4 region shows a synteny with a duplicated region on *M. extorquens* DM4 chromosome. Circles represent (from the outside): 1, percentage of deviation of the GC content in a 1 kb window; 2, predicted CDS transcribed in the clockwise direction; 3, predicted CDS transcribed in the counterclockwise direction; 4, GC skew (G+C/G-C) in a 1 kb window; 5, transposable elements (in pink) and pseudogenes (in grey) (adapted from the PhD manuscript of Sandro Roselli, 2009).

This region contains 23 genes involved in biosynthesis and transport of cobalamin, those involved on H_4F biosynthesis, both cofactors are essential for chloromethane utilization via the *cmu* pathway (Figure 1.15). Some of these plasmid-borne genes have chloromsomal homologs (Roselli *et al*., 2013). Genome comparative analysis of strains harboring *cmu* genes, demonstrated two types of *cmu* gene organization; one found in *M. extorquens* CM4 where

genes are localized in two clusters spaced by 30 kb and a second one, found in others bacteria, where *cmu* are contiguous (Figure 1.15). A high degree of similarity in *cmu* genes organization where found in chloromethane-degrading strains (Figure 1.15, Schäfer *et al*., 2007), all affiliated to the class of *Alphaproteobacteria*. The genes order into the *cmu* cluster is highly conserved (Nadalig *et al*., 2011). In *M. extorquens* CM4, *cmuB* and *cmuC* genes are part of the same transcriptional unit, unlike gene *cmuA*, which has its own promoter (Studer *et al*., 2002). The genome of some chloromethane degrading-strains lack detectable *cmu* genes as in *Roseovarius* sp. 217 and *Leisingera methylohalidovorans* MB2 (Schäfer *et al*., 2007). Chloromethane stable isotope analysis revealed different isotopic fractionation signatures for the carbon and hydrogen in *cmu*-containing strains and *L. methylohalidovorans* MB2 which not possess those genes (Nadalig *et al*., 2014). The *cmu*-independent chloromethane degradation pathway in *L. methylohalidovorans* MB2, needs to be solved.

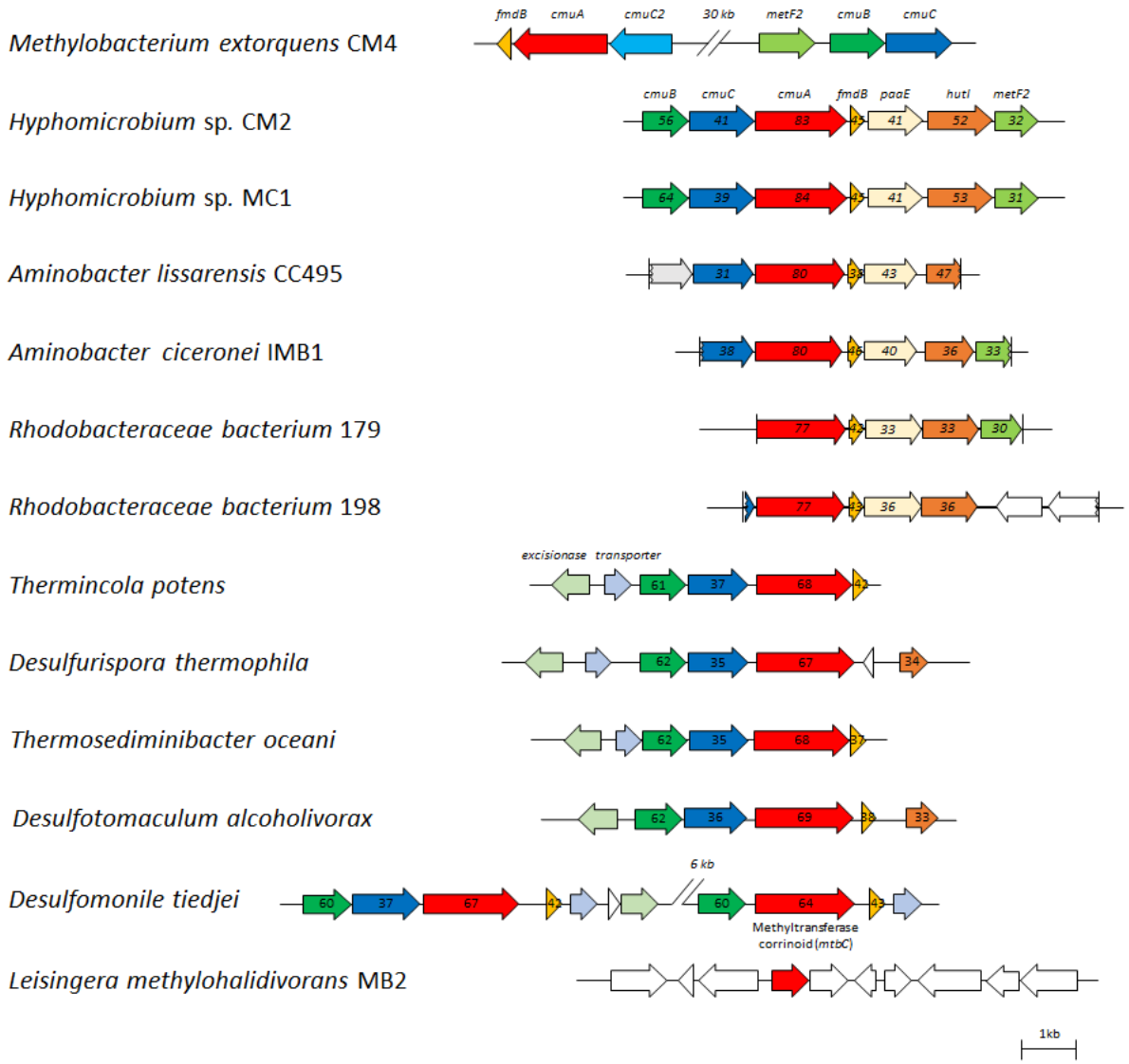


Figure 1.16. Comparison of *cmu* gene organization in bacteria harboring gene *cmuA*
Encoding genes are represented with arrows. Genes with an incomplete determined sequenced are represented by incomplete arrows. The gene reference colors are based on those from *M. extorquens* CM4 representation, where they are separated by 30 kb. Homologous genes are represented by the same color, the percentage of homology with proteins encoded by *M. extorquens* CM4 genes are written into arrows. The genetic representation is represented at the scale. To note that in the strain MB2, only one part of *cmuA*, corresponding to corrinoid cofactor binding-domain is present.

4. The *Methylobacterium* model

The genus *Methylobacterium* was first described in 1976 (Patt *et al.*, 1976), and now encompasses 20 species. This genus is affiliated to *Alphaproteobacteria*, of the order *Rhizobiales* and of the *Methylobacteriacaea* family as deduced from nucleotidic sequence analysis of 16S rRNA encoding genes (Lidstrom, 2006). They are facultative methylotrophs, strictly aerobic, mesophiles with an optimal temperature between 25 and 30°C, Gram

negative. They test positive with the catalase and oxidase assays. Cells often display poly-β-hydroxybutyrate inclusions involved in energy and carbon storage. They display a pink coloration due to carotenoids (Korotkova and Lidstrom, 2001; Van Dien *et al*., 2003). *Methylobacterium* strains have been isolated from contrasting environments either handmade (road borders, water tanks) or natural (soil, leaf surface) (Tableau 1.6) (Radajewski *et al*., 2002; Nadalig *et al*., 2011; Bai *et al*., 2015). They metabolize C_1 compounds such as methanol and methylamine, and a few strains also degrade toxic chlorinated methanes such as dichloromethane or chloromethane (Dourado *et al*., 2015). Methylotrophs have been among the first bacteria to be studied in the environment using molecular approaches (Holmes *et al*., 1995; Kolb et Stacheter, 2013).

4.1. *Methylobacterium*, a model for methylotrophy studies

One of the most studied species of the genus *Methylobacterium* is *M. extorquens* which has been isolated from various environments such as plants, soil, waste waters or cloud water (Bai *et al*., 2015; Bringel et Couée, 2015; Nadalig *et al*., 2011; Ochsner *et al*., 2015; Temkiv *et al*., 2012). *M. extorquens* AM1 is the most studied strain which has been used along with *Methylobacterium organophilum*, as *Methylobacterium* reference strains, in plant-microorganism interaction studies (Dourado *et al*., 2015). *M. extorquens* AM1 harbours five replicons and 174 insertion sequences, which complicates its utilization in molecular biology studies. Thus, strain AM1 has been replaced by *M. extorquens* PA1 whose genome consists of a unique chromosome and only 20 insertion sequences (Nayak et Marx, 2014). *M. extorquens* strains (AM1, BJ001, CM4, DM4, DSM 13060 and PA1) are able to utilize methanol, but only a few of those are able to utilize chlorinated methanes as a sole carbon and energy source (Vuilleumier *et al*., 2009; Marx *et al*., 2012). *M. extorquens* CM4 has been isolated from a Russian petrochemical industry soil, and is able to grow with chloromethane (Doronina *et al*., 1996). On the other hand, *M. extorquens* DM4 able to grow on dichloromethane, has been isolated from a chlorinated alkane-polluted Swiss soil (Gälli et Leisinger, 1985). *M. extorquens* genomes display high GC content of at least 68% with a large chromosome sized between 5.5 and 5.9 Mb (Tableau 1.6).

Unlike strains PA1 and DSM 13060, further *M*. *extorquens* strains harbour at least one plasmid, including *M. extorquens* strain AM1, which remarkably harbors of a megaplasmid (Ochsner *et al*., 2015). At the end of December 2015, 24 additional *Methylobacterium* genomes have been

sequenced from strains isolated from plant phyllosphere and rhizosphere (Bai *et al.*, 2015). These recently sequenced genomes harbour no *cmuA* gene homologue.

Table 1.6. Characteristics of the first *Methylobacterium* sequenced genomes

Parameters[a]	*M. extorquens* AM1	*M. extorquens* BJ001[b]	*M. extorquens* CM4	*M. extorquens* DM4	*M. extorquens* DSM13060	*M. extorquens* PA1	*M. nodulans* ORS 2060	*M. radiotolerans* JCM 2831	*M.* sp. 4-46
Genome									
Size (Mb)	6,88	5,85	6,18	6,12	6,73	5,47	8,84	6,90	7,73
% GC	68,5	69,4	68,0	68,0	68,23	68,2	68,4	71,0	71,5
Number of replicon	5	3	3	3	1	1	8	9	3
Number of predicted CDS	6335	5945	6276	5706	7132	5153	10333	7356	8479
Genbank accession number	NC_012808	NC_010725	NC_011757	NC_012988.1	NZ_AGJK00000000.1	NC_010172	NC_011894	NC_010505	NC_010511
Chromosome									
Size (Mb)	5,51	5,8	5,77	5,94	6,73	5,47	7,78	6,1	7,70
% GC	68,7	69,4	68,2	68,1	68,23	68,2	68,9	71,5	71,6
Number of plasmids	4	2	2	2	0	0	7	8	2
Sequencing (date)	UW Genome Sequencing Center, (2003/2006)	JGI (2008)	JGI (2008)	Génoscope (2006)	JGI (2010)	JGI (2007)	JGI (2009)	JGI (2008)	JGI (2008)
Genome publication	Vuilleumier *et al.*, 2009	Marx *et al*., 2012	Marx *et al*., 2012	Vuilleumier *et al.*, 2009	Koskimäki *et al*., 2015	Marx *et al*., 2012	Marx *et al*., 2012	Marx *et al*., 2012	Marx *et al*., 2012

[a] Génoscope (http://www.cns.fr/agc/microscope/home/index.php) or Craig Venter institute (http://www.jcvi.org/)

[b] Originally named *Methylobacterium populi* (Van Aken *et al*., 2004), but based on 16S rRNA identity, was affiliated to *M extorquens* (Marx *et al*., 2012)

4.2. Genetic tools available for *M. extorquens*

Different genetic tools are available for *M. extorquens* AM1. They include tools for random mutagenesis, allelic gene replacement for targeted gene deletion and insertion, but also for modulated gene expression (constitutive or inducible). A complete list of genetic tools for *M. extorquens* AM1 is detailed in a recent review (Ochsner *et al*., 2015).

4.3. Global approaches in *M. extorquens*

Our knowledge of methylotrophy in bacteria belonging to the genus *Methylobacterium* and in particular in the species *M. extorquens* has increased considerably over last years with high-throughput *Next Generation Sequencing* (NGS) technologies, allowing the study of genes (genomics), their expression at the RNA level (transcriptomics), at the protein level (proteomics) as well as of cell metabolites (metabolomics).

4.3.1. Genomic studies in *M. extorquens*

In 1998, the first *Methylobacterium* genome sequence was obtained from strain AM1 (Chistoserdova *et al*., 2003). It has been completed and assembled only in 2009 when the genome of strain DM4 was available (Vuilleumier *et al*., 2009). Later, in 2012, the genome of strain *M. extorquens* CM4 was sequenced simultaneously with the genomes of other *M. extorquens* strains unable to degrade chlorinated methanes (BJ001; PA1) (Marx *et al*., 2012). When I started my PhD thesis, the MaGe platform hosted by the Génoscope (Evry, France), offered easy annotation and comparative analysis of 6 *M. extorquens* genomes (AM1, BJ001, CM4, DM4, DSM 13060 and PA1), with characterized growth abilities for methanol and chlorinated methanes. Intra-species *M. extorquens* comparative genomic study has detected the presence of a high number of insertion sequences (IS) in the vicinity of gene clusters involved in methanol utilization, suggesting a role of these mobile elements in the evolution of phylogenetically closely-related genomes of strains AM1 and DM4 (Vuilleumier *et al*., 2009). Nevertheless, the history of the evolution of *M. extorquens* genome remains unclear. As a matter of facts, despite the fact that *M. extorquens* AM1 and PA1 have similar GC content (68.2% and 68.5%, respectively) and share a large number of common genes involved in methylotrophy (90 genes with at least 95% identity between their corresponding encoded proteins), a recent study highlighted differences in growth rates for a variety of C_1 compounds

such as methanol, methylamine or formaldehyde, which would suggest the implication of different mechanisms (Nayak et Marx, 2014).

Combined random mutagenesis and comparative proteomic studies have evidenced the growth with chlorinated methanes involved genes localized on genomic islands, in addition to housekeeping genes involved in stress response, and central metabolism (Michener *et al*., 2014a; Muller *et al*., 2011a; Roselli *et al*., 2013). Chloromethane or dichloromethane utilization generates a cellular stress that requires specific adaptive responses (Kayser et Vuilleumier, 2001; Kayser *et al*., 2002; Roselli *et al*., 2013). To decipher those adaptive mechanisms, Michener *et al.* (2014) have tested the ability to utilize dichloromethane of naïve non-degrading strains. The plasmid-borne *dcmA* gene has been transferred to strain AM1 (unable to utilize dichloromethane for its growth) (Michener *et al*., 2014b). Transconjugants displayed poor growth with dichloromethane and but had a detectable dehalogenation activity. The subsequent genome sequencing of adapted clones enabled the identification of mutation within gene *clcA* encoding an antiporter Cl^-/H^+. When a plasmid harbouring genes *dcmA* and the mutated *clcA* were introduced in *M. extorquens* PA1 (unable to utilize dichloromethane for its growth), it readily conferred the ability to growth with dichloromethane.

Complementary studies of chloromethane utilization and adaptation in naïve strains (unable to utilize chloromethane for their growth) have been performed (Michener *et al*., submitted article). A plasmid harbouring genes of the *cmu* pathway (*folD*, *metF2*, *paaE-like*, *purU*) and of associated genes (*fmdB*, *hutI*), has been inserted by conjugation in *M. extorquens* AM1, PA1, BJ001, *M. radiotolerans* and *M. nodulans* strains. The resulting transconjugants acquired the ability to growth with chloromethane, although cultures were of low optical densities. When the mutated gene *clcA*, which has been demonstrated to increase cell fitness for growth with dichloromethane, was also present in the transconjugants, no improvement of the fitness was observed for cells grown with chloromethane. This suggest that the effectiveness of heterologous catabolism of chloromethane and dichloromethane are uncorrelated in *Methylobacterium* strains (Michener *et al*., submitted article).

There remains a need for a better understanding of the mechanisms involved in methylotrophic growth in *M. extorquens* using complementary approaches such as transcriptomics, proteomics or metabolomics in addition to the studies already available (Tableau 1.7 completed from Ochsner *et al*., 2015).

Table 1.7. "Omics" studies in *M. extorquens* strains

« omic » study	Reference	Description
Genome	Vuilleumier *et al*., 2009	AM1, DM4
	Roselli *et al*., 2013	CM4
	Muller *et al*., 2011	DM4
	Bai *et al*., 2015	Isolates from phyllosphere and rhizosphere (24 genomes)
Modelling	Peyraud *et al*., 2011	AM1
Proteomic	Bosch *et al*., 2008; Laukel *et al*., 2004	AM1, methanol and succinate comparison
	Muller *et al*., 2011	DM4, dichloromethane et methanol comparison
	Roselli *et al*., 2013	CM4, chloromethane et methanol comparison
	Guo et Lidstrom, 2008	AM1, « Profiling » of metabolites in methanol and succinate
Transcriptomic	Okubo *et al*., 2007	AM1, methanol and succinate comparison (microarray)
	Francez-Charlot *et al*., 2009	AM1, *phyR* mutant, general stress regulation (microarray)
Metabolomic	Kiefer *et al*., 2008, 2011	AM1, central metabolites concentration in methanol and succinate
	Peyraud *et al*., 2012	AM1, [^{13}C]-fluxomic, [^{13}C]-CH_3OH in co-utilization with succinate
	Yang *et al*., 2013	AM1, ^{13}C methanol, metabolite labeling during growth
	Cui *et al*., 2016	AM1 with overexpression of ethylmalonyl-CoA mutase in methanol
	Reaser *et al*., 2016	AM1, monitoring over time of metabolite after incorporating [^{13}C]-CH_3OH

4.3.2. Transcriptomic studies in *M. extorquens*

Transcriptomics is a qualitative and quantitative study of the transcriptome, which encompasses in principle all RNAs produced by transcription (Wang *et al*., 2009). In different growth conditions, the spectrum of transcripts and their abundance is modified. Thus, the quantitative inventory of transcripts enables a better understanding of the adaptive processes involved in response to growth in various conditions.

Today, few transcriptomic studies have been performed in *M. extorquens*, and when performed, it was using DNA chips (microarrays) in strain AM1. These studies demonstrated the role of genes involved in C_1 metabolism in comparative studies of cultures grown with methanol versus succinate (Okubo *et al*., 2007; Francez-Charlot *et al*., 2009), as well as the involvement in the stress response of the transcriptional gene regulator *phyR* (Francez-Charlot *et al*., 2009). Sequencing and analysis of RNA (RNA sequencing, i.e. a global transcriptome analysis) has not been described so far for *M. extorquens,* and I provide in this doctoral work the first transcriptome study for strains CM4 and DM4.

4.3.3. Proteomic studies in *M. extorquens*

There are more published proteomic than transcriptomic studies in *M. extorquens* (Tableau 1.7, Gourion *et al*., 2006; Muller *et al*., 2011; Roselli *et al*., 2013). In *M. extorquens* AM1, proteins involved in methylotrophy have been detected in cultures grown with methanol (the reference growth substrate for methylotrophy studies) compared to cultures grown with succinate ($C_4H_6O_4$) (Laukel *et al*., 2004; Bosch *et al*., 2008). Other studies were performed in *M. extorquens* strains able to degrade chlorinated methanes. In *M. extorquens* DM4, the protein content of cultures grow with dichloromethane compared to methanol was used to identify proteins involved in dichloromethane utilization (see chapter 3) (Muller *et al*., 2011a). Similarly, *M. extorquens* AM1 colonization mechanisms of

A. thaliana phyllosphere has been better understood using proteomic approaches (Tableau 1.7, Knief *et al*., 2010, 2012; Vorholt, 2012). Overexpression of the regulatory protein PhyR was employed to evidence its role in plant colonization (Gourion *et al*., 2006).

Proteomic analysis in *M. extorquens* CM4 chloromethane utilization was used to demonstrated that 49 proteins more abundant in cultures grown with chloromethane compared tomethanol (Figure 1.17).

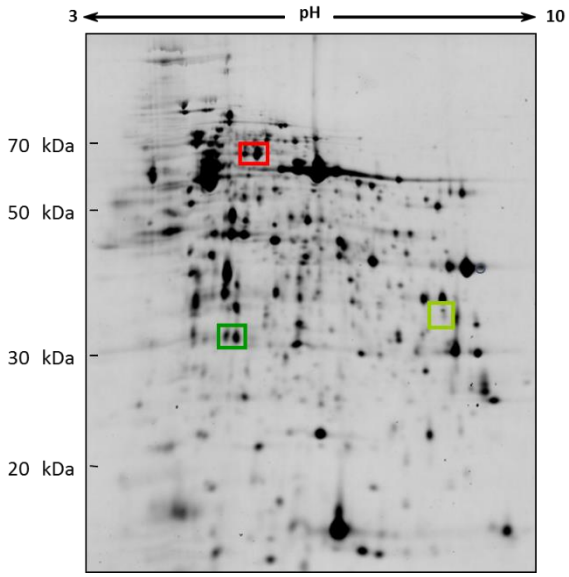


Figure 1.17. 2D gel picture of protein extracts of *M. extorquens* CM4 grown with chloromethane or with methanol
CmuA, CmuB and PurU framed in red, dark green and light green respectively, were identified by mass spectrometry and were more abundant with chloromethane (Sandro Roselli PhD thesis, 2009).

By combining genomic and proteomic approaches, genes involved in chloromethane utilization highlighted the central implication of plasmid pCMU01 in growth with chloromethane in *M. extorquens* CM4 (Roselli *et al*., 2013).

4.3.4. Metabolomic studies in *M. extorquens*

Most of the metabolomic studies were performed with cultures of *M. extorquens* AM1 to better characterize the methylotrophic and central metabolism of carbon assimilation. In cultures grown with methanol compared tosuccinate, metabolites specifically associated to growth in methylotrophic or involved in glyoxylate regeneration such as β-hydroxybutyrate, methylsuccinyl-CoA or ethylmalonyl-CoA have been detected (Figure 1.11) (Guo et Lidstrom, 2008). Metabolites profiling were also performed in cultures of
M. extorquens AM1 grown with compounds in C_2 (ethylamine) and in C_4 (succinate) (Yang *et al*., 2009). Similarly, *M. extorquens* AM1 grown with acetate (CH_3COO^-) as the sole source of energy and carbon was used to demonstrated that the ethylmalonyl-CoA pathway replaced the isocitrate lyase pathway for the regeneration of glyoxylate, needed for carbon assimilation in biomass via the serine cycle (Schneider *et al.,* 2012, Figure 1.11). These results illustrate metabolic adjustments to growth with different carbon sources.

5. PhD thesis aims and questions

Since the Montreal protocol in 1987, limiting the utilization of chlorinated compounds, chloromethane is responsible for 16% of ozone layer destruction related to chlorinated compounds (Montzka *et al*., 2011). The global budget of chloromethane remains uncertain even if sources and sinks are identified (Tableau 1.4 and Tableau 1.5, (Montzka and Fraser, 2003; Clerbaux *et al*., 2007). The role of the soil, both source and sink of chloromethane, in this global budget is shown (Gribble, 2003; Harper *et al*., 2003; Harper, 2000; Miller *et al*., 2004). Otherwise uncertainties remain regarding the importance of the soil contribution in chloromethane exchanges with the atmosphere (Saito *et al*., 2008). One of those uncertainties comes from the bacterial capacity to play a role as a chloromethane emission filter, limiting chloromethane into the atmosphere. Otherwise, low chloromethane concentrations available in forest soil (1 pM. g^{-1}) (Harper *et al*., 2003) could not allow enough to support an efficient growth for chloromethane degrader communities, requiring other carbon and energy sources to grow. In fact, methanol concentration, estimated at 1 nM. g^{-1}, in forest soils, is higher than those of chloromethane (Figure 1.18, Stacheter and Kolb, 2013). In the environment, methylotrophic microorganisms that use chloromethane could also use methane and methanol. Methane and methanol emissions form terrestrial ecosystems are within same range (10^{12} mol per year) (Kolb, 2009a). Methanol is a central compound of methylotrophic metabolism, as more than 83% of methylotrophic strains growing in aerobic conditions were isolated from soil and were able to grow using methanol (Kolb, 2009). It is also shown in the strain *M. extorquens* CM4, which is unable to use methane (Van Aken, 2004). One of the objectives of my PhD thesis was to characterize bacterial populations within forest soil that are able to use chloromethane, and to test the hypothesis of chloromethane and methanol co-utilizations. My PhD project was articulated around two major axes. First, an in depth study of chloromethane utilization in controlled laboratory conditions. The sequencing of transcripts (RNAseq), was done to identify genes differentially expressed in *M. extorquens* CM4 in methylotrophic growth conditions with chloromethane and methanol. The second part of my PhD work, based on a Stable Isotope Probing (SIP) approach (Neufeld *et al*., 2007), consisted of highlighting methylotrophic bacteria active in forest soil, and its ability to degrade chloromethane. In this case, incubations with chloromethane and/or methanol labelled with [^{13}C], were created to detect bacteria able to assimilate those compounds, and to incorporate the [^{13}C] into the biomass. Enriched [^{13}C]-DNA (“Heavy” DNA), corresponding to the DNA of

microorganisms assimilating the labelled-carbon source, was separated from the DNA not labelled ("Light" DNA constituted only by [^{12}C]) by ultracentrifugation. Apart from the heavy DNA, the bacterial diversity is estimated by sequencing of PCR products of taxonomic and functional markers.

Three gene markers were used targeting:

- Aerobic degradation of the chloromethane (*cmuA* gene encoding a chloromethane methyltransferase belonging to the *cmu* pathway)
- Aerobic degradation of the methanol (*mxaF* and *xoxF* genes encoding a methanol dehydrogenase subunit)
- Central methylotrophy metabolism (*mch* gene) (Stacheter and Kolb, 2013).

The main goal of those combined approaches was:

To have a better understanding of chloromethane utilization in model strains cultivated in the laboratory, to identify new biomarkers of chloromethane degradation, and to detect more efficiently new biomarkers of chloromethane degradation in forest soils.

Identify new bacterial methylotrophy taxa, active in aerobic chloromethane consumption in sub-surface forest soil microcosms, and the study of their functions within the bio-capture of terrestrial chloromethane emissions.

Figure 1.18. Sinks and emissions of chloromethane and methanol in forest soil
Chloromethane emissions data from Harper *et al.*, 2003. Methanol fluxes from Stacheter *et al.*, 2013.

Chapitre 1

Introduction

Chapter 1. Introduction (en français)

Ce premier chapitre débutera par la présentation des méthanes chlorés en général, et en particulier du chlorométhane, avec la description de leurs propriétés chimiques, leurs utilisations industrielles, et leurs impacts sur la santé et l'environnement. Suivront des paragraphes dédiés spécifiquement au chlorométhane, à ses sources de production et ses puits de consommation, notamment par les microorganismes. Les connaissances sur la dégradation microbienne du chlorométhane seront ensuite présentées dans le contexte du métabolisme des composés à un atome de carbone et la génétique de souches modèles méthylotrophes. Enfin, le dernier paragraphe posera les problématiques et les objectifs de mes travaux de thèse.

1. Les méthanes chlorés

Les méthanes chlorés sont des composés organiques ne comportant qu'un seul atome de carbone relié à un ou plusieurs atomes de chlore. Un atome de carbone possède quatre liaisons covalentes, il existe donc quatre méthanes chlorés, le chlorométhane (CH_3Cl), le dichlorométhane (CH_2Cl_2), le chloroforme ($CHCl_3$) et le tétrachlorométhane (CCl_4). A l'exception du chlorométhane qui se trouve à l'état de gaz à température ambiante, les autres méthanes chlorés sont liquides et incolores (Huang *et al*., 2014). Ces composés ont des propriétés physico-chimiques différentes et sont, depuis le 20^{e} siècle, produits et utilisés en grande quantité dans l'industrie (Tableau 1.1). Les méthanes chlorés peuvent également être produits naturellement. Les productions naturelles de chloroforme et de dichlorométhane sont estimées respectivement à 0,56 et 0,25 Tg.an^{-1} (Cox *et al*., 2003; Xiao, 2008). Cependant, le chlorométhane est le méthane chloré le plus présent dans l'atmosphère avec 2,8 Tg.an^{-1} (Ruecker *et al*., 2014) (Tableau 1.2).

Tableau 1.1. Propriétés et utilisations des méthanes chlorés

Composé	Apparence	Solubilité dans l'eau ($g.L^{-1}$)	Effet sur la santé[a]	Polluant[b]	Utilisation dans l'industrie[c]
chlorométhane	gaz incolore	5,3	groupe 3	non	production de silicone, produits de construction
dichlorométhane	liquide incolore	13,0	groupe 2B	oui	décapant pour peinture, produits pharmaceutiques
chloroforme	liquide incolore	8,0	groupe 2B	oui	solvants, synthèse de produits chimiques
tétrachlorométhane	liquide incolore	0,8	groupe 2B	oui	produits pharmaceutiques, produits de nettoyage, produits agrochimiques

[a] Données issues de l'Agence internationale de recherche contre le cancer (IARC). Le groupe 2B regroupe les composés probablement cancérigènes pour l'Homme. Le groupe 3 regroupe les composés non classifiables comme carcinogène pour l'Homme (Huang *et al*., 2012).
[b] Composés reconnus comme polluants par les Etats-Unis, la Chine et pour le dichlorométhane et le chloroforme reconnus également par l'Union Européenne (Huang *et al*., 2014).
[c] Données issues de Frascari *et al*., 2015.

Les méthanes chlorés sont présents dans les sols, l'eau et l'air (http://www.epa.gov). A l'exception du chlorométhane, ils ont été classés par l'IARC (International Agence for Research on Cancer) comme étant potentiellement cancérigènes, et comme polluants par de nombreux pays (Tableau 1.1). L'impact des composés chlorés dans l'environnement fait qu'ils sont très étudiés (Harper, 2000).

Tableau 1.2. Les méthanes chlorés dans la troposphère

	Troposphère		Flux (Gg. an^{-1})	
	Conc. (pptv)	Demi-vie (année)	Industrie	Naturel
chlorométhane	550 ± 30[a]	1,3	10[a]	1 900[a]
dichlorométhane	25 ± 5[d]	0,38	530[e]	160[e]
chloroforme	15 ± 5[c]	0,41	60[b]	595[b]
tétrachlorométhane	95 ± 5[f]	26	46[g]	~0[g]

D'après manuscrit de thèse de S. Roselli (2009)
[a] Montzka et Fraser, 2003
[b] McCulloch *et al.*, 2003
[c] O'Doherty *et al*., 2001
[d] Schauffler, 2003
[e] Kurylo et Rodriguez, 1998
[f] Prinn *et al*., 2000
[g] UNEP, 2005

Dans la stratosphère, les méthanes chlorés sont activés par les ultra-violets. Il en résulte la formation de radicaux $Cl^{\cdot}$, qui vont contribuer à la déplétion de la couche d'ozone en catalysant la formation d'O_2 (Figure 1.1). Il est estimé qu'un atome de chlore peut détruire 100 000 molécules d'ozone (http://www.epa.gov/ozone/science/process.html), avec pour exemple de cet impact, la formation d'un trou dans la couche d'ozone au-dessus de l'Antarctique à chaque printemps (Clerbaux *et al*., 2007).

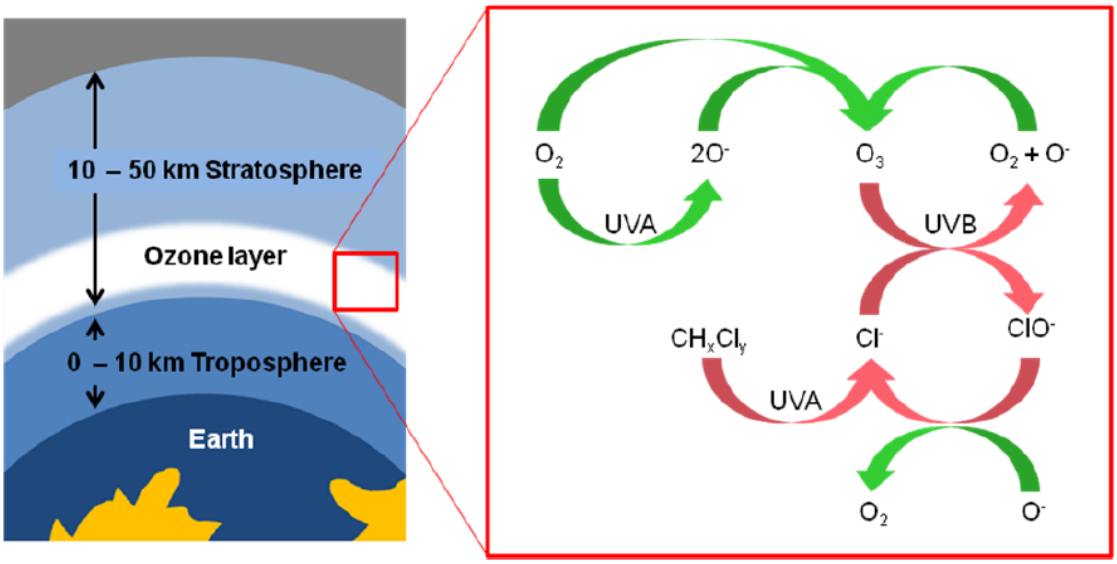


Figure 1.1. Réactions de formation et de destruction de l'ozone dans l'atmosphère
En présence de méthanes chlorés (CH_xCl_x), les réactions naturelles de l'ozone (O_3) sont perturbées. Rayons UVA entre 315 nm et 400 nm et rayons UVB entre 280 nm et 315 nm (Figure modifiée de http://www.cec.org)

En raison de leur impact sur l'environnement, l'utilisation des composés halogénés a été régulée par le protocole de Montréal en 1987, diminuant ainsi grandement leur utilisation dans l'industrie, allant jusqu'à un arrêt total de la production de composés comme le tétrachlorométhane en 1995. Ceci a eu pour conséquence d'augmenter la part du chlorométhane, le C_1 chloré le plus produit naturellement, dans l'impact environnemental des méthanes chlorés. En effet, parmi les méthanes chlorés retrouvés dans la troposphère, le chlorométhane est le plus abondant (Tableau 1.2). Aujourd'hui, on estime qu'il est responsable de 16 % de la dégradation de la couche d'ozone liée aux composés chlorés (Montzka *et al.*, 2011).

2. Le chlorométhane

Le budget global du chlorométhane, c'est-à-dire sa concentration dans l'atmosphère, est défini par la somme de ses sources de production, soustrait de l'ensemble des puits de dissipation. La diversité des sources et puits abiotiques et biotiques du chlorométhane est détaillée ci-dessous. Ce budget reste cependant incertain notamment du fait de la part des bactéries dans le filtrage du chlorométhane émis dans différents environnements.

2.1. Les sources de chlorométhane

La production totale annuelle est estimée à 2,8 Tg (Ruecker *et al*., 2014; Sailaukhanuly *et al*., 2014). Comme plus de 5 000 substances halogénées, le chlorométhane est produit majoritairement naturellement. La production naturelle de chlorométhane repose sur des sources abiotiques et biotiques (Tableau 1.3). Sa faible production d'origine anthropogénique est due à la combustion de matières fossiles comme le charbon, la combustion de déchets municipaux, ou à d'autres activités industrielles (Tableau 1.3, McCulloch et Aucott, 1999).

Tableau 1.3. Estimation des sources de chlorométhane dans l'atmosphère

Origine	Estimation (Gg. an^{-1})[a]	Valeurs basses et hautes (Gg. an^{-1})[b]
Senescence végétale	1800[d]	30/2500
Feux de forêts	911	655/ 1125
Plantes tropicales	910	820/ 8200
Océans	600	325/ 1300
Marais salants	170	65/ 440
Champignons	160	43/ 470
Combustion de matière organique fossile	105	5/ 205
Incinération	45	15/ 75
Zones humides	40	6-270[c]
Industrie	10	10
Rizières	5	2,4/ 4,9
Total des sources	*4746*	*2019/ 14378*

[a] Données issues de (Keppler *et al.*, 2005)
[b] Données issues de (Clerbaux *et al*., 2007)
[c] Données issues de (Montzka et Fraser, 2003)
[d] Données issues de (Montzka *et al*., 2011)

2.1.1. Production abiotique

Les feux de forêts constituent la plus grande source abiotique de chlorométhane (Sailaukhanuly *et al*., 2014). Le chlorométhane produit est émis dans l'atmosphère mais également retrouvé à la surface des sols, où des processus abiotiques de production de méthanes halogénés ont également été caractérisés (Keppler *et al*., 2005). Une production abiotique de chlorométhane par des réactions redox (Figure 1.2) et des réactions de substitution ont été démontrées (Hamilton, 2003; Keppler *et al*., 2005; Wishkerman *et al*.,

2008) bien que leur part dans le bugdet du chlorométhane n'ait pas été évaluée. Les océans représentent la 2e source abiotique de chlorométhane (Xiao, 2008; Hu *et al*., 2010).

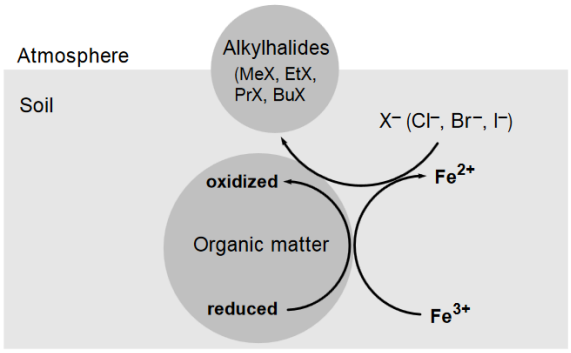


Figure 1.2. Formation abiotique du chlorométhane dans les sols
Formation des alkyls halogénés par la réduction du Fe (III) avec la matière organique en présence d'ions chlorure dans les sols (issue de Keppler *et al.*, 2003).

2.1.2. Production biotique

La production principale de chlorométhane résulte de processus biotiques chez les plantes (Hamilton, 2003; Saito *et al*., 2008), les champignons (Moore *et al*., 2005), les algues marines (Traunecker *et al*., 1991) ou encore les marais salants (Rhew *et al*., 2003) (Tableau 1.3). Les écosystèmes terrestres contribuent largement à cette production, avec plus de
50 % des émissions de chlorométhane qui en sont issues (Redeker et Kalin, 2012). Dans l'environnement, la présence d'ions chlorure est très variable et pourrait influencer les taux d'émissions de chlorométhane. Les premiers résultats d'une corrélation possible évoquée dans la littérature sont décrits ci-après ainsi que l'état des connaissances des productions du chlorométhane par les plantes d'une part et dans les sols d'autre part. On situera également les taux d'émissions du chlorométhane par rapport aux autres Composé Organiques Volatiles (COVs) et le développement d'un bio-rapporteur pour sa quantification.

– Production par les plantes

Le mécanisme génétique de production du chlorométhane a été caractérisé en détails chez la plante modèle *Arabidopsis thaliana* (Rhew *et al*., 2003). Cette étude montre la production de chlorométhane mais aussi de bromométhane (CH_3Br) et d'iodométhane (CH_3I), par l'action d'une méthyltransférase S-adénosylméthionine (SAM)-dépendante nommée *HOL* pour « *Harmless to Ozone Layer* ». Le gène *HOL* est conservé chez les plantes, laissant penser que la production de chlorométhane par les plantes est un processus répandu (Rhew *et al*., 2003;

Nagatoshi et Nakamura, 2007). Le chlorométhane est produit par transfert du groupement méthyl du SAM à un ion chlorure (Cl^-) (Figure 1.3). La SAM intervient dans de nombreuses réactions de synthèse de la lignine, un constituant de la paroi cellulaire végétale (Campbell et Sederoff, 1996). Rhew *et al*., 2003 a démontré que la méthyltransférase codée par le gène *HOL* catalyse la méthylation du chlorure, dépendamment du SAM.

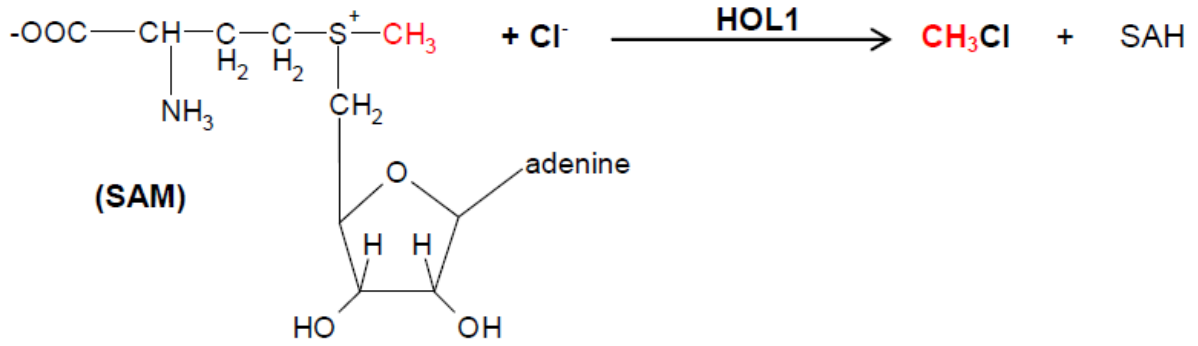


Figure 1.3. Production de chlorométhane chez *Arabidopsis thaliana*
(Figure issue et modifiée du manuscrit de thèse de Muhammad Farhan Ul Haque, Mai 2013)

De fait, différentes études ont démontré l'existence de plantes émettrices de chlorométhane. Par exemple, dans une région subtropicale, sur 187 plantes étudiées, 33 espèces appartenant à des familles différentes émettent du chlorométhane (Yokouchi *et al*., 2007). D'une manière similaire, l'étude des flux de chlorométhane en forêt tropicale a montré que 25 plantes sur les 117 étudiées émettent du chlorométhane (Saito *et al*., 2008). Les émissions les plus importantes ont été observées chez des fougères (*Osmunda banksiifolia*, *Cibotium balometz*, *Angiopteris palmiformis*), et des plantes halophiles (*Vitex rotundifolia*, *Vitex trifolia* et *Excoecaria agalloch*) avec des taux d'émission de l'ordre de 1 µg de chlorométhane.g^{-1} de matière sèche. h^{-1}. Toutefois, la capacité de production de chlorométhane dépend plus de l'espèce que de la famille à laquelle est affiliée la plante (Yokouchi *et al*., 2007).

– Production de chlorométhane par le sol et les matières végétales en décomposition

Les plantes en décomposition représentent la source majeure du chlorométhane (Hamilton, 2003, Montzka *et al*., 2011). La production de chlorométhane est surtout localisée dans l'horizon « O » des sols et semble être quasi nulle dans les horizons inférieurs (Figure 1.4) (Redeker et Kalin, 2012).

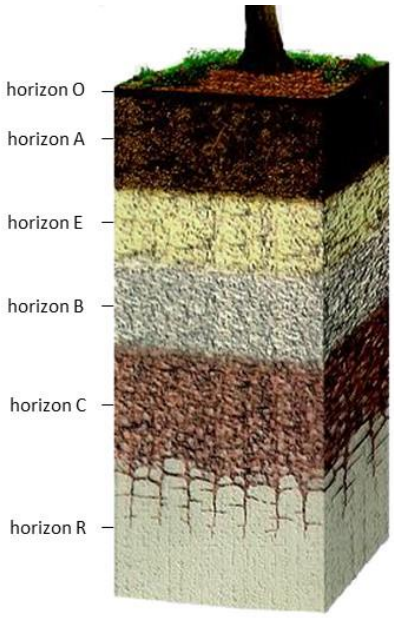


Figure 1.4. Organisation schématique du sol forestier
Le sol forestier type est organisé en différents horizons; O, A, E, B, C et R. L'horizon « O » est surtout constitué de matière organique en décomposition, l'horizon « A » constitué également de matière organique et de matière minérale, l'horizon « E » est un horizon lessivé, où les nutriments solubles ne sont plus présents en raison des précipitations ou de l'irrigation. L'horizon « B » est un horizon riche en minéraux issus du lessivage des horizons « A » et « E ». L'horizon « C » est un horizon de sous-sol, non consolidé. Enfin l'horizon R est un horizon rocheux, constitué de roche comme le granit ou encore du calcaire. Les horizons « O » et « A » sont biologiquement actifs et sont fortement oxygénées, tandis que les autres sont plus pauvres en matière organique et ont une activité microbienne plus faible. Dans le cadre de mon projet de thèse, un travail sur le sol forestier a été effectué au niveau de l'horizon « O ». (Figure modifiée de http://www.ctahr.hawaii.edu/mauisoil/a_profile.aspx)

Le sol est un environnement riche en ions chlorure, dont la biodisponibilité est essentielle à la formation du chlorométhane. Cependant, l'estimation de la concentration des ions chlorure dans les sols reste difficile car de nombreux mécanismes interviennent comme les flux entre la végétation et le sol, l'altération de la roche, ou encore la dégradation de la matière organique (Bastviken *et al*., 2007; Öberg, 2003). Il est admis qu'une partie des ions chlorure est considérée comme non biodisponible du fait de son adsorption ou de son absorption par la végétation ou les microorganismes (Bastviken *et al*., 2007). Les ions chlorure sont aussi contenus dans le bois et sont utilisés pour la synthèse de chlorométhane par les champignons les colonisant (Watling et Harper, 1998). On estime que le contenu en Cl^- dans le bois d'espèce de forêts tempérées varie entre 2,4 et 123 $mg.kg^{-1}$ de matière sèche (Watling et Harper, 1998).

Des études sur des champignons ont démontrées que des membres de la famille des *Hymenochaetaceae* sont capables de réémettre sous forme de chlorométhane une partie des ions chlorure présents dans leurs substrats (Harper et Kennedy, 1986). En effet, le bois mort en décomposition est colonisé par les champignons qui, via un ensemble de péroxydases, dégradent la lignine, un constituant majeur du bois, en SAM. La SAM, par interaction avec les ions Cl^- présents dans le milieu va, par l'intermédiaire d'une méthyltransférase, former du chlorométhane (Harper, 2000; Anke et Weber, 2006). Cette production a été étudiée et décrite chez le champignon *Phellinus tuberculosus* (White, 1982; Wuosmaa et Hager, 1990), cependant l'exemple le plus parlant est celui de *Phellinus pomaceus* (Figure 1.5) capable de réémettre 90 % des ions chlorure présents dans ses substrats sous forme de chlorométhane (Harper *et al*., 1988). Cette production de chlorométhane se fait via transfert d'un ion Cl^- sur le groupement méthyl de la SAM (Figure 1.5).

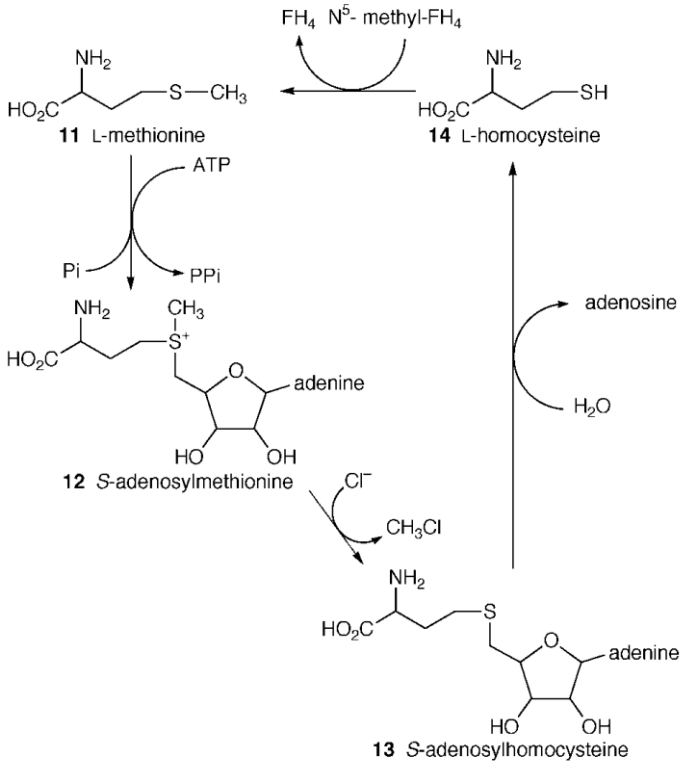


Figure 1.5. Production de chlorométhane par le champignon *Phellinus pomaceus*
(Figure modifiée de Harper, 2000)

Il existe également une production de chlorométhane par les champignons associés aux systèmes racinaires des sols forestiers. Les racines représentent une source importante de COVs (Lin *et al*., 2007). Dans l'étude de Redeker *et al*., 2004, 9 isolats de champignons ectomycorhiziens, dont *Cenococcum geophilum*, l'une des espèces les plus répandues et

communes (Trappe, 1977; Massicotte *et al*., 1992), ont démontré leur capacité à produire des méthanes halogénés dont le chlorométhane. Ils produisent entre 0,003 µg et 65 µg de chlorométhane par gramme de matière sèche (Redeker *et al*., 2004). Cependant, il reste difficile d'évaluer séparément le rôle du système racinaire, car il est étroitement lié au système microbien, lui-même influençant les flux de COVs dans les sols (Peñuelas *et al*., 2014). En effet, plus de 95 % des racines courtes des plantes sont colonisées par des champignons, eux-mêmes localisés à proximité des communautés bactériennes aidant les champignons de la rhizosphère « *mycorrhiza helper bacteria* » notamment lors des processus de développement mycorhyzien (Frey-Klett *et al*., 2007; Bonfante et Anca, 2009; Rigamonte *et al.*, 2010).

– Le chlorométhane est faiblement émis par rapport à d'autres COVs

Les plantes réémettent une partie du carbone disponible sous forme de Composés Organiques Volatiles (COVs) dont le méthanol (Bringel et Couée, 2015; Vorholt, 2012). Ces composés sont émis dans l'atmosphère ou les sols, où ils peuvent ensuite être directement assimilés par les microorganismes (Schade et Goldstein, 2001). Ainsi, la majeure partie du méthanol atmosphérique serait liée à la production par les plantes (Jacob, 2005; Galbally et Kirstine, 2002). D'autres COVs sont produits par les plantes, dont des composés aromatiques mais également des composés halogénés (Guenther *et al*., 2012; Forczek *et al*., 2015). La production par les plantes des 4 méthanes chlorés a été décrites, bien que le chlorométhane soit le plus produit avec une production estimée de 0,8 à 8 $Tg.an^{-1}$ (Clerbaux *et al*., 2007). Cette production reste inférieure à celle de méthanol (50-132 $Tg.\ an^{-1}$) (Jacob, 2005; Forczek *et al*., 2015).

Au niveau de sols forrestiers, le chlorométhane est loin d'être le seul COV produit (Keppler *et al*., 2005; Redeker et Kalin, 2012; Peñuelas *et al*., 2014). Le sol est un milieu hétérogène, constitué de microenvironnements complexes (Peñuelas *et al*., 2014) avec une diversité microbienne importante. La diversité bactérienne et des champignons est évaluée par le nombre d'espèces présentes, soit respectivement de 10^5 et 10^6 espèces par gramme de sol (Peñuelas *et al*., 2014). De nombreux COVs sont émis par ces microorganismes (Figure 1.6). Cette production est directement liée à la disponibilité des substrats et des conditions de croissance (Kai *et al*., 2010; Blom *et al*., 2011). Une grande partie des COVs émis provient de la dégradation microbienne de la litière végétale (matière morte végétale), de la dégradation

de la matière organique (Kögel-Knabner, 2002; Peñuelas *et al*., 2014) ainsi que des exsudats racinaires.

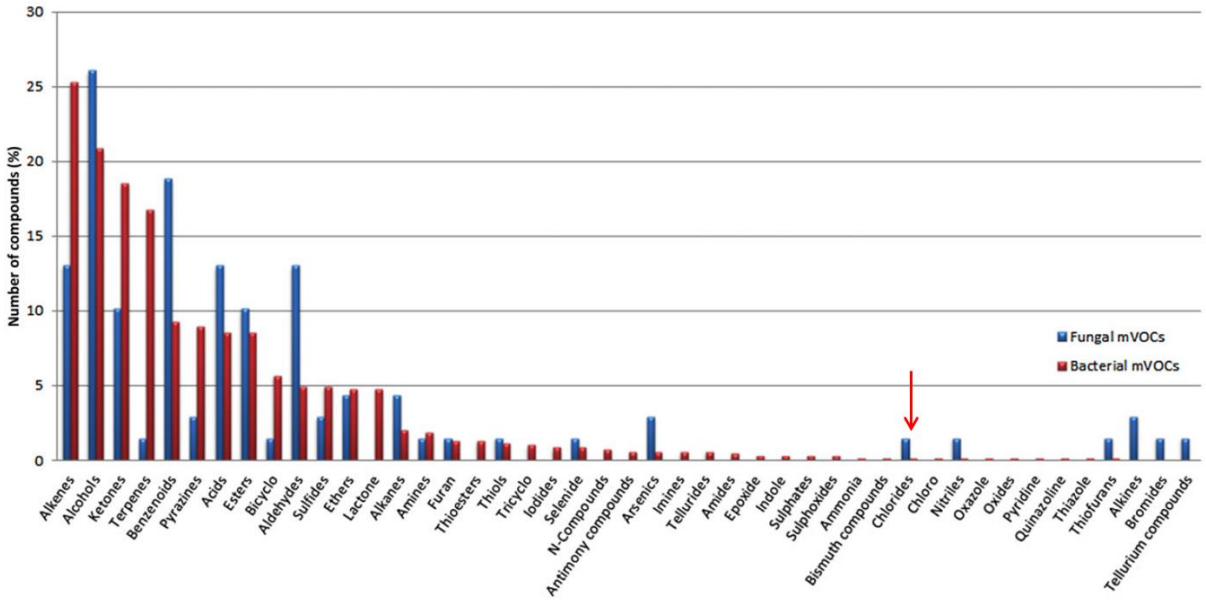


Figure 1.6. Les composés organiques volatiles émis dans les sols
Ceux émis par les bactéries sont représentés en rouge, et ceux émis par les champignons sont représentés en bleu (modifiée de Peñuelas *et al.,* 2014). L'ensemble de ces COVs est répertorié dans une base de données (http://bioinformatics.charite.de/mvoc/). La flèche rouge représente la catégorie de COVs dans laquelle se situe le chlorométhane.

De manière générale, les bactéries produisent plus d'alcènes, d'alcools, de cétones et de terpènes, tandis que les champignons produisent plus d'alcools, de benzénoïdes, d'acides et d'aldéhydes (Figure 1.6). Les composés chlorés sont moins produits, et cette production dans les sols est plus importante par les champignons que par les bactéries (Figure 1.6). Le chlorométhane produit est classé dans le groupe « chlorides » (indiqué par la flèche rouge).

– Dosage du chlorométhane à l'aide d'une souche bio-rapportrice

La quantification de la très faible production de chlorométhane par les plantes et le sol (de l'ordre du ppt) est difficile par des approches classiques de dosage GC « *gas chromatography* » pas assez sensible (Figure 1.7). Au laboratoire à Strasbourg, dans le but d'évaluer les émissions de chlorométhane par les plantes, *Methylobacterium extorquens* CM4 a été transformée pour être capable d'émettre de la fluorescence, proportionnellement à la faible concentration de chlorométhane (Farhan Ul Haque *et al*., 2013). Cette souche bio-rapportrice est basée sur la présence d'un plasmide portant un gène *yfp*, codant une protéine fluorescente sous contrôle du promoteur du gène *cmuA* (Figure 1.7) et permet d'apporter une réponse sensible et proportionnelle à des concentrations de chlorométhane, avec un seuil de détection de 2 pM.

A l'aide de ce bio-rapporteur, la production de chlorométhane chez la plante *A. thaliana* et chez *V. rotundifolia* a pu être démontrée et quantifiée avec des taux d'émission respectifs de 13 et 2 800 ng.g de matière séche^{-1}. h^{-1} (Farhan Ul Haque *et al*., 2013).

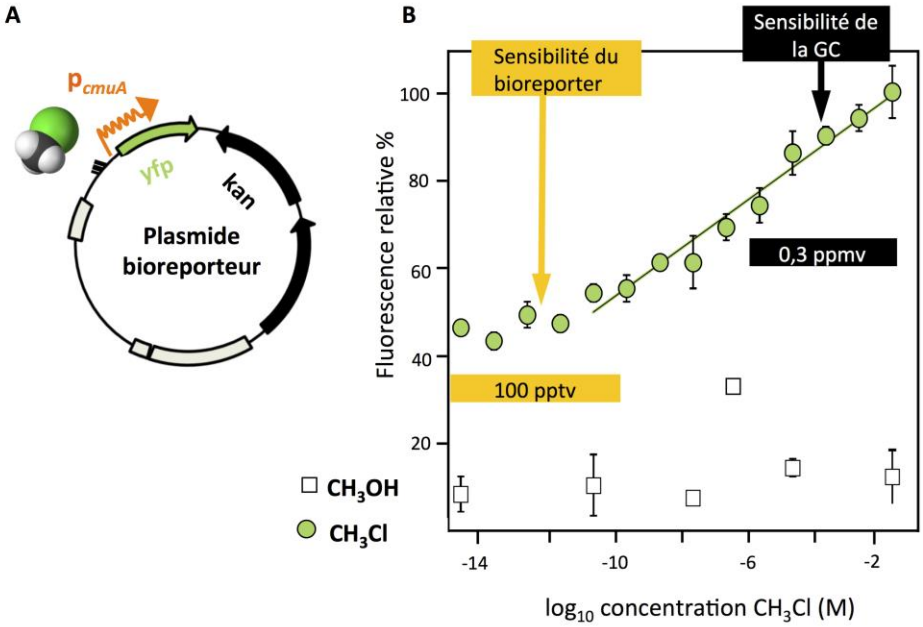


Figure 1.7. Souche bio-rapportrice pour la détection du chlorométhane
(A) Dans la souche bio-rapportrice, un plasmide porteur du gène *yfp* placé sous le contrôle du promoteur du gène *cmuA* (codant une protéine avec un domaine méthyltransférase et un de liaison corrinoïde, esentielle à l'utilisation du chlorométhane chez *M. extorquens* CM4 (Studer *et al*., 2001, 2002)), induit en présence de chlorométhane, est présent dans *M. extorquens* CM4 (Farhan Ul Haque *et al*., 2013). Le plasmide porte un gène de résistance à la kanamycine pour maintenir la présence de ce plasmide lors de la propagation de la souche bio-rapportrice. (B) L'émission de fluorescence de la protéine Yfp est spécifique à la présence de chlorométhane. En jaune, le seuil de détection avec la souche bio-rapportrice et en noir avec le dosage GC.

2.2. Les puits de chlorométhane

2.2.1. Dégradation abiotique du chlorométhane

Différents puits abiotiques de chlorométhane ont été identifiés et contribuent à des degrés plus ou moins importants à l'élimination du chlorométhane (Tableau 1.4). Bien qu'une dégradation abiotique de COVs ait été observée dans les sols, celle du chlorométhane n'a pu y être décrite (Miller *et al.*, 2004; Insam et Seewald, 2010).

Tableau 1.4. Estimation des puits de chlorométhane dans l'atmosphère

Nature des puits	Estimation (Gg. an^{-1})[a]	Valeurs basses et hautes (Gg. an^{-1})[b]

Puits		
Réaction avec les OH de la troposphère	-3 180	-3800/ -4100
Perte dans la stratosphère	-200	-100/ -300
Réaction avec les Cl^- à la surface des océans	-370	-18/ -550
Dégradation microbienne dans le sol	< -1 000	-100/ -1600
Perte dans les océans polaire	-75	-93/-145
Total des puits	< -4 875	-4273/ -6695

[a] Données issues de (Keppler *et al.*, 2005)
[b] Données issues de (Clerbaux *et al*., 2007)

2.2.2. Dégradation biotique du chlorométhane

La dégradation biotique du chlorométhane est principalement liée à l'activité microbienne même si son estimation reste très variable (entre 180 et 1600 $Gg.an^{-1}$) (Keppler *et al*., 2005).

– Dégradation bactérienne dans les environnements marins

Les environnements marins contribuent à la dégradation du chlorométhane (Tableau 1.4). Des bactéries issues d'environnements marins, capables d'utiliser le chlorométhane, ont été isolées. C'est le cas de *Leisingera methylohalidovorans* MB2, premier isolat des *Rhodobacteracea* à démontrer cette capacité (Goodwin *et al*., 1997; Schaefer; 2002, Tableau 1.5). Trois autres souches appartenant à cette famille et dégradant le chlorométhane ont été isolées d'environnements marins; *Roseovarius* sp. 179, *Roseovarius* sp. 217 et *Ruegeria* sp. 198 (Schäfer *et al*., 2005; Tableau 1.5).

Tableau 1.5. Souches bactérienne chlorométhane-dégradantes isolées de différents environnements

Souche bactérienne	Origine	Gram	Métabolisme	Type trophique	Présence de *cmuA*[a]	Références
Acetobacterium dehalogenans MC	Boues activées	positive	anaérobie	Homoacétogène	nd	Traunecker *et al.*, 1991
Aminobacter ciceronei IMB1	Champ de fraises fumigé	négative	aérobie	Méthylotrophe facultatif	oui	Hancock *et al.*, 1998
Aminobacter lissarensis CC495	Sol forestier	négative	aérobie	Méthylotrophe facultatif	oui	Coulter *et al.*, 1999
Hyphomicrobium sp. AT1	Phyllosphère	négative	aérobie	Méthylotrophe facultatif	oui	Nadalig *et al.*, 2011
Hyphomicrobium sp. AT2	Phyllosphère	négative	aérobie	Méthylotrophe facultatif	oui	Nadalig *et al.*, 2011
Hyphomicrobium sp. AT3	Phyllosphère	négative	aérobie	Méthylotrophe facultatif	oui	Nadalig *et al.*, 2011
Hyphomicrobium sp. AT4	Phyllosphère	négative	aérobie	Méthylotrophe facultatif	oui	Nadalig *et al.*, 2011
Hyphomicrobium sp. MC1	Station d'épuration industrielle	négative	aérobie	Méthylotrophe facultatif	oui	Hartmans *et al.*, 1986
Hyphomicrobium sp. MC2	Sol d'usine pétrochimique	négative	aérobie	Méthylotrophe facultatif	oui	Doronina *et al.*, 1996
Leisingera methylohalidovorans MB2	Bassin de marée	négative	aérobie	Méthylotrophe facultatif	non	Schaefer, 2002
Methylomicrobium album BG8	Eau douce	négative	aérobie	Méthylotrophe obligatoire	nd	Han et Semrau, 2000
Methylobacterium extorquens CM4	Sol d'usine pétrochimique	négative	aérobie	Méthylotrophe facultatif	oui	Doronina *et al.*, 1996
Pseudomonas aeruginosa NB1	Boues activées	négative	anaérobie	Méthylotrophe facultatif	nd	Freedman *et al.*, 2004
Roseovarius sp. 179	Eau de mer côtière	négative	aérobie	Méthylotrophe facultatif	oui	Schäfer *et al.*, 2005
Roseovarius sp. 198	Eau de mer côtière	négative	aérobie	Méthylotrophe facultatif	oui	Schäfer *et al.*, 2005
Roseovarius sp. 217	Eau de mer	négative	aérobie	Méthylotrophe facultatif	oui	Schäfer *et al.*, 2005

[a] nd; non determiné

– Dégradation bactérienne dans la phyllosphère

La phyllosphère correspond à la partie aérienne des plantes et constitue un des habitats microbiens les plus importants. La surface totale des faces inférieures et supérieures des feuilles est égale à deux fois la surface terrestre avec une densité bactérienne de 10^7.cm^{-2} (Vorholt, 2012). Peu d'études sont disponibles sur le rôle des microorganismes de la phyllosphère comme puits de chlorométhane. Des bactéries capables de l'utiliser ont été isolées à la surface des feuilles de la plante *A. thaliana* (Nadalig *et al*., 2011), mais le rôle potentiel de filtre de chlorométhane par les bactéries de la phyllosphère n'est pas pris en compte pour le moment.

– Dégradation bactérienne dans les sols

Le sol contient de nombreux composés, dont certains halogènes, et constitue un puits majeur pour certains d'entre eux. Ainsi, 70 % du bromométhane (CH_3Br) est consommé dans les sols (Shorter *et al*., 1995). La dégradation du chlorométhane dans les sols est mal évaluée mais est estimée comme étant supérieure à 1000 Gg. an^{-1} (Tableau 1.4). Cette consommation dans les sols est liée à une dégradation bactérienne mais aussi fongique et serait majoritaire dans l'horizon « O » et est plus faible dans les horizons inférieurs (Redeker et Kalin, 2012).

Les échanges de chlorométhane entre les écosystèmes terrestres et l'atmosphère sont modulés par les bactéries présentes dans les sols forestiers (Miller *et al*., 2004; Borodina *et al*., 2005; Keppler *et al*., 2005; Clerbaux *et al*., 2007; Schäfer *et al*., 2007; Rhew *et al*., 2010). Différentes bactéries appartenant aux alpha et béta-protéobactéries et capables d'utiliser le chlorométhane, ont été isolées de sols forestiers mais leur contribution dans le budget du chlorométhane n'est pas clairement définie (Miller *et al*., 2004; Borodina *et al*., 2005).

Cette consommation de chlorométhane est également liée aux champignons. En effet, ces derniers produisent des métabolites secondaires, dont plus de 200 sont des métabolites halogénés (Gribble, 2003), incluant des métabolites bromés, fluorés, iodés ou chlorés (Anke et Weber, 2006). Ces métabolites chlorés, dont le chlorométhane, peuvent être éliminés lors de réactions de méthylation chez les champignons (Figure 1.8). La méthylation est un processus fréquent de la biosynthèse de métabolites par les champignons (Anke et Weber, 2006).

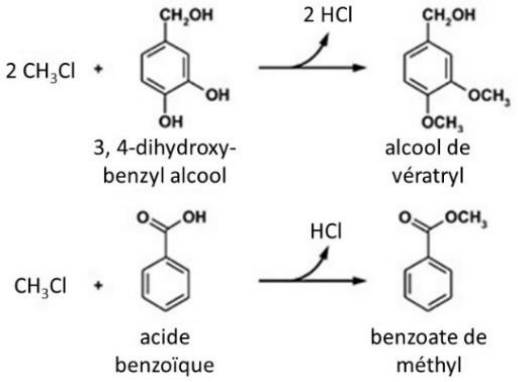


Figure 1.8. Réactions de dégradation du chlorométhane par les champignons
(Modifiée de Anke et Weber, 2006)

Cette méthylation peut se faire chez les champignons via le transfert du groupement méthyl de la SAM comme dans la synthèse du chlorométhane chez les plantes, mais il semble que l'utilisation de chlorométhane soit privilégiée par rapport au SAM, car moins coûteuse d'un point de vue énergétique (Harper, 2000). Plus particulièrement le chlorométhane sert de donneur de groupement méthyle sur des fonctions hydroxy (OH) ou carboxy (COOH), permettant ainsi la synthèse de métabolites, comme l'alcool de vératryl ou le benzoate de méthyl (Figure 1.8). L'alcool de vératryl est un métabolite secondaire intervenant dans la synthèse d'enzymes nécessaires à la dégradation de la lignine (Dekker *et al*., 2001). Le benzoate de méthyl est un composé olfactif synthétisé pour attirer les pollinisateurs ou encore jouant un rôle dans la réduction des racines chez *A. thaliana* (Horiuchi *et al*., 2007). Cependant, le bilan net du chlorométhane chez les champignons reste difficile à définir, le chlorométhane étant à la fois produit et dégradé par les champignons (Anke et Weber, 2006). Bien que l'importance du sol comme puits du chlorométhane ne soit plus à démontrer, de nombreuses incertitudes persistent quant au bilan net de chlorométhane dans les sols (voir paragraphe suivant).

2.3. Le budget global du chlorométhane du sol

Le sol est un environnement complexe rendant difficile une estimation nette du bilan global des puits et sources de chlorométhane. En plus de la grande hétérogénéité intrinsèque du sol (Figure 1.4), le sol est soumis à la variation de nombreux paramètres aussi bien physico-chimiques, climatiques que biologiques. Par exemple, un effet saisonnier sur la production de chlorométhane a été décrit (Peñuelas *et al*., 2014). En effet, une augmentation des émissions de chlorométhane a été observée durant l'été (Llusia *et al*., 2013; Oderbolz *et al*., 2013),

expliquée par le fait qu'en automne l'émission des COVs, dont le chlorométhane, vers l'atmosphère est atténuée par la chute des feuilles restant à la surface du sol (Figure 1.9). La température a également un impact sur la production du chlorométhane. La libération du chlorométhane par du matériel végétal en décomposition est plus importante lorsque la température augmente (Hamilton, 2003).

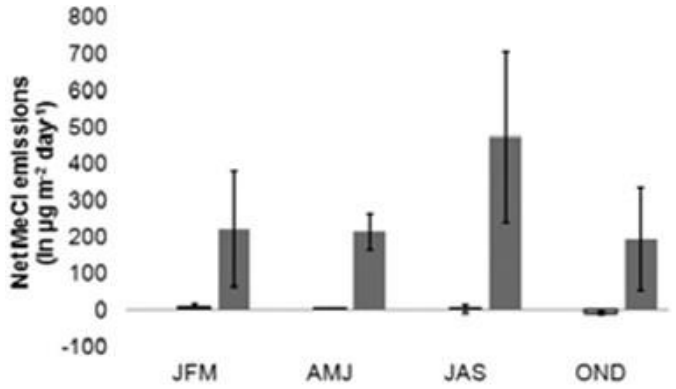


Figure 1.9. Flux du chlorométhane dans un sol forestier en fonction des saisons
Mesure du chlorométhane émis par un sol forestier ($\mu g.m^{-3}.$ $jour^{-1}$), tout au long d'une année. JFM, janvier, février et mars; AMJ, avril, mai et juin, JAS, juillet, août et septembre; OND, octobre, novembre et décembre (issue de Redeker *et al.,* 2012). Les barres d'erreur représentent les erreurs standards.

Ainsi les estimations de la production de chlorométhane peuvent varier de 30 % à 60 % (Saito et Yokouchi, 2008). De plus, les flux de chlorométhane sont bidirectionnels avec une émission depuis le sol vers l'atmosphère et inversement de l'atmosphère vers le sol, avec une importance des microorganismes du sol forestier dans la modulation de ces échanges rendant difficile cette estimation (Schäfer *et al*., 2005). En effet, les estimations du puits du chlorométhane dans le sol varient de 256 $Gg.an^{-1}$ (Clerbaux *et al*., 2007) à plus de 1000 $Gg.an^{-1}$ (Keppler *et al*., 2005) soit 6 % à plus de 20 % du puits troposphérique total du chlorométhane, d'où l'importance de mieux étudier et comprendre le cycle global du chlorométhane dans les sols forestiers à la fois source et puits de chlorométhane.

Les incertitudes sur les quantifications des différents puits et sources de chlorométhane ne permettent pas de fournir un budget global sûr. De plus, la double fonction de certains environnements, à la fois puits et sources, renforce ces incertitudes. Une amélioration des estimations de production et de consommation du chlorométhane passe par le développement de méthodes d'analyse performantes telle que l'isotopie du carbone et de l'hydrogène (Keppler *et al*., 2005; Greule *et al*., 2012; Nadalig *et al*., 2013) et la proposition de

modèle tridimensionnel des flux terrestres de chlorométhane (Xiao *et al.,* 2010). Des études de la dégradation du chlorométhane dans des sols forestiers ont mis en évidence la présence de bactéries actives dans la dégradation du chlorométhane (Borodina *et al*., 2005; Miller *et al*., 2004). Une étude non publiée, réalisée au sein du sol forestier de Steigerwalt en Allemagne (site d'étude pour l'expérience de *Stable Isotope Probing* décrite au chapitre 4) démontre également une dégradation du chlorométhane par des bactéries en condition oxique (Rüffer, 2013).

3. La méthylotrophie

Les microorganismes capables de croître sur le chlorométhane comme source de carbone et d'énergie sont des microorganismes méthylotrophes (Tableau 1.5). La méthylotrophie est définie comme la capacité d'un microorganisme à utiliser un composé sans liaison carbone-carbone comme unique source de carbone et d'énergie. Les substrats de la méthylotrophie regroupent des composés à un seul atome de carbone comme le méthane (CH_4), le méthanol (CH_3OH), le formaldéhyde (HCHO), la méthylamine (CH_3NH_2), ainsi que des composés halogénés comme le chlorométhane et le dichlorométhane, et également des composés à plusieurs atomes de carbone comme la diméthylamine (C_2H_7N) (Chistoserdova, 2011; Muller *et al*., 2011a). Les organismes méthylotrophes sont retrouvés de façon ubiquitaire dans l'air (DeLeon-Rodriguez *et al*., 2013), dans les eaux douces et marines (Neufeld *et al.,* 2008), dans les racines et les feuilles des plantes (Nadalig *et al*., 2011; Knief *et al*., 2012) ou encore dans le sol et les sédiments marins (Kolb, 2009a). Ces organismes sont retrouvés dans les trois domaines du vivant; les bactéries, les archées (dont des archées méthanotrophes, (Antony *et al*., 2012)) et les eucaryotes (dont des levures (Yurimoto, 2009)). Chez les eucaryotes, des levures du genre *Pichia*, *Candida* et *Hansenula* sont capables de métaboliser le méthanol via une alcool oxydase dans un compartiment cellulaire nommé peroxysome (van der Klei *et al*., 2006). Le métabolisme méthylotrophe est toutefois mieux étudié chez les bactéries (Kolb, 2009; Chistoserdova, 2011; Peyraud *et al*., 2011; Peyraud *et al*., 2012). Depuis la découverte des bactéries capables d'utiliser le méthane ou le méthanol comme unique source de carbone et d'énergie (Loew, 1892; käserer, 1906; Söhngen, 1906), de nombreux genres bactériens méthylotrophes ont été identifiés, dont certains sont méthylotrophes facultatifs, capables aussi de croître sur des substrats multicarbonés comme le succinate (Whittenbury *et al*., 1970; Lidstrom, 2006; Boden *et al*., 2008; Hung *et al*., 2011). Les bactéries méthylotrophes

appartiennent aux classes des alpha, béta et gamma protéobactéries, sont positifs ou négatifs à la coloration de Gram, avec des métabolismes aérobies ou anaérobies facultatifs.

Les bactéries méthylotrophes sont classées en 2 groupes; le premier regroupant les bactéries méthanotrophes, capables d'utiliser le méthane comme seule source de carbone et d'énergie. Elles sont souvent méthylotrophes strictes, comme *Methylomonas*, *Methylobacter*, *Methylosinus* et *Methylocystis*. Le 2^{e} groupe comprend celles qui sont incapables d'oxyder le méthane, appelées bactéries méthylotrophes non-méthanotrophes. Ces dernières sont souvent méthylotrophes facultatives, comme les bactéries du genre *Hyphomicrobium* ou *Methylobacterium* (Chistoserdova et Lidstrom, 2013).

3.1. Les voies de la méthylotrophie

La méthylotrophie comprend 3 étapes :

- la transformation initiale des composés en formaldéhyde (HCHO) (pas pour l'utilisation du chlorométhane) ;
- l'oxydation complète de la source de carbone en CO_2 pour la génération d'énergie;
- l'assimilation de carbone pour la production de biomasse qui peut se faire par 3 voies distinctes, la voie du ribulose biphosphate (RuBP) utilisant du CO_2, la voie du ribulose monophosphate (RuMP) à partir de formaldéhyde ou par le cycle de la sérine à partir du méthylène-tetrahydrofolate (CH_2=H_4F) et de CO_2 (Chistoserdova, 2011).

Pour la majorité des souches méthylotrophes, la première étape du métabolisme est l'oxydation ou l'hydrolyse de la source de carbone et d'énergie en formaldéhyde. Les enzymes catalysant cette étape sont spécifiques au substrat : la méthane monooxygénase MMO soluble (sMMO) ou particulaire (pMMO) pour le méthane, la méthylamine déshydrogénase (MADH, périplasmique) pour la méthylamine, la méthanol déshydrogénase (Mxa, périplasmique) pour le méthanol, ou encore une dichlorométhane déshydrogénase pour le dichlorométhane (DcmA, cytoplasmique). Le gène *mxaF* est utilisé comme marqueur moléculaire de l'oxydation du méthanol (Neufeld *et al.,* 2008). Cependant, il existe des systèmes alternatifs pour la dégradation du méthanol, comme les gènes *xox* plus répandus dans l'environnement ou encore *mdh* (Kalyuzhnaya *et al*., 2008). Parmi ces gènes, *xoxF* code une sous-unité d'une méthanol déshydrogénase qui présente plus de 50 % d'identité au niveau des acides aminés avec MxaF (Schmidt *et al*., 2010; Taubert *et al*., 2015). Plusieurs

systèmes d'oxydation du méthanol peuvent être présents simultanément au sein d'un organisme comme par exemple chez *M. extorquens* (Figure 1.10, Skovran *et al*., 2011). Plusieurs systèmes d'oxydation pour l'utilisation de la méthylamine ont également été mis en évidence chez certaines souches de *M. extorquen*s (Vuilleumier *et al*., 2009; Gruffaz *et al*., 2014; Nayak et Marx, 2014). Des composés halogénés sont également utilisés par les bactéries méthylotrophes comme chez *M. extorquens* et *Hyphomicrobium* où deux souches différentes peuvent utiliser soit le dichlorométhane soit le chlorométhane (Gälli et Leisinger, 1985; La Roche et Leisinger, 1990; Vannelli *et al*., 1998; McAnulla *et al*., 2001). La déshalogénation du dichlorométhane par la dichlorométhane déshalogenase codée par le gène *dcmA* conduit à la formation de formaldéhyde.

Chez *Methylobacterium*, deux voies sont actuellement caractérisées pour transformer le formaldéhyde en formiate : l'une est dépendante du tétrahydrofolate (H_4F) et l'autre utilise la tétrahydrométhanoptérine (H_4MPT). La première de ces deux voies n'est pas la voie majeure pour la génération d'énergie, et met en jeu des inter-conversions entre le méthylène-H_4F, le méthényl-H_4F, le formyl-H_4F et enfin le formiate (Marison et Attwood, 1982). Cette voie joue probablement un rôle important dans le maintien de concentrations suffisantes des différents intermédiaires métaboliques contenant du H_4F, nécessaires pour l'assimilation du carbone via le cycle de la sérine (Marx *et al*., 2005; Crowther *et al*., 2008). La seconde voie, découverte plus tardivement en 1998, utilise la H_4MPT qui, condensée au formaldéhyde, est successivement transformée en méthyléne-H_4MPT, méthényl-H_4MPT, formyl-H_4MPT, puis formiate (Chistoserdova *et al*., 1998). Cette voie, initialement décrite chez les *Archaea*, est plus efficace pour l'oxydation du formaldéhyde en CO_2 que la voie dépendante du H_4F grâce aux fortes activités de ses enzymes (Vorholt, 2002). Cependant, tous les substrats de la méthylotrophie ne sont pas oxydés en formaldéhyde; c'est le cas du chlorométhane. La déshalogénation du chlorométhane va mener à la formation de méthylène-tétrahydrofolate (CH_2=H_4F), qui va directement être assimilé dans la biomasse via le cycle de la sérine mais également oxydé en formiate pour fournir de l'énergie (Figure 1.10). Ainsi, *M. extorquens* CM4 convertit directement le chlorométhane en méthylène-tétrahydrofolate (CH_2=H_4F) sans passer par le formaldéhyde grâce aux gènes *cmuA*, *cmuB*, *metF2* codant respectivement une protéine avec un domaine méthyltransférase et un de liaison corrinoïde, une méthyltransférase méthylcobalamine : H_4F et une méthylène-H_4F réductase (Studer *et al*., 1999 ; Studer *et al*., 2001; Studer *et al*., 2002) (voir section 1.3.4.2.).

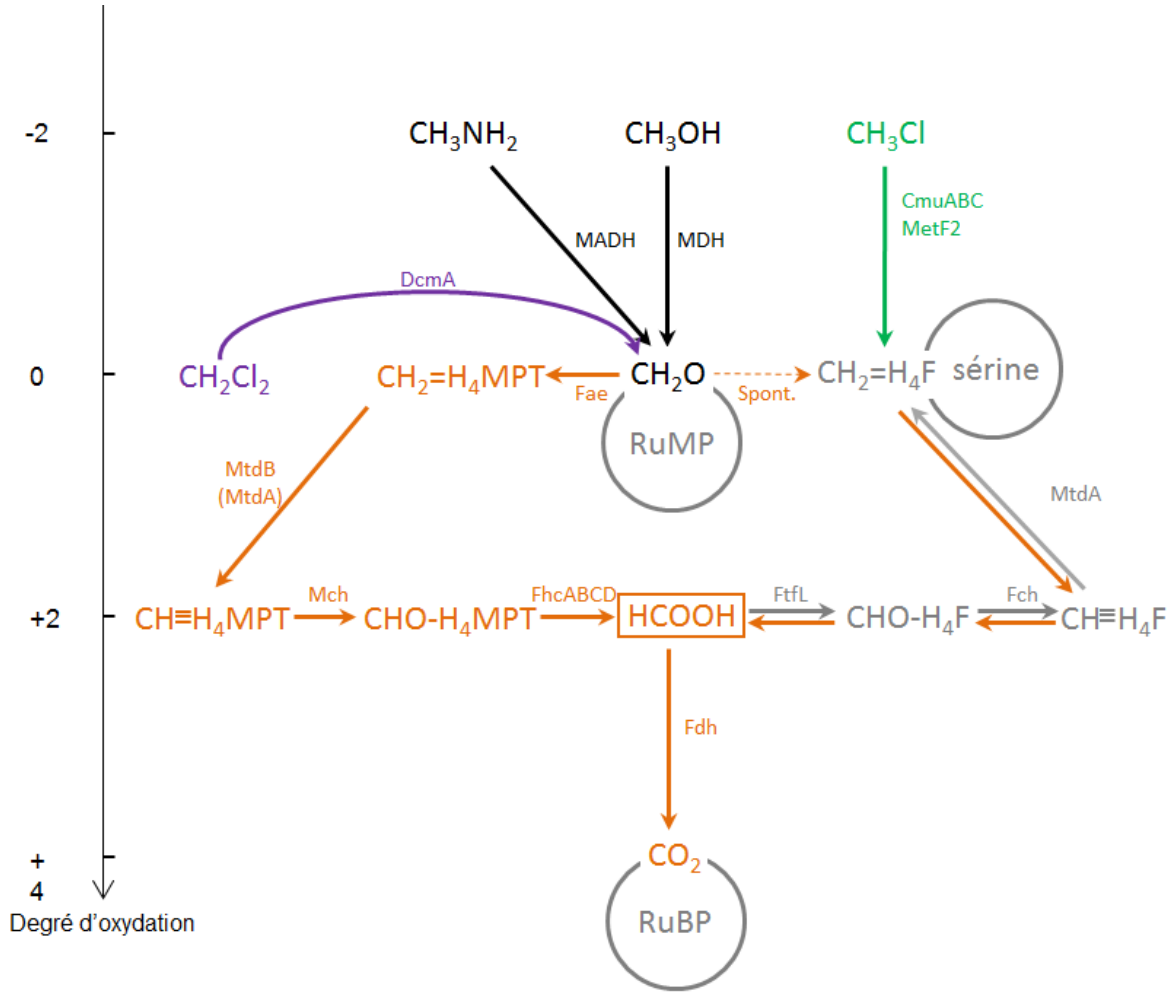


Figure 1.10. Schéma synthétique des différentes voies de la méthylotrophie chez *M. extorquens*

Le degré d'oxydation des molécules est représenté par l'échelle de gauche. Les réactions spécifiques liées à l'utilisation du chlorométhane (CH_3Cl) et du dichlorométhane (CH_2Cl_2) comme unique source de carbone et d'énergie sont représentées respectivement en vert et en violet : CmuABC, « *chloromethane utilizing genes* », MetF2, méthylène- H_4F réductase, DcmA, dichlorométhane déshalogénase. Les autres réactions spécifiques aux substrats utilisés sont en noir : MDH, méthanol déshydrogénase; MADH, méthylamine déshydrogénase. Les réactions conduisant à l'oxydation complète du carbone pour la production d'énergie sont en orange : Fae, enzyme d'activation du formaldéhyde; MtdAB, méthylène-tétrahydrométhanoptérine déhydrogénase; Mch, méthenyl-tétrahydrométhanoptérine cyclohydrolase; Fch, méthylène-tétrahydrofolate cyclohydrolase; FtfL, formyl-tétrahydrofolate ligase; FhcABCD, complexe formyltransférase-hydrolase; Fdh, formiate déshydrogénase. Les réactions de l'assimilation du carbone pour produire de la biomasse sont en gris : FtfL, formyl-tétrahydrofolate ligase; Fch, méthylène-tétrahydrofolate cyclohydrolase; MtdA, méthylène-tétrahydrométhanoptérine déhydrogénase; RuMP, cycle du ribulose monophosphate; RuBP, cycle du ribulose biphosphate.

Plus de 86 gènes impliqués dans le métabolisme des composés en C_1 chez *M. extorquens* ont été identifiés (Marx *et al*., 2003). Ces gènes sont impliqués dans l'oxydation du méthanol, de

la méthylamine, du formaldéhyde, du formiate ou encore dans le cycle de la sérine (Figure 1.11).

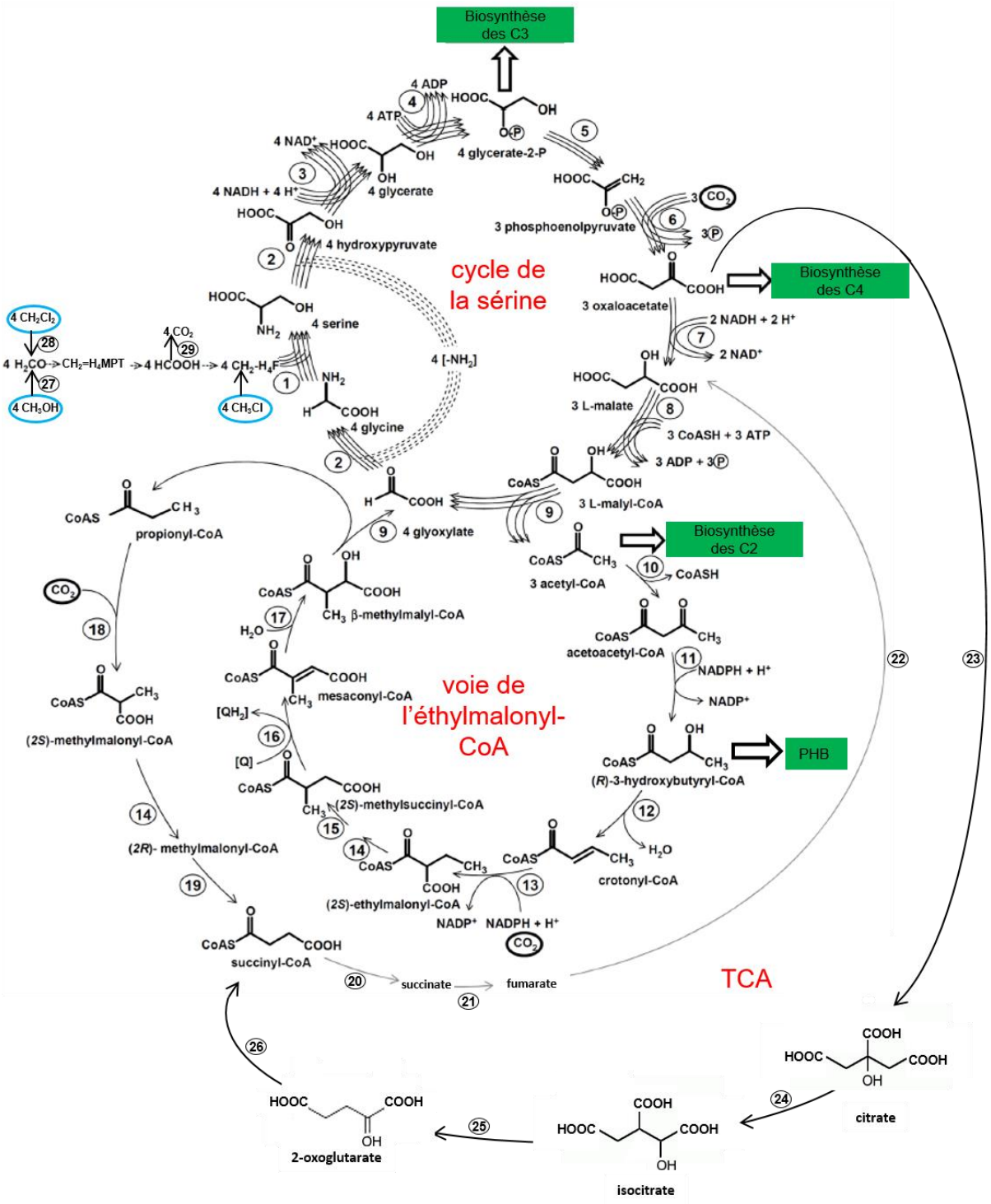


Figure 1.11. Assimilation du carbone chez *M. extorquens*

Les molécules d'entrée sont encadrées en bleu. Les principales connexions vers d'autres voies métaboliques sont indiquées par des rectangles verts. Les bactéries transforment le méthanol (CH_3OH) ou le dichlorométhane (CH_2Cl_2) en formaldéhyde (HCHO) qui est condensé avec une molécule de tétrahydrométhanoptérine (H_4MPT), puis oxydé en formiate (HCOOH). Le formiate peut être transformé en CO_2, ou assimilé via le méthylène-tétrahydrofolate (H_4F) par le cycle de la sérine. L'assimilation du chlorométhane (CH_3Cl), se fait directement par transfert

du groupement méthyl sur le H_4F, qui est alors assimilé via le cycle de la sérine pour fournir de la biomasse, ou oxydé en formiate (HCOOH). L'assimilation de l'acétyl-CoA et la régénération du glyoxylate se font au travers de la voie de l'éthylmalonyl CoA. Le cycle TCA (cycle des acides tricarboxyliques) est quant à lui principalement utilisé pour l'assimilation des composés multicarbonés (Šmejkalová *et al*., 2010). Le nombre de flèches correspond au nombre de molécules impliquées dans la réaction. 1, sérine hydroxyméthyl transférase; 2, sérine-glyoxylate aminotransférase; 3, hydroxypyruvate réductase; 4, glycérate kinase; 5, énolase; 6, phosphoénolpyruvate carboxylase; 7, malate déshydrogénase; 8, malate-CoA ligase (malate thiokinase); 9, L-malyl-CoA/β-méthylmalyl-CoA lyase; 10, β-cétothiolase; 11, acétoacétyl-CoA réductase; 12, crotonase (R-spécifique); 13, crotonyl-CoA carboxylase réductase; 14, éthylmalonyl-CoA/méthylmalonyl-CoA épimerase; 15, éthylmalonyl-CoA mutase; 16, méthylsuccinyl-CoA déshydrogénase; 17, mésaconyl-CoA hydratase; 18, propionyl-CoA carboxylase; 19, méthylmalonyl-CoA mutase; 20, succinyl-CoA synthase; 21, succinate déshydrogénase; 22, fumarase; 23, citryl-CoA lyase; 24, aconitate hydratase; 25, NADP-dépendent isocitrate déshydrogenase; 26, complexe 2-oxoglutarate déshydrogénase; 27, méthanol déshydrogénase; 28, dichlorométhane déshalogénase; 29, formiate déshydrogenase. Pour une même réaction le nombre de flèche correspond au nombre de molécules impliquées dans cette réaction. PHB, polyhydroxybutyrate; Q, quinone (figure modifiée d'après Šmejkalová *et al*., 2010).

Chez *Methylobacterium*, une partie du formaldéhyde produit intègre le cycle de la sérine via le formiate et le méthylène-H_4F pour la production de biomasse. L'enzyme GlyA (sérine hydroxyméthyltransférase) permet la condensation du méthylène-H_4F avec la glycine pour produire la sérine (Figure 1.11). La sérine (composé à trois atomes de carbone) est ensuite transformée en hydroxypyruvate, glycérate, glycérate-2-phosphate et en phosphoénolpyruvate. Le phosphoénolpyruvate est converti en oxaloacétate puis en malate, composés à quatre atomes de carbone par incorporation de CO_2. Le malate, couplé au co-enzymeA (CoA), est scindé en deux composés, le glyoxylate permettant de refermer le cycle de la sérine et l'acétyl-CoA qui est intégré dans la voie de l'éthylmalonyl-CoA où une molécule de CO_2 est incorporée (Erb *et al*., 2008; Peyraud *et al*., 2009; Šmejkalová *et al*., 2010). Ce cycle peut également produire de l'hydroxybutyryl-CoA, matière première de la synthèse de polyhydroxybutyrate (PHB), formant des granules de réserve en carbone et phosphate pour la cellule.

3.2. La dégradation bactérienne du dichlorométhane (CH_2Cl_2)

L'utilisation bactérienne du dichlorométhane est mieux connue en condition oxique qu'anoxique.

3.2.1. Utilisation du dichlorométhane en condition anoxique

Un processus de fermentation du dichlorométhane a été décrit chez la bactérie *Dehalobacterium formicoaceticum* (Mägli *et al*., 1998). Le métabolisme anaérobie de cette bactérie fait intervenir une combinaison de méthyltransférases dépendantes de la cobalamine et du H_4F. Le dichlorométhane et le H_4F sont convertis en méthylène-tétrahydrofolate (CH_2=H_4F), et en chlorure inorganique. Une partie du CH_2=H_4F est transformée en formiate, tandis qu'une autre partie du CH_2=H_4F forme de l'acétate par incorporation de CO_2 présent dans le milieu (Mägli *et al*., 1998).

3.2.2. Utilisation du dichlorométhane en condition oxique

La première souche capable d'utiliser le dichlorométhane comme substrat de croissance à avoir été isolée et caractérisé est une bactérie du genre *Methylopila* (Brunner *et al*., 1980). Trois voies de dégradation du dichlorométhane en aérobie sont connues. Cependant, une seule de ces voies permet l'utilisation du dichlorométhane comme source de carbone et d'énergie. L'utilisation du dichlorométhane comme source de carbone et d'énergie est possible par l'expression d'une dichlorométhane déshalogénase appartenant à la famille des glutathion-S-transférase (GST), codée par le gène *dcmA* (Figure 1.12). Dans la majorité des cas, les souches capables d'utiliser le dichlorométhane comme substrat de croissance possèdent le gène *dcmA* (Vuilleumier *et al*., 2001; Muller *et al*., 2011b). De plus, ce gène est fortement conservé avec des identités au niveau nucléotidique comprises entre 98 % et 100 %. La protéine DcmA catalyse la réaction de déchloration d'une mole de dichlorométhane en deux moles de protons (H^+), deux moles de chlorure (Cl^-), et une mole de formiate, via la formation consécutive de S-chlorométhylglutathion et de formaldéhyde, deux composés génotoxiques (Figure 1.12).

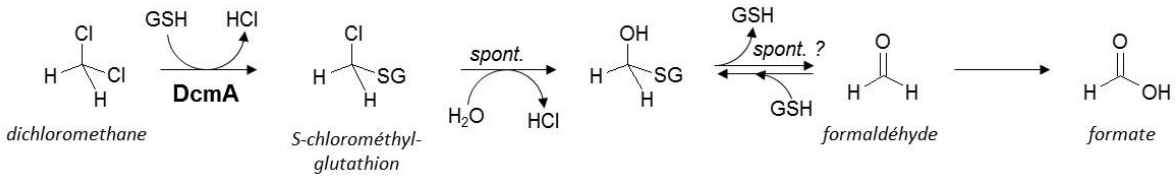


Figure 1.12. Dégradation du dichlorométhane par la voie hydrolytique de la glutathion S-transférase
(Figure adaptée de Muller *et al*., 2011). spont.: spontanée, réaction abiotique; GSH, glutathion; SG, glutathionyl.

La dégradation bactérienne aérobie du dichlorométhane a été largement décrite chez *M. extorquens* DM4 qui utilise le dichlorométhane comme unique source de carbone et d'énergie (Roche et Leisinger, 1991; Kayser *et al*., 2002; Muller *et al*., 2011a). Les gènes d'utilisation du dichlorométhane sont portés par un transposon catabolique possédant quatre gènes *dcm* encadrés par des séquences d'insertion. Ces gènes sont la dichlorométhane déshalogénase (*dcmA*), *dcmR* qui code un facteur de transcription, ainsi que deux gènes de fonction encore inconnue (*dcmB* et *dcmC*). Ce transposon est lui-même localisé sur un îlot génomique d'une taille de 126 kb (Vuilleumier *et al*., 2009). L'utilisation du dichlorométhane par *M. extorquens* DM4 aboutit à la formation de formaldéhyde, qui est ensuite oxydé en formiate via la voie de la tétrahydrométhanoptérine (H_4MPT). Le formiate est soit réduit en CH_2=H_4F pour fournir de la biomasse via le cycle de la sérine, soit oxydé en CO_2 par une formiate déshydrogénase (FDH) pour produire de l'énergie (Figure 1.10).

3.3. La dégradation bactérienne du chlorométhane

Plusieurs genres bactériens, isolés de différents environnements, sont capables de dégrader le chlorométhane (Tableau 1.5). Le chlorométhane peut être utilisé comme source de carbone et d'énergie ou dégradé sans servir de source de carbone en conditions oxique ou anoxique (Traunecker *et al*., 1991; Han et Semrau, 2000; Freedman *et al*., 2004).

3.3.1. Dégradation du chlorométhane en condition anoxique

La déshalogénation du chlorométhane en anaérobie est catalysée par une enzyme qui transfère le groupement méthyl (CH_3) au H_4F via une protéine corrinoïde, produisant des ions chlorure et du méthyl-H_4F, qui est ensuite oxydé en formiate puis en CO_2 ou assimilé par la voie de l'acétyl CoA (Harper, 2000). La bactérie *Acetobacterium dehalogenans* MC utilise le chlorométhane en anaérobie (Tableau 1.5) (Traunecker *et al*., 1991; Harper, 2000). Cette

déshalogénation est basée sur une méthyl transférase corrinoïde-dépendante qui contrairement à d'autres méthyl transférase corrinoïde-dépendantes, nécessite le maintien du cobalt dans la protéine corrinoïde par un système ATP-dépendant (Studer *et al*., 1999). La souche *Pseudomonas aeruginosa* NB1, quant à elle, est capable d'utiliser le chlorométhane comme unique source de carbone et d'énergie en condition de réduction des nitrates (Freedman *et al*., 2004). Cependant, le mécanisme de déchloration reste inconnu.

3.3.2. Dégradation du chlorométhane en condition oxique

La première souche décrite comme utilisant le chlorométhane comme unique source de carbone et d'énergie est *Hyphomicrobium* sp. MC1 (Hartmans *et al*., 1986). Comme cette souche, la majorité des souches chlorométhanes-dégradantes isolées à ce jour sont des microorganismes se développant en condition oxique. Les souches ont été isolées d'environnements variés comme des sols d'usines pétrochimiques, des stations d'épuration, des sols forestiers ou des milieux aquatiques marins et dulçaquicoles (Tableau 1.5). Elles appartiennent à la classe des *Alphaproteobacteriacea* et aux genres *Methylobacterium*, *Hyphomicrobium* et *Aminobacter*. A ce jour, la dégradation du chlorométhane la mieux caractérisée est celle de *M. extorquens* CM4 selon la voie *cmu*, mais elle n'est pas la seule voie de dégradation (voir paragraphes ci-dessous).

– Voie de dégradation du chlorométhane chez *Methylomicrobium album* BG8

A part la voie *cmu*, le chlorométhane peut aussi être dégradé en condition oxique par une bactérie méthanotrophe, *Methylomicrobium album* BG8, mais seulement en présence d'une autre source carbonée telle que le méthanol (Figure 1.13) (Han et Semrau, 2000).

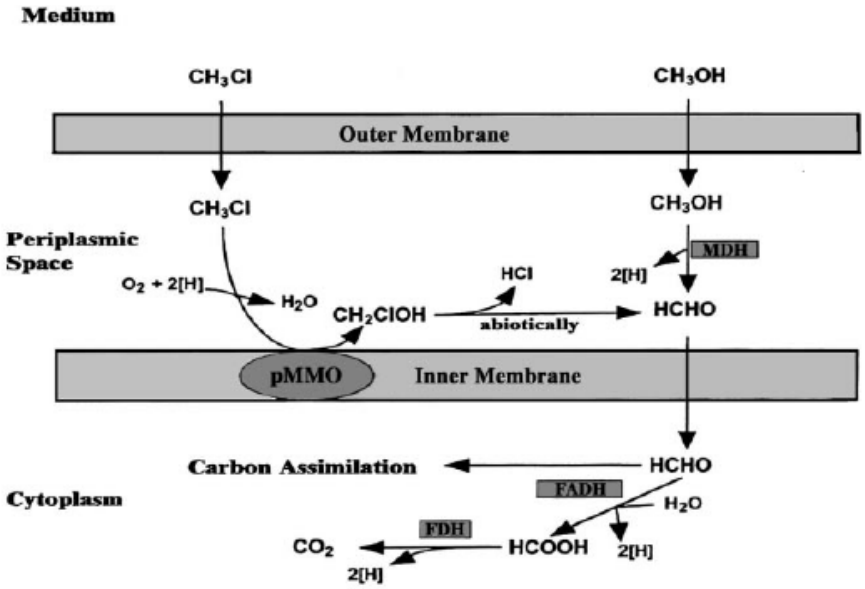


Figure 1.13. Mecanisme proposé d'oxydation du chlorométhane par *Methylomicrobium album* BG8
En présence d'acéthylène, inhibiteur de l'activité de la pMMO, la souche ne dégrade plus le chlorométhane, démontrant que le chlorométhane est oxydé par la méthane monooxygénase particulaire (pMMO). Le méthanol est oxydé en formadéhyde (HCHO) par la méthanol deshydrogénase (MDH). Le formaldéhyde rejoint le métabolisme en C_1 pour la synthèse de biomasse et pour la production d'énergie sous forme d'équivalents réducteurs (2[H]) en impliquant la formaldéhyde déshydrogénase (FADH) et la formiate déshydrogénase (FDH). On notera que par rapport au méthanol, le chlorométhane génère deux équivalents réducteurs en moins. (D'après Han et Semrau, 2000).

Dans une culture de *M. album* BG8 en présence de méthanol, le marquage au C^{14} du CH_3Cl a permis de démontrer que le chlorométhane stimule la croissance et l'assimilation du carbone dans la biomasse jusqu'à 38 % avec 50 % du chlorométhane transformé en CO_2 suggérant que le chlorométhane peut servir comme source de carbone et d'énergie (Han and Semrau, 2000). Cependant, comme le chlorométhane ne peut pas servir comme unique source de carbone, les auteurs de ce travail proposent que *M. album* BG8 ne puisse pas tirer assez d'équivalents réduits de la dégradation du chlorométhane pour permettre sa croissance sur le chlorométhane. En présence d'une autre source carbonée en quantité suffisante par rapport au chlorométhane, l'utilisation du chlorométhane est possible (2,6 mM de chlorométhane et 5 mM de méthanol) (Han et Semrau, 2000).

– La voie *cmu*

Des études biochimiques et génétiques chez la souche *M extorquens* CM4 ont permis d'identifier une voie d'utilisation du chlorométhane en condition oxique, appelée voie *cmu*

(chlorométhane utilisation) (Vannelli *et al*., 1998, 1999) (Figure 1.14). *M extorquens* CM4 métabolise 1 mole de chlorométhane et 1,5 moles d'O_2 pour produire 1 mole de CO_2 et 1 mole de chlorure (Vannelli *et al*., 1998). L'inactivation de gènes par mutagénèse aléatoire avec le minitransposon *Tn*5 a permis d'identifier 9 mutants dont la croissance était maintenue en méthanol ou en méthylamine, mais impossible sur chlorométhane (Vannelli *et al*., 1998). L'étude de ces mutants a mis en évidence le caractère essentiel des gènes *cmuABC,* des méthyltransférases et *purU* et *metF2*, codant respectivement une méthylène-H_4F réductase et une formyl- H_4F, pour la croissance en présence de chlorométhane (Studer *et al*., 2001, 2002). De plus, les gènes *cmuAB* sont essentiels pour la déchloration du chlorométhane (Studer *et al*., 2001b, 2002).

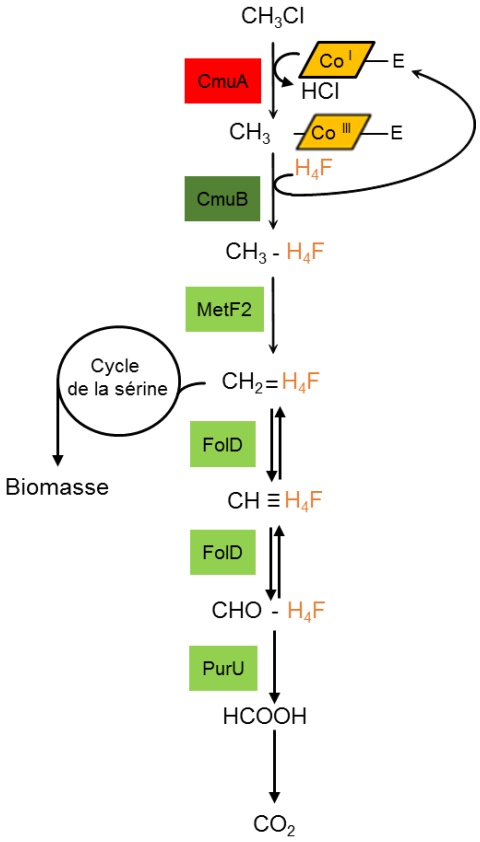


Figure 1.14 Voie *cmu* chez *Methylobacterium extorquens* CM4
Le groupement méthyl du chlorométhane est transféré au tétrahydrofolate (H_4F) par *cmuA* et *cmuB*. Le méthényl-tétrahydrofolate produit par oxydation du méthyle-tétrahydrofolate peut soit incorporer le cycle de la sérine pour produire de la biomasse, ou être finalement oxydé en CO_2 pour produire de l'énergie. CmuA, protéine avec un domaine méthyltransférase et un de liaison de groupe corrinoïde (Studer *et al.,* 2001), CmuB, méthyltransférase méthylcobalamine:H_4F (Studer *et al.,* 1999); MetF2, méthylène réductase; FolD, enzyme bifonctionnelle méthyléne-H_4F déshydrogénase/ méthényl-H_4F cyclohydrolase; PurU, 10-formyl-H_4F hydrolase.

Les protéines CmuA et CmuB vont catalyser le transfert du groupement méthyl du chlorométhane au H_4F pour former le méthyl-H_4F (CH_3-H_4F) (Studer *et al*., 2001). Le CH_3-H_4F est ensuite oxydé en formiate et CO_2, avec une assimilation du carbone au niveau du méthylène-H_4F (Vannelli *et al*., 1999). Chez *M. extorquens* CM4, l'ensemble des gènes de la voie *cmu* sont localisés sur le plasmide pCMU01 dans une région de 180 kb (Figure 1.15). Les gènes sont regroupés en 2 clusters distants de 30 kb, contenant un taux de GC atypique et dépourvus d'éléments mobiles. Directement en aval de *purU* et *folD* se trouve le gène *folC2*, codant une enzyme bifonctionnelle folylpolyglutamate synthase et dihydrofolate synthase. Cette région comporte aussi des gènes non caractérisés fonctionnellement : *paaE-like*, *fmdB* et *hutI*, codant respectivement une putative oxydoréductase, un régulateur transcriptionnel putatif, et une imidazole hydrolase putative.

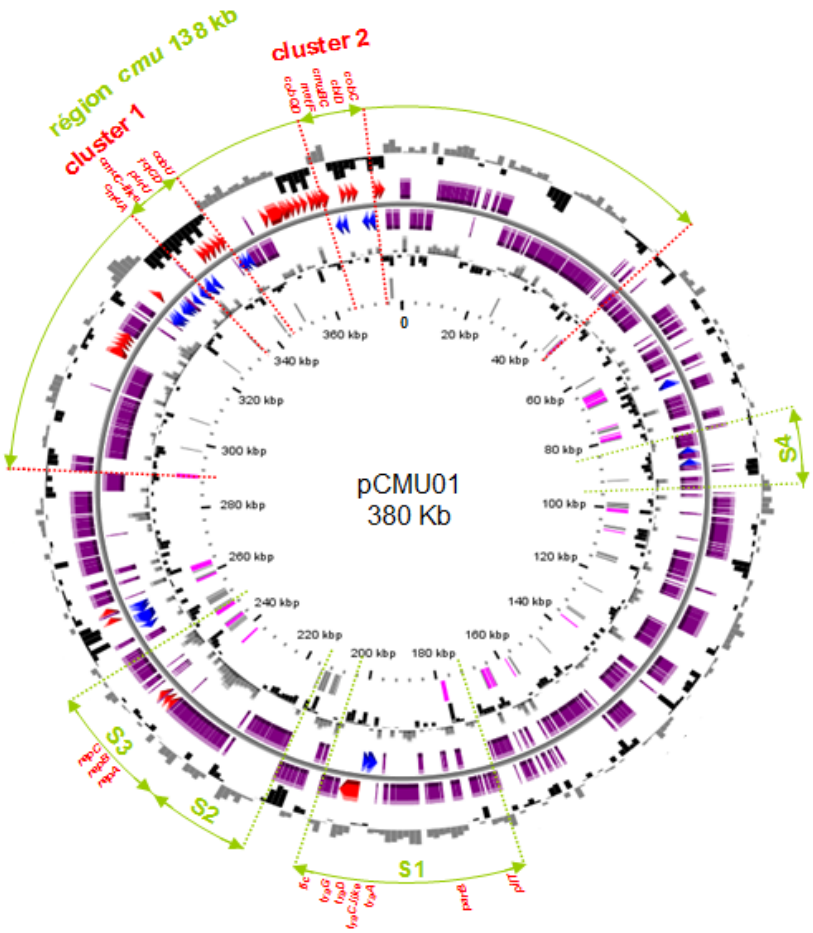


Figure 1.15. Plasmide pCMU01 d'utilisation du chlorométhane chez *M. extorquens* CM4
Les clusters 1 et 2 distants de 21 kb contiennent tous les gènes essentiels d'utilisation du chlorométhane (*cmu*) connues avant le séquençage du génome (Vannelli *et al*., 1999; Studer *et al*., 2002). L'ensemble de cette région couvre 40 kb de la région *cmu* du plasmide de 380 kb. Les zones S1, S2, S3 correspondent à une région de 69 kb conservée entre le plasmide pCMU01

et le plasmide p1METDI de *M*. *extorquens* DM4. La région S4 présente une synténie avec une région dupliquée du chromosome de la souche dichlorométhane-dégradante DM4. Les cercles représentent depuis l'extérieur : 1, pourcentage de déviation GC (GC fenêtre – GC moyen) dans une fenêtre de 1000 pb; 2, CDS prédits transcrits dans le sens des aiguilles d'une montre; 3, CDS prédits transcrits dans le sens contraire des aiguilles d'une montre; 4, GC skew (G+C/G-C) dans une fenêtre de 1000 pb; 5, éléments transposables (rose) et pseudogènes (gris) (Figure issue du manuscrit de thèse de Sandro Roselli, 2009)

Cette région contient également 23 gènes impliqués dans la biosynthèse et le transport de la cobalamine, ainsi que ceux nécessaires à la biosynthèse du tétrahydrofolate, deux cofacteurs essentiels à la voie *cmu* d'utilisation du chlorométhane (Figure 1.15). Une partie d'entre eux présente une copie chromosomique (Roselli *et al*., 2013). Toutefois, l'analyse des génomes de bactéries possédant les gènes *cmu* a fait apparaître deux organisations génétiques : une retrouvée chez *M. extorquens* CM4, où les gènes sont distribués en deux clusters éloignés de 30 kb, et une deuxième organisation, présente chez les autres bactéries, où les gènes *cmu* sont situés sur le même cluster (Figure 1.15). Un degré important de conservation dans l'organisation des gènes *cmu* chez les bactéries chlorométhane-dégradantes a été observé (Figure 1.15, (Schäfer *et al*., 2007)). Ces bactéries appartiennent toute à la classe des *Alphaproteobacteria*. L'ordre des gènes dans le cluster *cmu* est fortement conservé (Nadalig *et al*., 2011). Cependant chez *M. extorquens* CM4, les gènes *cmuB* et *cmuC* appartiennent à la même unité transcriptionnelle, contrairement à *cmuA* sous contrôle d'un promoteur propre (Studer *et al*., 2002).

Des souches chlorométhane-dégradantes ne possédant pas les gènes *cmu* ont été isolées comme *Roseovarius* sp. 217 et *Leisingera methylohalidovorans* MB2 (Schäfer *et al*., 2007). Des analyses en isotopes stables du chlorométhane ont montré que les fractionnements isotopiques du carbone et de l'hydrogène étaient différents pour les bactéries chlorométhane-dégradantes possédant les gènes *cmu* et *L. methylohalidovorans* MB2 ne les possédant pas (Nadalig *et al*., 2014). La voie de dégradation du chlorométhane par *L. methylohalidovorans* MB2, indépendante des gènes *cmu*, reste à élucider.

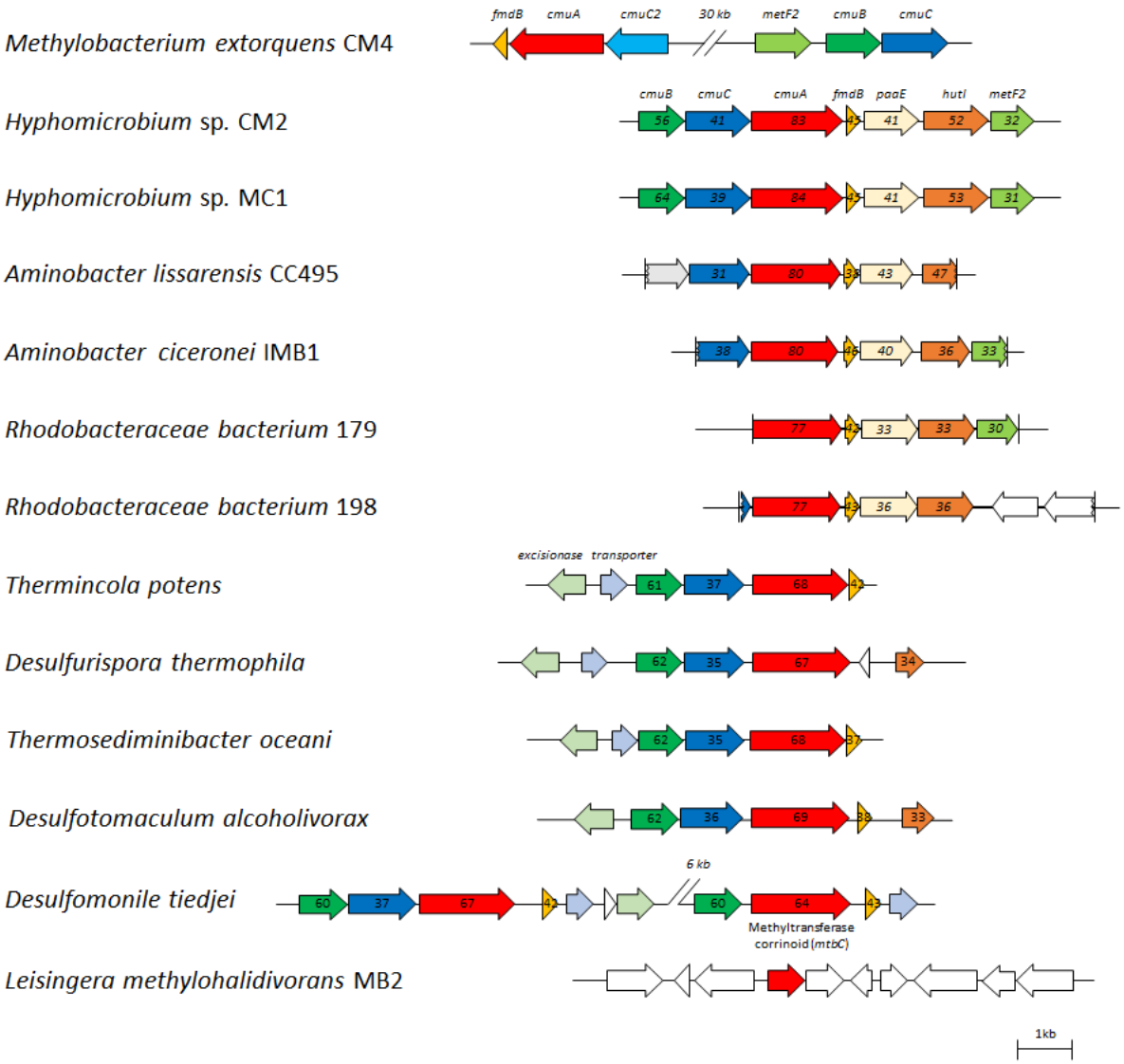


Figure 1.16. Comparaison de l'organisation des gènes *cmu* chez des bactéries contenant le gène *cmuA*

Les flèches représentent les gènes codant des protéines. Les flèches incomplètes représentent la présence de gène dont la séquence n'est pas entièrement déterminée. La couleur de référence des gènes est celle indiquée chez *M. extorquens* CM4, où ces gènes sont retrouvés sur deux clusters séparés de 30 kb. Les gènes homologues sont représentés de la même couleur; Le pourcentage d'homologie avec les protéines codées par les gènes de *M. extorquens* CM4 est indiqué à l'intérieur des flèches. L'organisation génétique est représentée à l'échelle. Chez la souche MB2, on notera que seulement une partie du gène *cmuA* correspondant au domaine de liaison du co-facteur corrinoide est présente.

4. Le modèle *Methylobacterium*

Le genre *Methylobacterium* a été décrit pour la première fois par Patt, Cole et Hanson en 1976 (Patt *et al.*, 1976). Ce genre compte une vingtaine d'espèces appartenant, selon la séquence nucléotidique du gène codant l'ARNr 16S, aux Alpha-protéobactéries, à l'ordre des *Rhizobiales* et à la famille des *Methylobacteriacaea* (Lidstrom, 2006). Elles sont méthylotrophes

facultatives, aérobies strictes, mésophiles avec un optimal de température compris entre 25 et 30°C, de coloration Gram négative et elles montrent une réaction positive aux tests de catalase et d'oxydase. Les cellules possèdent souvent des inclusions de poly-β-hydroxybutyrate, réserves d'énergie et de carbone, et ont en général une coloration rose en raison de la présence de caroténoïdes (Korotkova et Lidstrom, 2001; Van Dien *et al*., 2003). Les souches de *Methylobacterium* ont été isolées de lieux très divers, anthropisés (bords de route, réservoirs d'eau) mais aussi naturels (sol, surface de feuilles) (Tableau 1.6)
(Radajewski *et al*., 2002; Nadalig *et al*., 2011; Bai *et al*., 2015). Ces bactéries ont un rôle environnemental important de par leur capacité à métaboliser des composés toxiques comme le méthanol, le dichlorométhane ou encore le chlorométhane (Dourado *et al*., 2015). Les bactéries méthylotrophes ont été parmi les premières étudiées dans l'environnement à l'aide d'outils moléculaires (Holmes *et al*., 1995; Kolb et Stacheter, 2013).

4.1. *Methylobacterium*, bactérie modèle pour l'étude de la méthylotrophie

Une des espèces les plus étudiées du genre *Methylobacterium* est *M. extorquens* retrouvée dans de nombreux environnements comme les plantes, le sol, les eaux usées ou les nuages (Bai *et al*., 2015; Bringel et Couée, 2015; Nadalig *et al*., 2011; Ochsner *et al*., 2015; Temkiv *et al*., 2012). *M. extorquens* AM1 est la souche la plus étudiée et a été considérée, avec *Methylobacterium organophilum*, comme le modèle d'étude du genre *Methylobacterium*, pour la méthylotrophie ou les intéractions plantes-microorganismes (Dourado *et al*., 2015). La présence de cinq réplicons et de 174 séquences d'insertion chez *M. extorquens* AM1 complique son utilisation en biologie moléculaire et elle est en voie d'être remplacée par *M. extorquens* PA1 qui ne comporte qu'un chromosome et seulement 20 séquences d'insertion (Nayak et Marx, 2014). L'ensemble des souches de *M. extorquens* (AM1, BJ001, CM4, DM4, DSM 13060 et PA1) sont capables d'utiliser le méthanol, mais seules *M. extorquens* CM4 et DM4 sont capables d'utiliser des substrats chlorés comme unique source de carbone et d'énergie (Vuilleumier *et al*., 2009; Marx *et al*., 2012). *M. extorquens* CM4, isolée d'un sol d'usine pétrochimique en Russie, est capable de croître avec du chlorométhane (Doronina *et al*., 1996), tandis que *M. extorquens* DM4, capable de croître sur dichlorométhane, a été isolée d'un sol pollué par des alcanes chlorés (Gälli et Leisinger, 1985). Les génomes des souches de

M. extorquens sont caractérisés par un taux de GC supérieur à 68 %, avec un chromosome de taille importante (entre 5,5 et 5,9 Mb) (Tableau 1.6).

A l'exception de la souche PA1 et DSM 13060, les souches de *M*. *extorquens* possèdent un ou plusieurs plasmides, avec pour la souche *M. extorquens* AM1 la présence d'un mégaplasmide (Ochsner *et al*., 2015). Fin décembre 2015, 24 génomes de *Methylobacterium* isolés de la phyllosphère et de racines de plantes ont été séquencées (Bai *et al*., 2015), mais aucun gène similaire à *cmuA* n'a été détecté.

Tableau 1.6. Caractéristiques des premières souches de *Methylobacterium* dont le génome a été séquencé

Paramètres[a]	*M. extorquens* AM1	*M. extorquens* BJ001[b]	*M. extorquens* CM4	*M. extorquens* DM4	*M. extorquens* DSM13060	*M.* extorquens PA1	*M. nodulans* ORS 2060	*M. radiotolerans* JCM 2831	*M.* sp. 4-46
Génome									
Taille (Mb)	6,88	5,85	6,18	6,12	6,73	5,47	8,84	6,90	7,73
Contenu en G+C (%)	68,5	69,4	68,0	68,0	68,23	68,2	68,4	71,0	71,5
Nombre de réplicons	5	3	3	3	1	1	8	9	3
Nombre de CDS prédites	6335	5945	6276	5706	7132	5153	10333	7356	8479
N° d'accession Genbank	NC_012808	NC_010725	NC_011757	NC_012988.1	NZ_AGJK00000000.1	NC_010172	NC_011894	NC_010505	NC_010511
Chromosome									
Taille (Mb)	5,51	5,8	5,77	5,94	6,73	5,47	7,78	6,1	7,70
Contenu en G+C (%)	68,7	69,4	68,2	68,1	68,23	68,2	68,9	71,5	71,6
Nombre de plasmides	4	2	2	2	0	0	7	8	2
Séquençage (date)	UW Genome Sequencing Center, (2003/2006)	JGI (2008)	JGI (2008)	Génoscope (2006)	JGI (2010)	JGI (2007)	JGI (2009)	JGI (2008)	JGI (2008)
Publication du génome	Vuilleumier *et al.*, 2009	Marx *et al.*, 2012	Marx *et al.*, 2012	Vuilleumier *et al.*, 2009	Koskimäki *et al.*, 2015	Marx *et al.*, 2012	Marx *et al.*, 2012	Marx *et al.*, 2012	Marx *et al.*, 2012

[a] Génoscope (http://www.cns.fr/agc/microscope/home/index.php) ou Craig Venter institute (http://www.jcvi.org/)

[b] Originellement nommée *Methylobacterium populi* (Van Aken *et al.*, 2004), mais sur la base de l'identité de l'ARNr 16S a été affiliée à *M extorquens* (Marx *et al.*, 2012)

4.2. Outils génétiques disponible chez *M. extorquens*

De nombreux outils génétiques sont disponibles chez *M. extorquens* AM1. Ces outils génétiques sont utilisés dans des approches de mutagénèse aléatoire, de délétion ou d'insertion ciblée de gènes, ainsi que d'expression génique constitutive ou inductible. Une liste des outils génétiques disponibles chez *M. extorquens* AM1 a été mise à jour récemment dans une revue (Ochsner *et al*., 2015).

4.3. Approches globales chez *M. extorquens*

Les connaissances portant sur le métabolisme méthylotrophe chez les bactéries du genre *Methylobacterium* et plus particulièrement de l'espèce *M. extorquens* ont considérablement augmenté ces dernières années grâce aux apports de séquençage nouvelle génération à haut-débit « *Next Generation Sequencing* » ou NGS, permettant l'étude des gènes (génomique), de leurs expressions (transcriptomique), ainsi que l'étude du protéome (protéomique) et des métabolites (métabolomique).

4.3.1. Etudes génomiques chez *M. extorquens*

Un projet de séquençage du génome complet de AM1 a été entrepris en 1998 (Chistoserdova *et al*., 2003). L'assemblage du génome n'a été complété qu'en 2009 lorsque celui de la souche DM4 a été disponible (Vuilleumier *et al*., 2009). Le génome de la souche *M. extorquens* CM4 a été disponible en même temps que le génome d'autres souches non déchlorantes de la même espèce (BJ001; PA1) (Marx *et al*., 2012). Au début de ma thèse, la plateforme MaGe, mise en place par le Génoscope (Evry, France), permettait l'accès à l'annotation et aux analyses comparatives de six génomes de *M. extorquens* (AM1, BJ001, CM4, DM4, DSM 13060 et PA1) dont on connaissait bien le phénotype de croissance sur le méthanol et les méthanes chlorés. La première analyse génomique comparative de souches de *M. extorquens* a démontré la présence de nombreux éléments d'insertion (IS) ainsi qu'une organisation groupée de ces gènes, notamment ceux du métabolisme du méthanol, suggérant ainsi le rôle prédominant des IS dans l'évolution des génomes de souches phylogénétiquement proches comme AM1 et DM4 (Vuilleumier *et al*., 2009a). Cependant, l'histoire évolutive des génomes de *M extorquens* n'est pas encore élucidée. En effet, bien que *M. extorquens* AM1 et PA1 aient un taux de GC proche (respectivement 68,2 % et 68,5 %), et partagent plus de 90 gènes

impliqués dans la méthylotrophie avec au moins 95 % d'identité, une étude récente montre que ces souches présentent des taux de croissance différents avec des substrats en C_1, comme le méthanol, la méthylamine ou encore le formaldéhyde suggérant l'implication d'autres mécanismes (Nayak et Marx, 2014).

Des études de mutagénèse aléatoire et de protéomique comparative ont mis en évidence qu'en croissance sur un méthane chloré, *M. extorquens* requière la présence de gènes localisés sur des îlots génomiques, mais également des gènes de ménage impliqués dans la réponse au stress, ainsi que dans le métabolisme central (Michener *et al*., 2014a; Muller *et al*., 2011a; Roselli *et al*., 2013). L'utilisation de chlorométhane ou de dichlorométhane génère un stress cellulaire auquel la cellule doit s'adapter (Kayser et Vuilleumier, 2001; Kayser *et al*., 2002; Roselli *et al*., 2013). Pour comprendre ces adaptations, Michener *et al.,* 2014 ont testé la capacité d'utilisation du dichlorométhane chez des souches non dichlorométhane-dégradantes. Pour cela, le gène *dcmA*, inséré sur un plasmide, a été transféré dans la souche AM1 (incapable d'utiliser le dichlorométhane pour sa croissance) (Michener *et al*., 2014b). Les transconjugants obtenus ont une croissance difficile en présence de dichlorométhane, mais montrent une activité de déshalogénation. Le séquençage du génome de ces souches a permis l'identification de mutation dans le gène *clcA* codant un antiporteur Cl^-/H^+. L'introduction d'un plasmide contenant les gènes *dcmA* et *clcA* mutés dans la souche *M. extorquens* PA1 (incapable de croître en présence de dichlorométhane) a permis d'établir une croissance efficace en présence de dichlorométhane chez cette souche.

Des travaux complémentaires, portant sur l'adaptation à l'utilisation du chlorométhane par des souches non chlorométhane-dégradantes, ont été réalisés et font l'objet d'un article en cours de soumission (Michener *et al*., soumis). Un plasmide contenant des gènes de la voie *cmu* (*folD*, *metF2*, *paaE-like*, *purU*), et des gènes associées (*fmdB*, *hutI*) ont été transféré dans des souches non chlorométhane-dégradantes, *M. extorquens* AM1, PA1, BJ001, *M. radiotolerans* et *M. nodulans*. Les transconjugants de chacune de ces souches montrent une croissance sur chlorométhane, bien que la croissance en présence de chlorométhane soit faible. Si le gène *clcA* muté, identifié comme bénéfique à l'adaptation à la croissance sur le dichlorométhane, est inséré dans le plasmide en plus des autres gènes du cluster *cmu*, cela n'améliore pas le fitness des transconjugants sur le chlorométhane. Ceci suggère que l'adaptation à la croissance sur le chlorométhane et le dichlorométhane impliquerait des réponses adaptatives distinctes (Michener *et al*., submitted article).

Ainsi, malgré de nombreuses études sur la méthylotrophie, de nombreux points ne sont pas élucidés, d'où l'importance de l'utilisation d'approches diverses mais néanmoins complémentaires comme la transcriptomique, la protéomique ou la métabolomique pour comprendre l'ensemble des mécanismes impliqués dans la méthylotrophie chez *M. extorquens* (Tableau 1.7 complété à partir de Ochsner *et al*., 2015).

Tableau 1.7. Etudes «omiques» réalisées dans des souches de *M. extorquens*

Etude « omique »	Réfèrence	Description
Génome	Vuilleumier *et al*., 2009	AM1, DM4
	Roselli *et al*., 2013	CM4
	Muller *et al*., 2011	DM4
	Bai *et al*., 2015	Isolats de la phyllosphère et de la rhizosphère (24 génomes)
Modélisation	Peyraud *et al*., 2011	AM1
Protéomique	Bosch *et al*., 2008; Laukel *et al*., 2004	AM1, comparaison méthanol et succinate
	Muller *et al*., 2011	DM4, comparaison dichlorométhane et méthanol
	Roselli *et al*., 2013	CM4, comparaison chlorométhane et méthanol
	Guo et Lidstrom, 2008	AM1, « Profiling » des métabolites en méthanol et succinate
Transcriptomique	Okubo *et al*., 2007	AM1, comparaison méthanol et succinate (microarray)
	Francez-Charlot *et al*., 2009	AM1, mutant *phyR* de la régulation du stress général (microarray)
Métabolomique	Kiefer *et al*., 2008, 2011	AM1, Concentration des métabolites centraux en méthanol et succinate
	Peyraud *et al*., 2012	AM1, [^{13}C]-fluxomique, [^{13}C]-CH_3OH en co-utilisation avec le succinate
	Yang *et al*., 2013	AM1, ^{13}C méthanol, marquage de métabolite au cours la croissance
	Cui *et al*., 2016	AM1 avec surexpression d'éthylmalonyl-CoA mutase en méthanol
	Reaser *et al*., 2016	AM1, suivi au cours du temps de métabolite après incorporation de [^{13}C]-CH_3OH

4.3.2. Approches de transcriptomique chez *M. extorquens*

La transcriptomique est l'étude qualitative et quantitative du transcriptome, c'est-à-dire, l'ensemble des ARNs produits lors de la transcription (Wang *et al*., 2009). En fonction des conditions de croissance, la cellule modifie le type de transcrits produits ainsi que leurs abondances. Ainsi l'étude de ces transcrits permet de mieux comprendre les adaptations mises en place par la cellule en réponse à diverses conditions de croissance.

Peu d'études de transcriptomique ont été réalisées chez *M. extorquens*. Toutes ont été réalisées chez AM1 et sont basées sur des approches de puces à ADN (microarray). Ces études ont mis en évidence des gènes intervenant dans le métabolisme des C_1, par comparaison des transcriptomes de AM1 en croissance sur du méthanol ou du succinate
(Okubo *et al*., 2007; Francez-Charlot *et al*., 2009), ou encore le rôle de *phyR*, un régulateur transcriptionnel de la réponse au stress (Francez-Charlot *et al*., 2009). Le séquençage et

l'analyse des transcrits (RNA sequencing) réalisés dans ce projet de thèse constituent la première approche de ce genre chez *M. extorquens*, mais aussi la première étude des transcriptomes chez les souches CM4 et DM4.

4.3.3. Etudes de protéomique chez *M. extorquens*

Les études de protéomique chez *M. extorquens* sont plus nombreuses que celles de transcriptomique (Tableau 1.7, Gourion *et al*., 2006; Muller *et al*., 2011; Roselli *et al*., 2013). Des travaux sur la méthylotrophie, notamment chez *M. extorquens* AM1, ont mis en évidence les protéines induites lors de croissance en présence de méthanol (substrat de référence dans l'étude de la méthylotrophie), en comparaison avec des substrats comme le succinate ($C_4H_6O_4$) (Laukel *et al*., 2004; Bosch *et al*., 2008). Comme décrit précédemment, *M. extorquens* est capable de dégrader de nombreux composés dont les méthanes chlorés. Des études réalisées chez *M. extorquens* DM4, en croissance sur du méthanol versus du dichlorométhane, ont permis de mieux comprendre les mécanismes impliqués dans l'utilisation de ces composés (voir chapitre 3) (Muller *et al*., 2011a). De même, les mécanismes de colonisation de la phyllosphère de *A. thaliana* par *M. extorquens* AM1 sont mieux compris grâce à l'utilisation de ces techniques (Tableau 1.7, (Knief *et al*., 2010, 2012; Vorholt, 2012)). Ainsi, la surexpression de la protéine régulatrice PhyR, nécessaire à la colonisation de la plante a été mise en évidence (Gourion *et al*., 2006). Concernant l'utilisation de chlorométhane par *M. extorquens* CM4, 49 protéines plus exprimées en condition de croissance sur chlorométhane par rapport à du méthanol ont été identifiées (Figure 1.17).

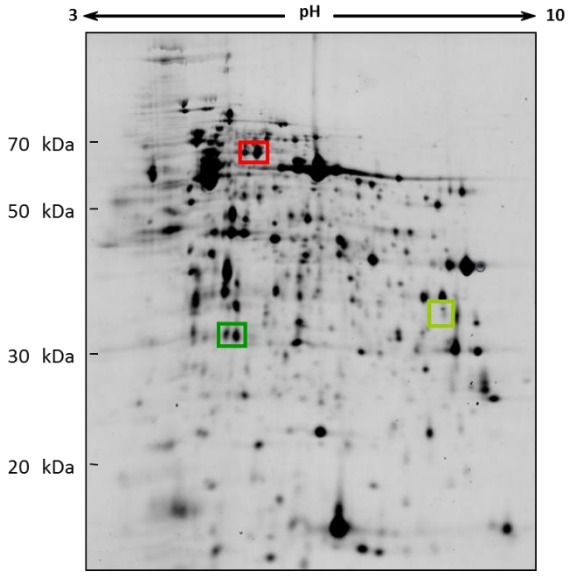


Figure 1.17. Image d'un gel 2D des protéines extraites après croissance de *M. extorquens* CM4 en présence de chlorométhane ou de méthanol
Les protéines CmuA, CmuB et PurU, encadrées respectivement en rouge, vert foncé et vert clair, ont été identifiées par spectrométrie de masse et sont plus abondantes en condition chlorométhane (Thèse de Sandro Roselli, 2009).

Cette approche de protéomique combinée à la génomique a permis de mettre en évidence les gènes requis pour l'utilisation du chlorométhane par *M. extorquens* CM4 et également l'importance des gènes portés par le plasmide pCMU01 dans la croissance avec chlorométhane (Roselli *et al*., 2013).

4.3.4. Etudes de métabolomique chez *M. extorquens*

L'ensemble des études de métabolomique réalisées chez *M. extorquens* AM1 avaient pour but de mieux caractériser le métabolisme méthylotrophe et le métabolisme central d'assimilation du carbone. La croissance sur du méthanol ou sur du succinate met en évidence la présence de métabolites spécifiquement liés à la croissance en condition méthylotrophe ou encore des métabolites intervenant dans la régénération du glyoxylate comme le β-hydroxybutyrate, le méthylsuccinyl-CoA ou encore l'éthylmalonyl-CoA (Figure 1.11) (Guo et Lidstrom, 2008). Des différences dans les métabolites produits ont été observées lors de la croissance de AM1 sur un composé en C_2 (l'éthylamine) et en C_4 (le succinate) (Yang *et al*., 2009). De même, chez *M. extorquens* AM1 en croissance avec l'acétate (CH_3COO^-) comme unique source de carbone et d'énergie, la voie de l'éthylmalonyl-CoA remplace la voie de l'isocitrate lyase pour régénérer le glyoxylate, nécessaire à l'assimilation du carbone dans la biomasse via le cycle de la sérine

(Schneider *et al.,* 2012, Figure 1.11). Ces résultats ont permis de préciser comment la cellules modifie son métabolisme pour s'adapater aux sources de carbone disponible.

5. Objectif et questions du projet de thèse

Depuis le protocole de Montréal, en 1987, réduisant l'utilisation des composés chlorés, le chlorométhane est responsable de 16 % de la dégradation de la couche d'ozone liée aux composés chlorés (Montzka *et al*., 2011). Le budget atmosphérique du chlorométhane reste mal évalué même si de nombreuses sources et puits ont été identifiés (Tableau 1.3 etTableau 1.4, Clerbaux et Cunnold, 2007; Montzka et Fraser, 2003). Le rôle du sol, à la fois source et puits de chlorométhane, dans ce budget atmosphérique du chlorométhane n'est plus à démontrer (Harper, 2000; Harper *et al*., 2003; Gribble, 2003; Miller *et al*., 2004). Cependant de nombreuses incertitudes persistent quant à l'importance des sols forestiers dans les échanges de chlorométhane avec l'atmosphère (Saito et Yokouchi, 2008). Une des incertitudes est notamment liée à la capacité des bactéries du sol à jouer le rôle de filtre de chlorométhane limitant ainsi son relargage dans l'atmosphère. Toutefois, en théorie, les faibles concentrations de chlorométhane disponibles dans les sols (1 pM. g^{-1}) (Harper *et al*., 2003) ne pourraient pas à elles seules permettre une croissance suffisante pour le maintien de communautés chlorométhane-dégradantes, qui nécessiteraient l'utilisation d'autres sources de carbone et d'énergie bio-disponibles pour leur croissance. De fait, la concentration du méthanol, estimée à 1 nM.g^{-1} dans les sols, est nettement supérieure à celle du chlorométhane (Figure 1.18, Stacheter et Kolb, 2013). Dans l'environnement, les microorganismes méthylotrophes utilisant le chlorométhane pourraient utiliser le méthane et le méthanol, en plus du chlorométhane. Les émissions de méthane et de méthanol à partir d'écosystèmes terrestres sont du même ordre de grandeur (10^{12} mol par an) (Kolb, 2009a). Le méthanol est un composé central dans le métabolisme méthylotrophe et 83 % des souches méthylotrophes poussant en aérobiose qui ont été isolées du sol sont capables d'utiliser le méthanol (Kolb, 2009). C'est le cas de la souche de référence de la dégradation aérobie du chlorométhane, *M. extorquens* CM4, incapable d'utiliser le méthane (Van Aken, 2004). Un des objectifs de mon projet de thèse était de caractériser les populations bactériennes méthylotrophes de sols forestiers capables d'utiliser le chlorométhane et de tester l'hypothèse de co-utilisation du chlorométhane et du méthanol.

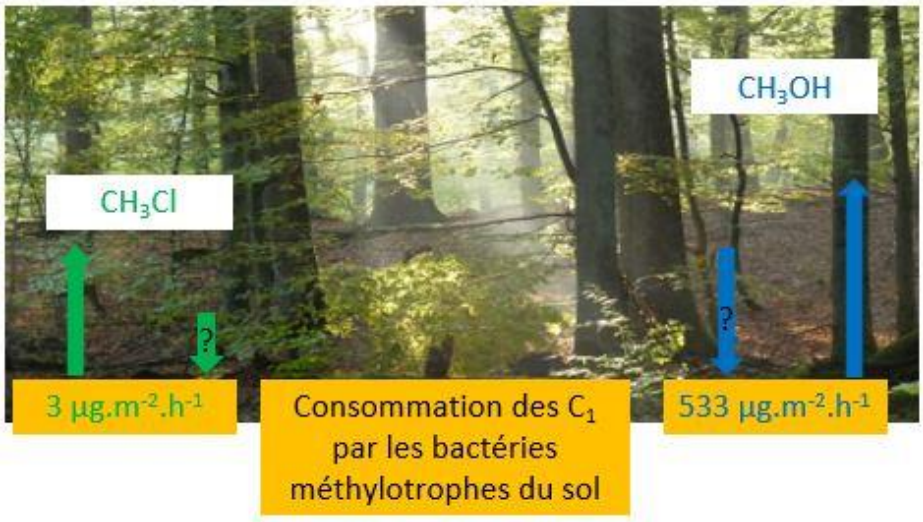


Figure 1.18. Emissions et consommations de chlorométhane et de méthanol dans les sols forestiers
Données pour les émissions de chlorométhane issues de Harper *et al*., 2003. Celles pour les flux de méthanol sont issues de Stacheter *et al*., 2013.

Mon projet s'est articulé autour de deux volets. Dans un premier temps, une étude approfondie de l'utilisation du chlorométhane en conditions contrôlées a été réalisée au laboratoire. Le séquençage de transcrits à haut-débit (RNAseq) visait à identifier l'ensemble des gènes différentiellement exprimés chez *M. extorquens* CM4 en condition de croissance méthylotrophe avec le chlorométhane et le méthanol. Le deuxième volet de mes travaux, basé sur une approche de type Stable Isotope Probing (SIP) (Neufeld *et al*., 2007) a consisté à mettre en évidence les bactéries méthylotrophes chlorométhane-dégradantes actives dans des microcosmes de sols forestiers. Ainsi des incubations avec du chlorométhane et/ou du méthanol marqués avec du [^{13}C] ont été réalisées afin de détecter les bactéries capables d'assimiler ces composés, et donc d'incorporer le [^{13}C] dans leurs ADNs. L'ADN enrichi en [^{13}C] (ADN « lourd »), correspondant à l'ADN des microorganismes ayant utilisé la source de carbone marquée, a été séparé de l'ADN non marqué (« léger » constitué de [^{12}C]) par ultracentrifugation. A partir de l'ADN « lourd », la diversité bactérienne est estimée par séquençage de produits PCR d'un marqueur taxonomique (gène codant l'ARNr 16S) et de marqueurs fonctionnels. Trois gènes marqueurs ont été utilisés ciblant :

- la dégradation aérobie du chlorométhane (gène *cmuA* codant une chlorométhane méthyltransférase de la voie *cmu*) ;
- la dégradation aérobie du méthanol (gènes *mxaF* et *XoxF* codant une sous-unité de méthanol déshydrogénase) ;

- le métabolisme méthylotrophe central (gène *mch*) (Stacheter et Kolb, 2013).

Ces approches combinées avaient pour but :

- de mieux comprendre l'utilisation du chlorométhane chez une souche modèle cultivée au laboratoire pour éventuellement pouvoir identifier de nouveaux marqueurs de la dégradation du chlorométhane en vue de détecter plus efficacement les bactéries capables de dégrader le chlorométhane dans les sols de forêts tempérées ;

- d'identifier les taxa de bactéries méthylotrophes actifs dans l'utilisation aérobie du chlorométhane en microcosmes de sols forestiers de surface et l'étude de leur rôle dans la bio-capture des émissions de chlorométhane terrestres.

Chapitre 2

Methodological section

Chapter 2. Methodological section

1. Methods used to study *M. extorquens* model strains

During this thesis, eight cDNA banks were constructed using RNA extracted from *M. extorquens* CM4, the model strain for growth with chloromethane (2 biological independent replicates), and from *M. extorquens* DM4, the model strain for growth with dichloromethane (2 biological independent replicates). In addition, RNAs were extracted from both strains grown using a methylotrophic reference substrate, methanol (2 biological independent replicates per strain).

1.1. Aerobic methylotrophic growth

Cultured of 220 mL, were performed into M3 medium defined for *Methylobacterium* with 10 mM of carbon (Muller *et al*., 2011a). To prepare 1 liter of M3 medium, 1.9 g of KH_2PO_4, 6.39 g of $Na_2HPO_4 \cdot 2\ H_2O$, 0.1 g of $MgSO_4 \cdot 7\ H_2O$ and 0.2 g de $(NH_4)_2SO_4$ are dissolved into 800 mL of milliQ water. After autoclaving, 1 mL of $Ca(NO_3)_2 \cdot 4\ H_2O$ at 25 $g.L^{-1}$ is added with 1 mL of a trace solution. This solution contains H_2SO_4 concentrated (95 %) at 9.2 $g.L^{-1}$, $Fe(II)SO_4 \cdot 7\ H_2O$ at 1.0 $g.L^{-1}$, $MnSO_4 \cdot H_2O$ at 1.0 $g.L^{-1}$, $Na_2MoO_4 \cdot H_2O$ at.25 $g.L^{-1}$, H_3BO_3 at 0,1 $g.L^{-1}$, $CuCl_2 \cdot 2\ H_2O$ at 0.28 $g.L^{-1}$, $ZnSO_4 \cdot 7\ H_2O$ at 0.53 $g.L^{-1}$, NH_4VO_3 at 0.1 $g.L^{-1}$, $Co(NO_3)_2 \cdot 6\ H_2O$ à 0.25 $g.L^{-1}$, $NiSO_4 \cdot 7\ H_2O$ at 0.1 $g.L^{-1}$. Growth conditions and total RNA extractions are described in the article “Contribution of the Core and Variable Genome to the Transcriptomes of two *M extorquens* Strains Grown with Chloromethane, Dichloromethane or Methanol” (see Chapter 3, section 3.2.).

1.2. RNA preparation

The total RNA extraction protocol is described in the Chapter 3 in the Materials and Methods section. No kit was utilized to the RNAs extraction, but a protocol with a phenol/chloroform step. These methods have some advantages in comparison with kits; no restriction in the total amount of RNA extracted, and no size selection. Once total RNA has been extracted, 2 additional steps before cDNA bank construction are needed: rRNA removal and RNA quality check.

1.2.1. DNA removal

A DNase treatment was used to remove DNA contaminants (Turbo DNAse Ambion Thermo Fischer Scientific). A total of 3 µL of Turbo DNAse was added (2U/µL) with 2 µL of RNAsin (Promega), 10 µL of TURBO™ buffer (10 X; Ambion) and 78µL of milli DEPC-treated water.

The reactionnal mix is incubated two times, at 30 minutes at 37°C, with 2µL of TURBO™ DNase added between those two incubations. Evaluation of the efficiency of DNA-free RNA extracts was checked by PCR with three sets of primers targeting the 16S RNA-encoding gene *rrsA* and *cliA* present in both strains. Additional PCR targeted genes *cmuA* and *dcmA*, which are specific of strains CM4 and DM4 respectively, were also used (see article "Contribution of the Core and Variable Genome to the Transcriptomes of two *M. extorquens* Strains Grown with Chloromethane, Dichloromethane or Methanol" (see Chapter 3, section 3.2.). PCR conditions are summarized in Table 2.1.

Table 2.1. PCR amplification program to validate DNA removal

Mix	Volume (µL)
H2O milliQ water	17.6
PCR buffer (Biolabs, 10X)	2.5
dNTP 20mM each	0.25
Forward primer (20µM)	1
Reverse primer (20µM)	1
Taq polymerase	0.1
DNA	1

PCR program	
Pre-denaturation, 2 min at 95°C	
40 cycles	Denaturation, 30 sec at 94°C
	Annealing, 45 sec at 60°C
	Extension, 30 sec at 72°C
Final extension, 10 min at 72°C	

1.2.2. rRNA removal

The abundance of ribosomal RNA (rRNA) accounts for 95-98 % of total RNA in bacteria. Thus, a step to efficiently remove rRNA is necessary to achieve optimal coverage of mRNA, good detection sensitivity and reliable results. The necessity to optimize efficient rRNA removal methods for RNA sequencing in other GC-rich bacteria has been described previously (Peano *et al*., 2013). Two types of rRNA depletion approaches exist: the first one is based on a selective degradation of rRNAs by an enzyme recognizing specifically the 5' monophosphate extremity of rRNAs, whereas mRNAs are protected by their triphosphates extremities (Evguenieva-Hackenberg and Klug, 2011). The second method relies on subtractive hybridization with probes specifically designed to target rRNA conserved regions (van Dijk *et al*., 2014). This second method was the most efficient in rRNA depletion of samples extracted from *M. extorquens* strains as tested at the Génoscope in Evry (experiments performed by

Béatrice Ségurens and Adriana Albertini). Extracted RNA samples were treated with the RiBoZero Magnetic kit Gram-negative bacteria (Tebu-Bio). Evaluation of the efficiency of RNA removal was checked by PCR with primers targeting the 16S RNA-encoding gene *rrsA* and *cliA* as a control. Despite the validation of rRNA removal by PCR and bioanalyzer profil, 14 up to 34 % of the RNAseq reads mapped to rRNA (Chapter 3 (section 3.2, Table 1). Thus, the RiboZero Magnetic Kit Gram-negative bacteria is not optimized to target *M. extorquens* rRNA (data confirmed by the Tebu-Bio company).

1.2.3. RNA quality control

Enzymes that degrade RNA, ribonucleases (RNases), are so ubiquitous (reagents, skin, dust) that working in an RNase-free environment can be challenging. The integrity of the extracted RNA was systematically checked using the Agilent biotechnology 2100 Bioanalyzer System. This system is based on classical gel electrophoresis adapted to a chip format. The RNA nano and RNA pico kits are used to detect RNA in the range of 5-500 $ng.\mu L^{-1}$ and 50-5000 $pg.\mu L^{-1}$ respectively. Each RNA sample is mixed to a fluorescent dye and loaded into a well connected to micro-channels for capillary electrophoresis. The dye binds to nucleic acids, then RNA-dye complex are detected by laser-induced fluorescence. Fluorescence is recorded for nucleic acid visualization. RNA fragments are separated based on their size and fragment size assignment is deduced from the electrophoregram of a co-migrating RNA size ladder (Figure 2.1).

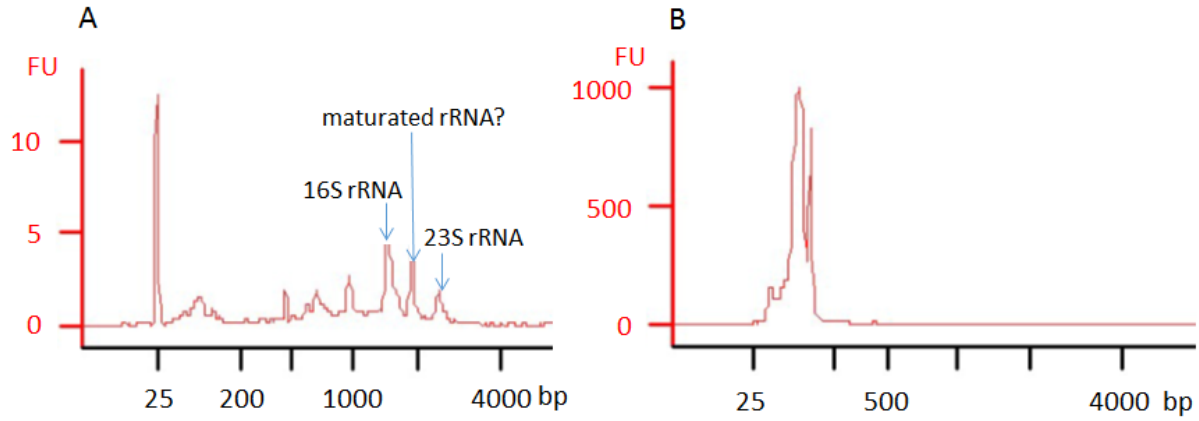


Figure 2.1. Electrophoregram of RNAs extracted from *M. extorquens* CM4 growing with chloromethane

Peak area is assigned in arbitrary fluorescence units (FU). (A) RNA nano kit-based electrophoregram of RNA before rRNA depletion (RIN value of 6.7). In *M. extorquens* an additional band is observed. (B) RNA pico kit-based electrophoregram of RNA after the rRNA depletion (RIN value of 2.6).

Profiles generated on the Agilent 2100 Bioanalyzer System yield information on concentration, allow a visual inspection of RNA integrity, and generate ribosomal ratios (16S versus 23S rRNA). Two kits are available for RNA analysis: RNA nano and RNA pico kits, depending on their sensibility (5-500 $ng.\mu L^{-1}$ and 50-5000 $pg.\mu L^{-10}$ respectively). The manufacturer Agilent Technologies has introduced the RNA Integrity Number (RIN). The RIN is an algorithm for assigning integrity values to RNA measurements: a value of 1 means that RNAs are totally degraded while a value of 10 means "perfect" RNAs integrity with no detected degradation. The RIN value is defined by a complex calculations fixed by the software using the ratio 16S/23S. Moreover, the RIN value should be use with caution. In some bacteria, RNAs can be naturally fragmented, as described in *M. extorquens* (Zahn *et al*., 2000; Evguenieva-Hackenberg, 2005). This maturation will lead to the 23S fractionation, illustrating by the apparition a new peak between those for the 16S and 23S, as observed in Figure 2.1.

1.3. Construction of the directional cDNA bank

To construct the cDNA banks, Béatrice Segurens and Adriana Albertini performed preliminary tests at the Génoscope in Evry. Three protocols were tested basing on the TruSeq Small RNA, the TruSeq stranded mRNA, and a Génoscope "house" protocol. The TruSeq stranded mRNA LT kit (Illumina) was chosen. The directional bank relies on dUTP incorporation during the second strand synthesis. This will avoid the amplification of the second strand, keeping only information from the first strand cDNA. The addition of two different adapters, in 3' and 5' extremities will keep the strand orientation. Fifty to sixty ng of rRNA-depleted RNA were used to construct directional cDNA banks as described in Figure 2.2.

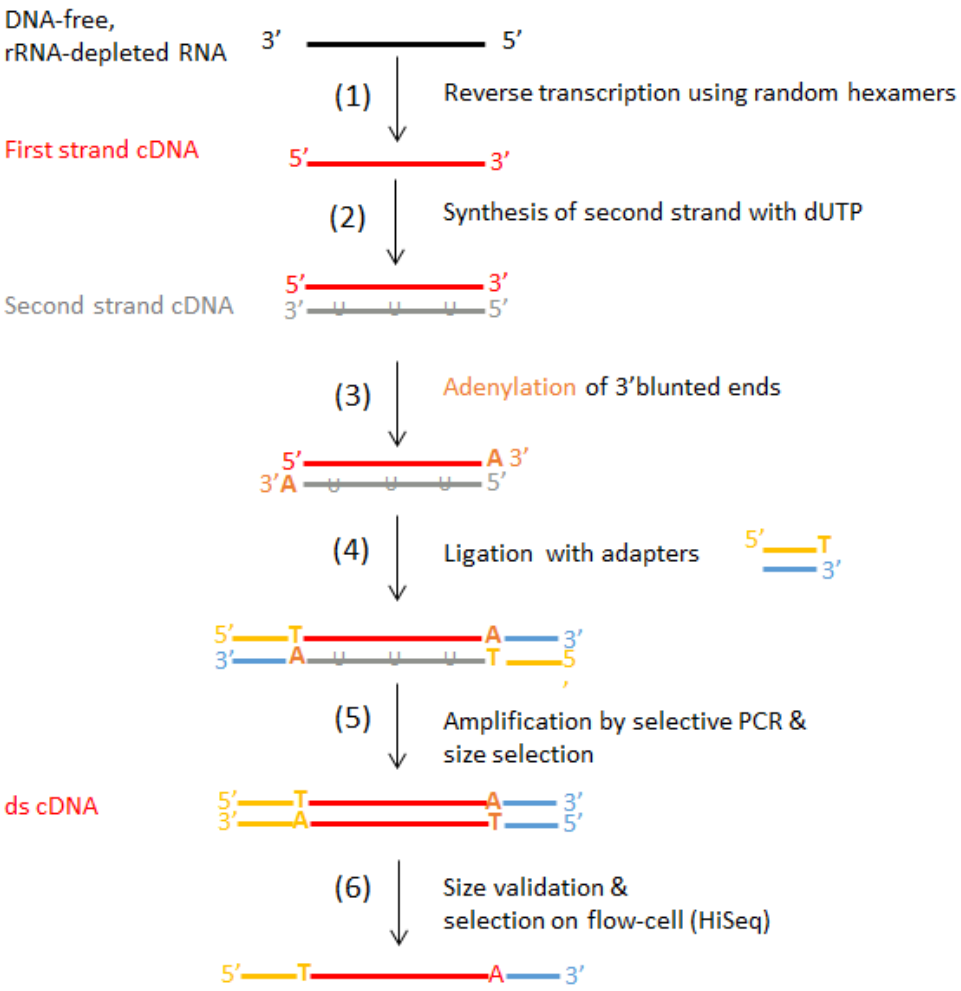


Figure 2.2. Directional cDNA bank construction
The TruSeq stranded mRNA kit (Illumina) was used. (1) RNAs were primed with random hexamers to synthetize the first cDNA strand by reverse transcription. (2) The complementary cDNA strand was synthetized with the incorporation of dUTP instead of dTTP, whereas the RNA template was degraded; (3) A single adenosine nucleotide (A) was added to the 3' ends of the blunt extremity of each cDNA fragment to prevent them from ligation with other cDNA fragments; (4) Ligation of multiple indexing adapters to both the 5' and 3' ends of the cDNA fragments; (5) Selectively enriched DNA fragments: PCR amplification and treatment with magnetic beads (AMP Xpure, Agencourt, Agilent) for PCR product size selection; (6) cDNAs with a size around 260 bp were hybridized to surface-bound primers in the flow cell. Then sequencing can start.

The cDNA bank constructions were validated if the final product had an expected size around 260 bp using the Agilent DNA 1000 kit. An example of an electrophoregram of cDNA with a DNA chip is shown (Figure 2.3). Validated cDNA banks were sent to the Génoscope to be sequenced in paired-end with the HiSeq2000.

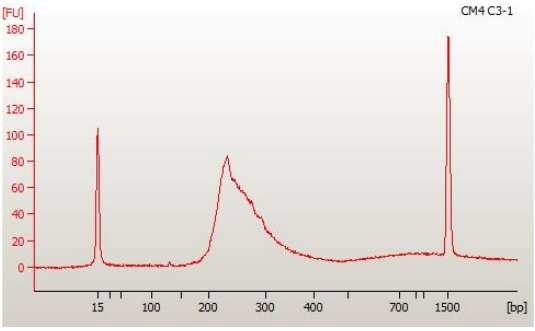


Figure 2.3. Validation of cDNA bank constructions with the 2100 Bioanalyzer

2. Methods used to study chloromethane-degrading bacterial communities in forest soils

The environment is a complex system, with a high bacterial diversity, and a lot of different metabolisms. The majority of bacteria found in the environment is not cultivable in the laboratory (Pham and Kim, 2012) so that the identification *in situ* of a specific bacterial associated-metabolism may be challenging (Rappé and Giovannoni, 2003). Cultivation-independent methods have been developed to identify which microorganisms are carrying out a specific set of metabolic processes in the natural environment (Coyotzi *et al.*, 2016). Among those methods, the stable isotope probing (SIP) method relies on the incorporation of a labeled substrate that is highly enriched in a stable isotope, such as [^{13}C], in order to identify the active microorganisms by the selective recovery and analysis of isotope-enriched cellular components (Dumont *et al.*, 2011). The stable isotope of carbon [^{13}C] is found naturally although it only represents about 1.1 % of all natural carbon on Earth while carbon [^{12}C] makes up to 98.9 %. The labeling of microorganisms with the [^{13}C]-carbon is the direct consequence of the [^{13}C]-labelled substrate consumption. The approach has been applied to lipids (Evershed *et al.*, 2006), nucleic acids (Neufeld *et al.*, 2007) and proteins (Jehmlich *et al.*, 2008). The incorporation of heavy isotopes into newly synthesized nucleic acids has been used to characterize the diversity of active bacteria in soil, after separation of newly synthesized from existing rRNA and DNA in RNA-SIP and DNA-SIP experiments, respectively (Rettedal and Brözel, 2015). RNA-SIP targets transcriptionally active cells producing new ribosomes after addition of the label, whereas DNA-SIP focuses on DNA replicating, yielding an inventory of all dividing cells (Chen and Murrell, 2010). Since rRNA-SIP is more sensitive and does not require cell division to label RNAs, the RNA-SIP approach was envisaged in addition to the DNA-SIP to characterize active chloromethane-degraders in forest soils. Nevertheless, reverse

transcribed-soil RNA, were used as template for PCR amplification of *cmuA*, *mch* and *mxaF/xoxF*, but no amplification was observed, that's why RNA were not further studied. Moreover, the DNA-SIP approach allowed subsequent metagenomic analysis.

2.1. Soil sampling and preincubation

The soil sampling was performed at a forest site that now serves as a model forest ecosystem for investigation of methylotrophic communities of temperate climate zone is the forest at Steigerwald (49°52'N, 10°28'E; sandy loam (soil type: Dystric Cambisol) with a pH of 4.6; mean annual temperature: 7.5°C; mean annual precipitation: 725 mm). The Steigerwald forest covers an area of 115,2 km^2 located between Würzburg and Nürnberg in Germany. And though the area is developed, there are no industrial activities nearby and access is limited. A total of 5 different samples were taken from the upper layer in an area of 20 m of diameter, in September 2014. Mycelium and leaves were removed. The soil samples were sieved through a 2 mm screen and 500 g of each sieved soil sample were mixed. Before the SIP experiment per se, a pre-incubation was performed during two weeks at 20°C in the dark in presence of 1 % of [^{12}C]-CH_3Cl per 270 g in 1L Müller-Krempel flasks. The resulting activated soils were exposed to air to remove remaining chloromethane gas, and mixed together. The average water content was determined to be of 18 % by measurement of the weight difference before and after drying (ten days at 60°C) of 1.2 g of soil placed into an Eppendorf tube.

2.2. Stable isotope probing (SIP): the [^{13}C]-chloromethane labeling step

Microcosms were set up using 100 g of the activated soil introduced into sterile 0.5 L Müller-Krempel flasks. Chloromethane and methanol were added alone or in combination in a total of 5 successive carbon pulses per condition. The next carbon pulse occurred when chloromethane concentration decreased below 10 % of the initial concentration. In parallel, each methanol pulse took place at the same time as the reference chloromethane pulse. To optimize carbon availability and homogenization without transforming the soil into mud, 1 mL of liquid phase (water or methanol) was added into each microcosm per pulse. Control conditions without added carbon sources were included in the study (Figure 2.4). Each pulse of chloromethane corresponded in the addition of 1 % of CH_3Cl. This resulted in a final

concentration of 18 mM (volume flask of 0.5L; chloromethane solubility of 5 g. L^{-1}). Equivalent amounts of chloromethane and methanol were added per each pulse (see Table 4.1 in Chapter 4). Thus, for methanol, 1mL of a methanol stock solution at 216 nM was added per pulse (70g of soil contains 12 mL of water) to obtain a final concentration of 18 mM.

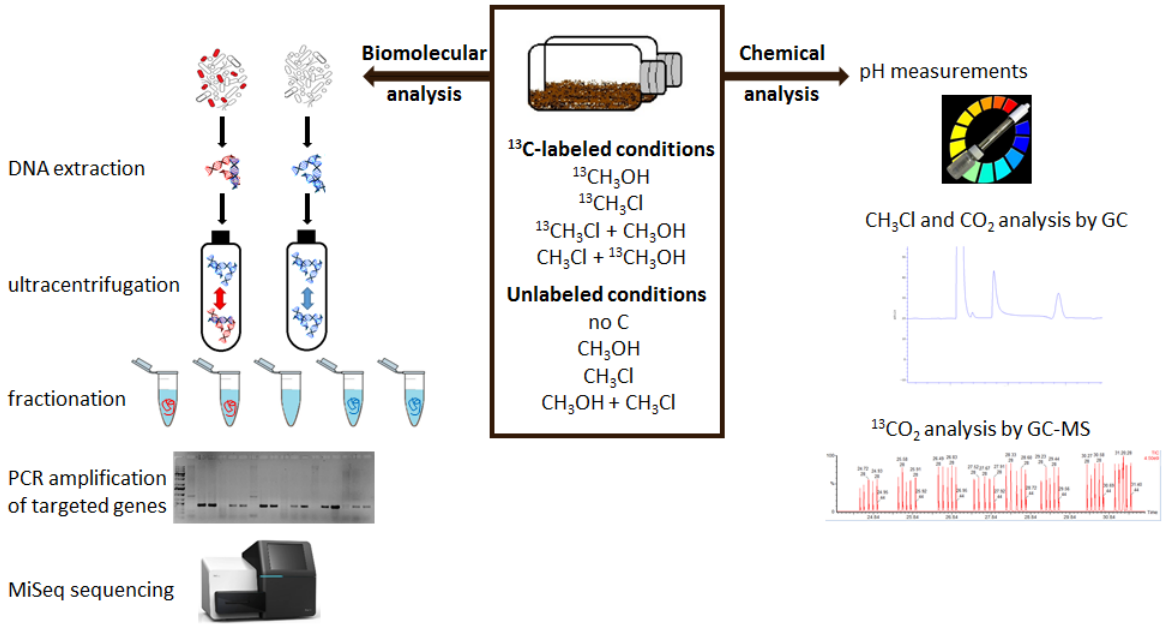


Figure 2.4. Overview of the stable isotope probing experiment

During the course of the labeling experiment, four parameters were monitored:

a) **pH**. CH_3Cl consumption or CO_2 reactions may produce protons and subsequently modify bacterial diversity and hinder the detection of the chloromethane-degrading communities. To be sure that pH values do not change during the experiment, the pH was checked on sieved soil, before the pre-incubation step, and after each carbon source pulse. The pH was measured by suspending 0.5 g of soil into 1 mL of water using a pH electrode (R422 InLab Semi-Micro; Mettler-Toledo).

b) **Chloromethane concentrations**. Chloromethane concentrations were monitored using gas chromatography (GC) directly after CH_3Cl has been added to the microcosms. To correlate the peak area and CH_3Cl concentration, a calibration curve was done, by using 500 µL of standard CH_3Cl concentrations prepared at 10 %, 2 %, 1 %, 0.5 %, 0.25 %, 0.001 % and 0.005 % in serum flasks (correlation factor R^2 of 0.95).

c) **Total CO_2 concentration.** The CO_2 was measured by using a thermal conductivity detector (TCD). For more details, see Chapter 4. As for CH_3Cl, a calibration curve was realized to correlate peak areas and CO_2 concentrations.

d) **[^{13}C]-CO_2.** A carbon source can be assimilated into biomass, or reduce to produce energy leading to a CO_2 production. To estimate the part of carbon mineralized, the [^{13}C]-CO_2 was measured by gas chromatography coupled with mass spectrometry (GC-MS).

2.3. DNA extraction from soil

The bacterial community DNA was extracted from soil after the third chloromethane addition pulse according to Griffith *et al*., (2000) with minor modifications. Elution buffer was prepared by mixing equal volume of potassium-phosphate buffer and hexadecyltrimethylammonium bromide (CTAB)/ NaCl solution. The potassium-phosphate buffer was prepared by mixing 39.30 g of K_2HPO_4 and 1.96 g of KH_2PO_4 qsp 1000 mL of milliQ H_2O and adjusted at pH 8. The CTAB/ NaCl solution was prepared with 10 g of CTAB with 4.09 g qsp 100 mL of DEPC water and then heated at 65°C. After mixing potassium-phosphate buffer and CTAB/NaCl solution, the elution buffer was mixed and heated at 60°C. Then, 0.5 mL was added to 0.5 g of soil previously placed into a tube with sterile zirconium beads (0.5 g of those of 0.1 mm and 0.5 g of 0.5 mm diameters, Roth) before to be mixed using a vortex. A volume of 0.5 mL of phenol/chloroform/isoamyl alcohol (ratio 25:24:1) was added before to place the tube in a bead beater for 30 seconds at 5.5 m.sec^{-1} and immediately placed on ice before to be centrifuged at 4°C at 13,000g during 5 minutes. The upper phase was transferred in another tube and mixed by inversion with chloroform/isoamylalcohol (24:1) and centrifuged at 4°C at 13,000 g during 5 min. The supernatant was transferred to another tube with 2-3 volume of precipitation buffer, mixed and incubated at room temperature for two hours. This buffer was prepared with 30 % PEG 3,000 at pH 7 in 0.1 M of 2-[4-(2-hydroxyethyl) piperazin-1-yl] ethane sulfonic acid (HEPES). The samples were centrifuged at 4°C during 10 min at 13,000 g then the pellet was washed with ethanol 70 % before to be centrifuged again, and dried at room temperature. The DNA pellet was suspended in 30 µL DEPC water and treated with RNAse A to remove contaminant RNAs. DNA was quantified by fluorescence with the Quant-iT™ PicoGreen® kit (Invitrogen).

2.4. Separation of isotopically-labeled DNA by isopycnic density gradient

The “heavy”-DNA is separated from the “light”-DNA by ultracentrifugation in a cesium-chloride (CsCl) gradient where DNA molecules migrate to the point where they have the same density as the gradient. The CsCl solution and gradient buffer were prepared as recommended

(Neufeld *et al*., 2007). To dissolve CsCl in milliQ water (1 g mL^{-1}), the solution was heated at 30°C. The density of the CsCl solution at 20°C absolutely needs to be between 1.88 and 1.89 g mL^{-1}. To prepare the gradient buffer, 3.75 g of KCl, 50 mL of Tris-HCl, 1 mL of 0.5 M EDTA were dissolved in a final volume to 500 mL of milliQ water. The CsCl solution and the gradient buffer were mixed in the ratio 4:1 to give an expected density of 1.725 $g.mL^{-1}$ at 20°C. This gradient solution (4.98 mL) was mixed to DNA (5 µg in 20 µL of RNA-free milliQ-water), and transfered into a column Quick-Seal (Beckman Coulter). A control (without added nucleic acids) was run in parallel to validate the density gradient after ultracentrifugation. Each tube was sealed and placed into the rotor Vti 65.2 (Beckman Coulter) for ultracentrifugation for 40 hours at 177,000 g at temperature 20°C (Figure 2.5). To avoid gradient destruction, a deceleration phase of 2 hours was programmed. After the centrifugation, the ''heavy"-DNA and the "light"-DNA were located at the bottom or the top of the tube, respectively. Immediately after the rotor stopped, the isotopically-labeled and unlabeled nucleic acids present along the isopycnic density gradient were separated in 10 fractions per tube. According to the literature, the collected fractions 6-7 are expected to have a density of 1.725 g. mL^{-1} and all other fractions a density comprised between 1.690 and 1.760 $g.mL^{-1}$ (Neufeld *et al*., 2007). The density of the collected fractions was established from a standard curve established by weighting each fraction of the control tube after the temperature has been stabilized at 20°C in a water bath.

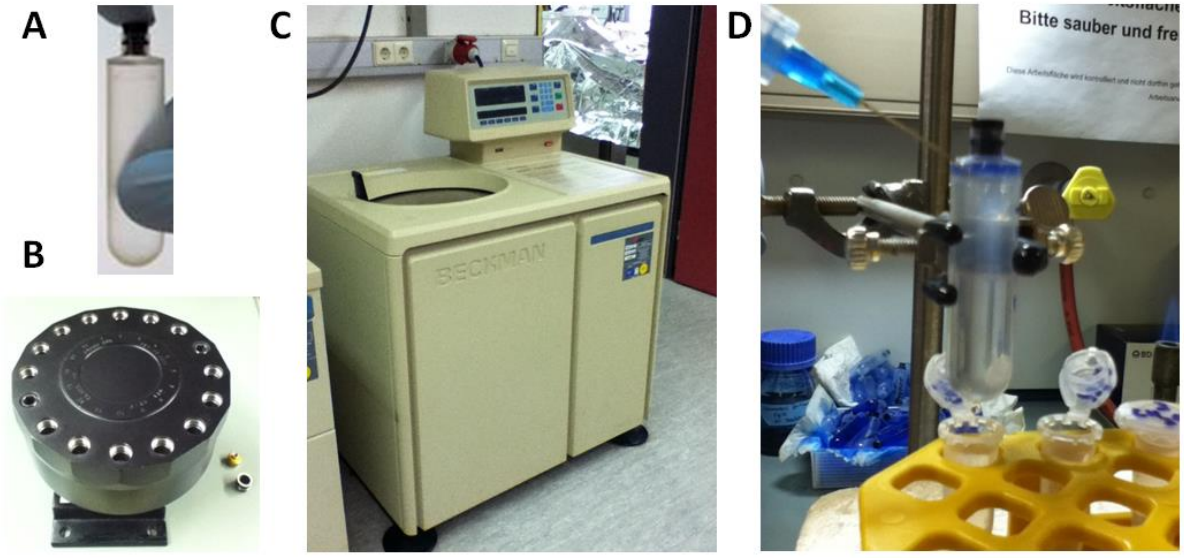


Figure 2.5. Material required for ultracentrifugation and fractionation
(A) Nucleic acids dissolved in the gradient solution are transferred to a microcentriguge tube (Tube, Quick-Seal, Beckman Coulter); (B) A total of 16 tubes can be placed vertically into the rotor Vti 65.2 (Beckman Coulter). Each tube is circumscribed in a rotor well with a screw-on adaptor; (C) Ultracentrifuge (LE-70, Beckman Coulter); (D) After centrifugation, each tube containing the gradient is fractionated to split heavy, middle and light fractions. Each tube is fixed and pierced carefully in the top and bottom with needles (30 x 0.63 mm, BD). Using a low-flow pump (BioRad Econo Pump), a solution of water containing-dye was injected into the top with a flux fixed at 450 µL.min^{-1}. Drops are collected during one minute before to change the tube to collect the next fraction. This technique provides smooth delivery of displacement liquid, which minimizes disruption of the gradient and ensures a consistent volume in the collected fractions.

The collected nucleic acid fractions were precipitated for 2 hours at 4°C in presence of 1 µL of glycogen (20 µg.µL^{-1}) and 2 volumes of polyethylene glycol PEG 6000 30 % / 1.6 M NaCl solution (150 g PEG 6000, 46.8 g of NaCl dissolved in total volume of 500 mL of milliQ water). Precipitated nucleic acids were harvested by centrifugation at 13,000 g at 20°C during 30 minutes, washed once with ethanol 70 %, dried at room temperature, and resuspended in 30 µL of TE Buffer previously prepared to have a final concentration of 10 mM Tris-HCl (pH 8), and 1 mM EDTA (pH 8).

2.5. Metagenomics of PCR-amplified markers

To assess the diversity of chloromethane-degraders in forests soil, the *rrsA* gene was amplified in the different microcosms prepared. In addition, genes encoding enzymes associated with mono carbon metabolisms are suitable to detect methylotrophs in the environment

(McDonald *et al*., 2005). The genes *cmuA*, *mxaF*/*xoxF* and *mch*, encoding respectively a chloromethane deshalogenase, a methanol dehydrogenase subunit and methenyl cyclohydrolase were targeted.

The corresponding PCR products were sent for sequencing at LGC Genomics GmbH (Berlin, Germany). To reduce sequencing costs, PCR fragments were extended by small nucleotides sequences called barcode and specific to each sample. A total of 48 barcode sequences were used; 24 barcodes for the 16S PCR amplicons (eight microcosms, with three fractions by microcosm (heavy, middle and light)), and 24 for functional genes.

2.5.1. Design of new primers targeting environmental chloromethane dehalogenase CmuA-encoding gene

PCR amplification targeting the diversity of *cmuA* in samples from contrasting environments such as soil, marine ecosystems, and phyllosphere have been performed previously using primer targeting the end of the methyltransferase domain (Miller *et al*., 2004; Borodina *et al*., 2005; Nadalig *et al*., 2011) (Table 2.2; box 1). These studies as well as sequencing of new chloromethane-degrading *cmuA* gene enriched the database allowing the design of new primers, which may be better in detecting *cmuA* genes in the environment for NGS and associated qPCR analysis.

Table 2.2. Set of primers available for *cmuA* amplification in the environment

Utilization	Product (bp)	Primer orientation, name, sequence 5'->3' (degeneracy rate)[b]	CM4 genome position (nt)[c]	GC (%)	Sequences with mismatch (number of aligned sequences)[d]	Reference
Box1 NGS	442	Forward, cmuA802F, TTCAACGGCGAYATGTATCCYGG (4)	336191-336213	52.2	*Vibrio orientalis, Thermosediminibacter oceani, Desulfotomaculum alcoholivorax* (32)	Miller *et al*., 2004
		Reverse, cmuA1244R, TABTCCATDATGGCYTCGAC (18)	335790-335771	56.3	Marine methyl halide utilising bacteria, *Vibrio orientalis*, *Thermosediminibacter oceani*, *Thermincola potens*, *Desulfomonile tiedjei*, *Desulfotomaculum alcoholivorax* (113)	PhD thesis, Farhan Ul Haque
ox1 qPCR	166	Forward, cmuA802F, TTCAACGGCGAYATGTATCCYGG (4)	336191-336213	52.2	*Vibrio orientalis, Thermosediminibacter oceani, Desulfotomaculum alcoholivorax* (32)	Miller *et al*., 2004
		Reverse, cmuA968R, CCRCCRTTRTAVCCVACYTC (166)	336066-336047	57.0	***Vibrio orientalis*** (113)	Nadalig *et al*., 2011
Box2 NGS	422	Forward, cmuAf422 GARGTBGGITAYAAYGGHGG (24)	336046-336066	52.5	***Vibrio orientalis*** (113)	This work
		Reverse, cmuAr422 TCRTTGCGCTCRTACATGTCICC (4)	335666-335644	52.2	*Thermosediminibacter oceani*, *Vibrio orientalis*, *Desulfomonile tiedjei*, uncultured bacterium (113)	This work
Box2 qPCR	121	Forward, cmuAf422 GARGTBGGITAYAAYGGHGG (24)	336046-336066	52.5	***Vibrio orientalis*** (113)	This work
		Reverse, cmuA1070R AYRTCCATCGGRTAYTTRAT (32)	335964-335945	38.0	***Vibrio orientalis*** (113)	This work
Box2 qPCR	144	Forward, cmuA1227F, CGARGCSATIATGGAITAYGA (8)	335767-335788	43.0	*Vibrio orientalis, Thermincola potens* (113)	This work
		Reverse, cmuAr422 TCRTTGCGCTCRTACATGTCICC (4)	335666-335644	52.2	*Vibrio orientalis, Thermincola potens* (113)	This work

[a] Sets of primers target either box 1 or box 2 of *cmuA* genes, which are localized at nucleotides 336191-335790 and 336191-335790, respectively for *M. extorquens* CM4 genome (see footnote [c]). NGS stands for Next Generation Sequencing (MiSeq from Illumina)

[b] Number of sequences in degenerate primers synthesized as mixtures of different sequences with B=C/G/T, D=A/G/T, H=A/C/T, I=inosine, R=A/G, S=G/C, Y=C/T, V=A/C/G, W=A/T

[c] MaGe database http://www.genoscope.cns.fr/agc/mage)

[d] Divergent sequences include sequences with a 3'-terminal mismatch (in bold) or one mismatch at the second last position of the 3'-terminal position or at least three mismatches. Sequence accession n° Desti_5447 (*Desulfomonile tiedjei* DSM 6799, Uniprot); TherJR_0143 (*Thermincola potens* JR, Uniprot); Toce_1533 (*Thermosediminibacter oceani* DSM 16646, Uniprot); VIA_000821 (*Vibrio orientalis* ATCC33934, Uniprot); WP_027363948 (Desulfotomaculum alcoholivorax, GenBank); Gi/66474906 (Marine methyl halide utilizing bacteria); Gi/664795/1 (Uncultured bacteria). *Desulfotomaculum kuznetsovii* DSM 6115 was excluded from this study.

A total of 113 *cmuA* sequences were retrieved from NCBI database after a BLAST search using *M. extorquens cmuA* (Mchl_5697) sequence, and subsequently aligned using clustalW-multialign, Mobyle (http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::clustalw-multialign alignment). Most divergent sequences shared 48.9 % of aa identity. Among those sequences, uncultured bacterium isolated from *cmuA* sequences from strains detected in soil (Miller *et al*., 2004; Borodina *et al*., 2005), marine environment (Cox *et al*., 2012) and also *cmuA* sequences from *Proteobacteria* isolates that belong to the class of:

- *Alphaproteobacteria* of the genus *Methylobacterium* (strain CM4), *Hyphomicrobium* (strains AT1, AT2, AT3, MC1), *Aminobacter* (strains *lissarensis* CC495, *ciceronei* IMB-1, TW3), *Rhodobacteraceae* (strains 179, 198);
- *Deltaproteobacteria* of the genus *Desulfotomaculum (D. alcoholivorax*, *D. kuznetsovii*) and *Desulfomonile tiedjei;*
- *Gammaproteobacteria* such as *Vibrio orientalis*.

Two *cmuA* sequences of *Clostridia* isolates were also added in this study (*Thermincola potens* JR and *Thermosediminibacter oceani*).

Based on the alignment, new sets of primers (box 2) were designed for NGS and qPCR primers (Table 2.2). The PCR program for Box 2 targeted primers starts with pre-denaturation of 10 min at 95°C, then 30 cycles of two steps: denaturation of 15 seconds at 94°C followed by an annealing/elongation of 60 seconds at 56°C. The program ends with a post-elongation of 5 minutes at 72°C. The program for the primer set cmuA802F/cmuA1244 (box 1) was performed as previously described (Miller *et al*., 2004). To validate those new primers, PCR amplifications were performed on three different forest soils, a native forest one, a [^{12}C]-CH_3Cl pre-incubated forest soil, and one incubated with [^{13}C]-CH_3Cl, and compared to the old primers (box 1). Amplifications were more specific with the new primers cmuAF422/cmuAR422 (box 2) than with cmuA802F/cmuA1244R (box 1), but also more efficient with primers cmuA802F/cmuA1244R ten time more sensitive than others (Figure 2.6). The new primer set is still unable to amplify the *cmuA* gene found in *Vibrio orientalis* (Table 2.2). This organism is unable to dechlorinate chloromethane as tested by Ludovic Besaury, a post-doctorate fellow in the laboratory (data not shown). To confirm that no *Vibrio orientalis cmuA*-like sequence is present into the studied forest soil, a PCR reaction with primer set specially designed to target

Vibrio orientalis cmuA sequence was performed. No amplification was detected in the DNA from soil while the positive control with *Vibrio orientalis* DNA template was amplified.

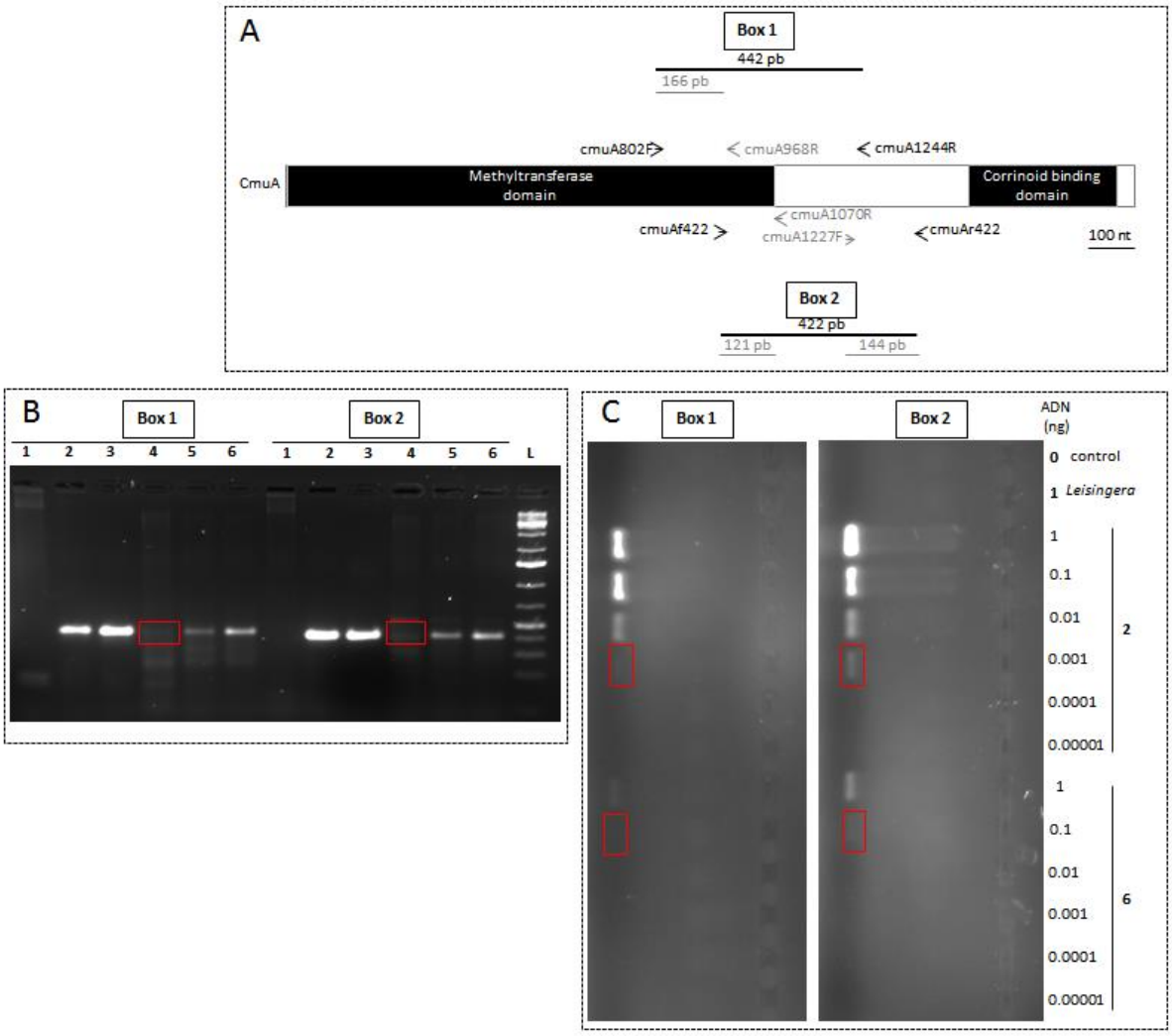


Figure 2.6. Comparing two sets of primers targeting environmental chloromethane dehalogenase CmuA-encoding gene

Gene *cmuA* PCR amplification targets the region between the conserved domains (in grey color) IPR006360 and IPR006158 of the encoded protein CmuA. A) Position of the primer box 1 (Miller *et al*., 2004) and primer box 2 (this study as detailed in Table 2.2). For each primer, its corresponding name, localization and orientation (arrow) is shown. Amplification DNA fragments are indicated with lines, in grey when designed for qPCR only. Number indicate product size in bp. B) PCR-amplified *cmuA* products using reference genomic DNA and environmental DNA. Amplifications were performed using 10 ng of DNA. Track 1-3, DNA from isolated chloromethane-degrading strains including the negative control *Leisingera methylohalidovorans* (track 1) (*cmuA* undetected, Schäfer *et al*., 2004); *M. extorquens* CM4 (2) and *Hyphomicrobium* sp. MC1 (3) are the positive controls. Tracks 4-6, environmental DNA from a native forest soil (4), a [^{12}C]-CH_3Cl pre-incubated forest soil (5), and a forest soil incubated with [^{13}C]-CH_3Cl (6). C) Sensitivity of two sets of primers sets, in PCR-amplification of *cmuA* using DNA from isolated chloromethane-degrading strains including the negative control *Leisingera methylohalidovorans* and a forest soil incubated with [^{13}C]-CH_3Cl (6).

2.5.2. Design of new primers targeting environmental methanol dehydrogenases MDH-encoding genes

In soil, species belonging to *Alpha-*, *Beta-*, and *Gammaproteobacteria*, *Verrucomicrobia*, *Firmicutes*, and *Actinobacteria* were described as involved in methanol oxidation (Kolb, 2009b). A pyrroloquinoline quinone (PQQ)-dependent methanol dehydrogenase encoded by *mxaFI* is found in Gram negative bacteria including Proteobacteria (Lidstrom, 2006; Chistoserdova *et al*., 2009). A similar gene, *xoxF*, encoding a (PQQ)-dependent methanol dehydrogenase, occurs in all known methylotroph genomes and has been described as essential for growth in the presence of methanol (Chistoserdova *et al*., 2011, Schmidt *et al*., 2010). It was described that *mxaF* and *xoxF* can be present simultaneously in genomes of methylotrophs (Skovran *et al*., 2011b). But *xoxF* and not *mxaF* is expressed even in samples where both genes have been detected (Beck *et al*., 2015). This prompted us to design new primers that target both *mxaF* and related *xoxF* gene (Figure 2.7).

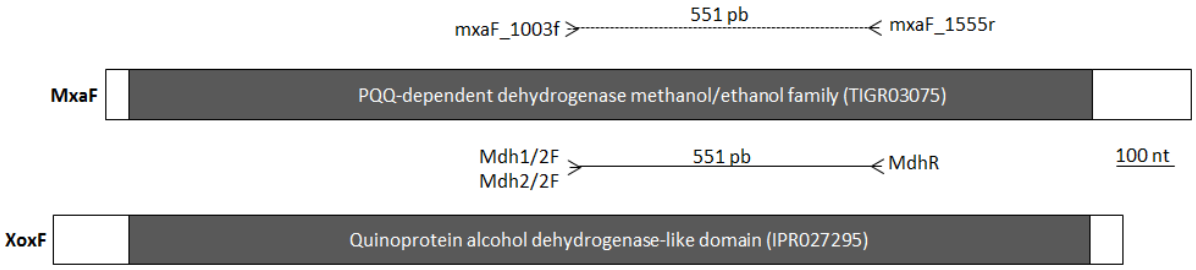


Figure 2.7. PCR amplification of methanol dehydrogenase-like MxaF and XoxF encoding genes
The dotted line shows the PCR product only detected with the "old primer set".

A total of 120 *mxaF/xoxF* sequences, mostly found in sequenced genomes, were aligned using clustalW-multialign, Mobyle (http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::clustalw-multialign alignment). Based on this alignment, primer 1555R previously used to target *mxaF* in soil (Stacheter *et al*., 2013) would poorly detect *xoxF* (Table 2.3), whereas the newly designed primer sets target all the aligned *mxaF* and *xoxF* genes (Table 2.3).

Table 2.3. Primer sets available for *mxaF_xoxF* amplification for NGS analysis in the environment

Product (bp)	Orientation, name, sequence 5'->3' (number of expected sequence)[a]	CM4 genome position (nt)[b]	GC content (%)	Primer occurrence: sequences with mismatch (number of aligned sequences)[c]	Reference
549	Forward, mxaF_1003f GCGGCACCAACTGGGGCTCGT	4786435-786452	67.6	***Hyhomicrobium denitrificans* ATCC 51888, *Hyphomicrobium sp*. MC1, *Hyphomicrobium nitrativorans* NL23, *Methylocella silvestris* BL2, *Methylobacterium extorquens* PA1**, *Methylobacillus flagellatus* KT, *Methylobacterium nodulans* strain ORS2060 (nd)	McDonald *et al*., 1997
	Reverse, mxaF_1555r CATGAABGGCTCCCARTCCAT (6)	4785922-785902	59.5	nd	Neufeld *et al*., 2007
551	Forward, Mdh1/2F GCGGIWSCAICTGGGGYT (8)	4786432-786452	69.0	***Methylocella silvestris* BL2** (97)	This work
	Reverse, MdhR GAASGGYTCSYARTCCATGCA (32)	4785925-785905	55.5	(120)	This work
551	Forward, Mdh2/2F GCGGIWSGAICTGGGGYT (8)	4786432-786452	69.0	(23)	This work
	Reverse, MdhR GAASGGYTCSYARTCCATGCA (32)	4785925-785905	55.5	(120)	This work

[a] Degenerate primers are synthesized as mixtures of different sequences with B=C/G/T, D=A/G/T, H=A/C/T, I=inosine, R=A/G, S=G/C, Y=C/T, V=A/C/G, W=A/T

[b] MxaF position, MaGe database http://www.genoscope.cns.fr/agc/mage)

[c] Divergent sequences include sequences with a 3'-terminal mismatch (in bold) or one mismatch at the second last position of the 3'-terminal position or at least three mismatches. Sequence accession n° Hden_1305/ Hden_1617/ Hden_2848 (*Hyhomicrobium denitrificans* ATCC 51888, Uniprot); HYPMCv2_0800/ HYPMCv2_3613 (*Hyphomicrobium* sp. MC1, Uniprot); Mext_0099 (*M. extorquens* PA1), Mfla_1451 (*Methylobacillus flagellatus* KT, Uniprot); Mnod_2344 (*Methylobacterium nodulans* strain ORS2060, Uniprot); Msil_2260 / Q8KMI4_METSI (*Methylocella silvestris* BL2, Uniprot); W911_13290 (*Hyphomicrobium nitrativorans* NL23, Uniprot).

3. Illumina sequencing

The Illumina technology relies on sequencing by synthesis (SBS) and occurs on a solid phase, called a flow-cell, where short DNA sequences are anchored (Figure 2.8). The DNA fragments are immobilized on the flow cell surface thanks to the ligation of a DNA sequence complementary to the fixed short sequences. A flow-cell has the dimension of a microscope slide (26 x 76 mm) with 8 lanes. It is designed to facilitate access to enzymes. Densities on the order of ten million single-molecule clusters per square centimeter are achieved and the tens of millions of clusters on the flow cell surface are sequenced in parallel. Solid-phase amplification creates up to 1,000 identical copies of each single template molecule in close proximity (diameter of one micron or less). The sequencing is based on the using of four fluorescent tagged reversible terminator-bound dNTP (A, T, C, G). During each sequencing cycle, a single labeled dNTP is added to the nucleic acid chain. After each dNTP incorporation, the fluorescent dye is imaged to identify the base added and then enzymatically cleaved to allow incorporation of the next labeled nucleotide. Since all four labeled dNTPs are present as single, separate molecules, natural competition minimizes incorporation bias. Base calls are made directly from signal intensity measurements during each cycle, which greatly reduces raw error rates compared to other next generation sequencing technologies. The end result is highly accurate base-by-base sequencing that eliminates sequence-context specific errors, enabling robust base calling across the genome, including repetitive sequence regions and within homopolymers. Fragments in a range size of 150-400 bp can be sequenced using the Illumina technology.

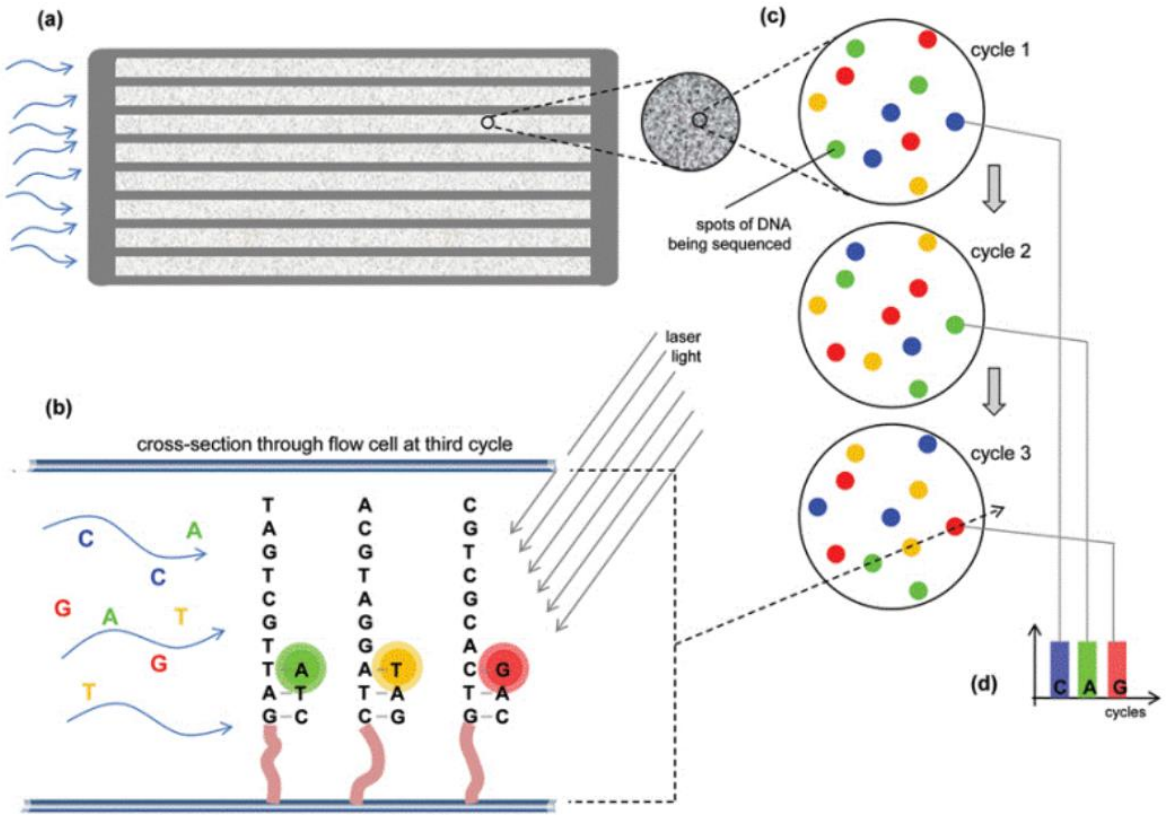


Figure 2.8. Schematic representation of Illumina 'Genome Analyser' flow cell for high throughput sequencing
(a) The flow cell contained many capillaries, where millions of DNA fragments are anchored for sequencing. (b) Cross section of flow cell showing second strand synthesis. Laser light activates base-specific emission color from modified nucleotides incorporated at the end of each cycle. (c) Imaging the surface of the flow cell after each cycle records the base added at each 'spot' of DNA. (d) Image analysis 'reads out' the sequence of incorporated second strand bases at each spot (from Mike Gilchrist and was first published in the 2010 *Mill Hill Essays,* National Institute for Medical Research, http://www.historyofnimr.org.uk/mill-hill-essays/essays-yearly-volumes/2010-2/bringing-it-all-back-home-next-generation-sequencing-technology-and-you/).

The data generated by HiSeq2000 and MiSeq Illumina platforms are similar, but differ in read lengths, scale and therefore support different applications. For large projects where time is less of an issue but cost per sequence is a major concern, the HiSeq platform allows massively parallel sequencing at the lowest cost. Comparable data can be generated on the MiSeq for smaller projects where it is important to process samples quickly, for example, in environmental microbial community assessment (Caporaso *et al*., 2012; Schmidt *et al*., 2013). The price per sequence decreases on Illumina HiSeq2000 and MiSeq platforms, democratizing the sequencing (Tringe and Hugenholtz, 2008). Multiplexing is also used to reduce sequencing

costs. The incorporation of a barcoded nucleotide sequence specific to each sample was performed for both the RNAseq and the SIP studies (see Chapters 3 and 4), which are specifically recognized by the sequencer (Figure 2.8, Shearer *et al.,* 2012).

4. Bioinformatics

4.1. MicroScope platform

The MicroScope platform was developed by Génoscope (Evry, France). It provides an access to easy-use-tools to study, annote, and perform comparative genomic analysis of bacterial genomes (Vallenet *et al*., 2013). This database relies on three major components; (i) the PKGDB (Prokaryotic Genome DataBase), used to access, store and modify information on organisms, sequences and genomic objects, (ii) a panel of bioinformatic tools linked to the PKGDB database to provide results of syntactic and functional annotation pipeline as metabolic pathway tools (KEGG, MicroCyc) and analysis for each genome; (iii), a graphic interface, MaGe (Magnifying Genome) allowing manual annotations and comparative genomic analysis (Vannellet *et al*., 2006 and 2009). The MaGe database interface combines graphical interfaces and the analysis of gene conservation (synteny), together with genome data sources and metabolic pathway tools to assist in data evaluation in order to manually assign the best possible annotation to a given gene product (Figure 2.9) (Vannellet *et al*., 2006; Vannellet *et al*., 2009). It offers useful tools to explore genome data according to the specifications and requests of the annotator. MicroScope can be used as an open or restricted access resource.

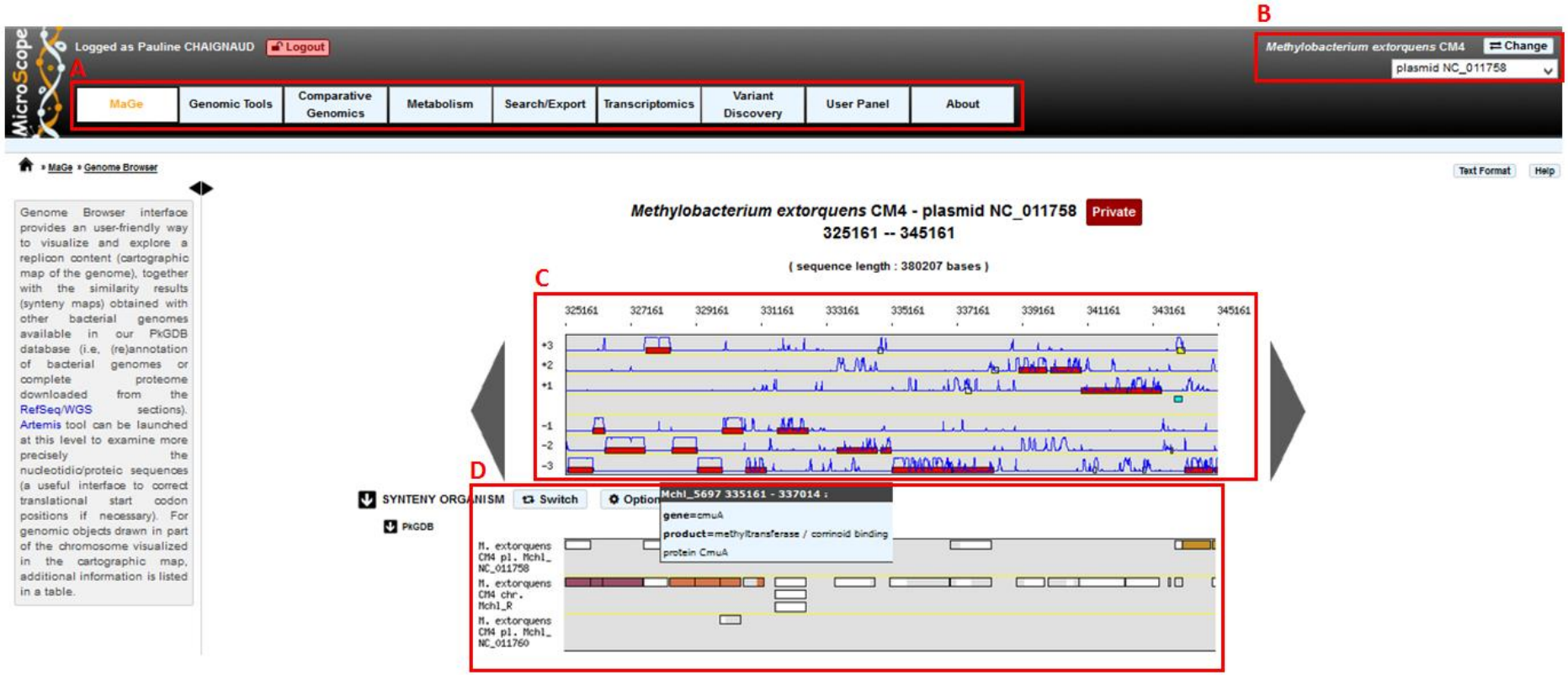


Figure 2.9. A screen capture of the "MicroScope" web platform

The screenshot was taken during analysis of the genome of *M. extorquens* CM4 in the "Genome Browser" window. Several tools (A) can be used to explore the genome of the selected (pivot) organism (B). (C) Genomic objects (predicted genes in the 6 frames) are displayed as rectangles, along with (D) genes in synteny in other organisms in the MicroScope database (PkGBD, http://www.genoscope.cns.fr/agc/microscope) and publically available genome sequences (Refseq).

4.2. Transcriptome Analyses based on MAssive sequencing of RnAs: RNAseq database

TAMARA is the acronym for "Transcriptome Analyses based on Massive sequencing of RnAs" included in the MicroScope platform, which was developed by the Génoscope team at Evry (France). TAMARA pipeline enables to preprocess raw sequencing reads, map reads on reference genomes implemented in the microscope platform, compute transcript coverage along genome and expression levels for annotated genes, and evaluates differential expression between biological tested samples (https://www.genoscope.cns.fr/agc/microscope/transcriptomic/NGSProjectRNAseq.php ?projType=RNAseq&wwwpkgdb=8da0d938cc9a2c3d80982c6a2e87f0cf). It is a user-friendly tool for RNAseq database analysis. This platform stores data from the DESeq pipeline. Of the different methods available to analyze differentially-expressed genes, DESeq is one of the most effective (Qin *et al.*, 2016). TAMARA platform converts data from DESeq and make them accessible for users and connect them with data from annotated genomes available on the MicroScope database. This facilitates data treatment and their modifications, making the analysis easy to do (Table 2.4).

Table 2.4. TAMARA project DehaloRNAseq: sample name reference

Strain	Group	Carbon source	Name of independent biological samples
CM4	C1	CH_3Cl	CM4 C2-2 CM4 C3-1
	C0	CH_3OH	CM4-4 CM4 M4-1
DM4	D2	CH_2Cl_2	C3-2C C3-3B
	D0	CH_3OH	DM4-3 DM4 M2-3

The TAMARA interface gives an access to a large set of tools:

- the sequence data overview (Figure 2.10)
- the raw read count (Figure 2.11)
- a differential analysis (Figure 2.12)

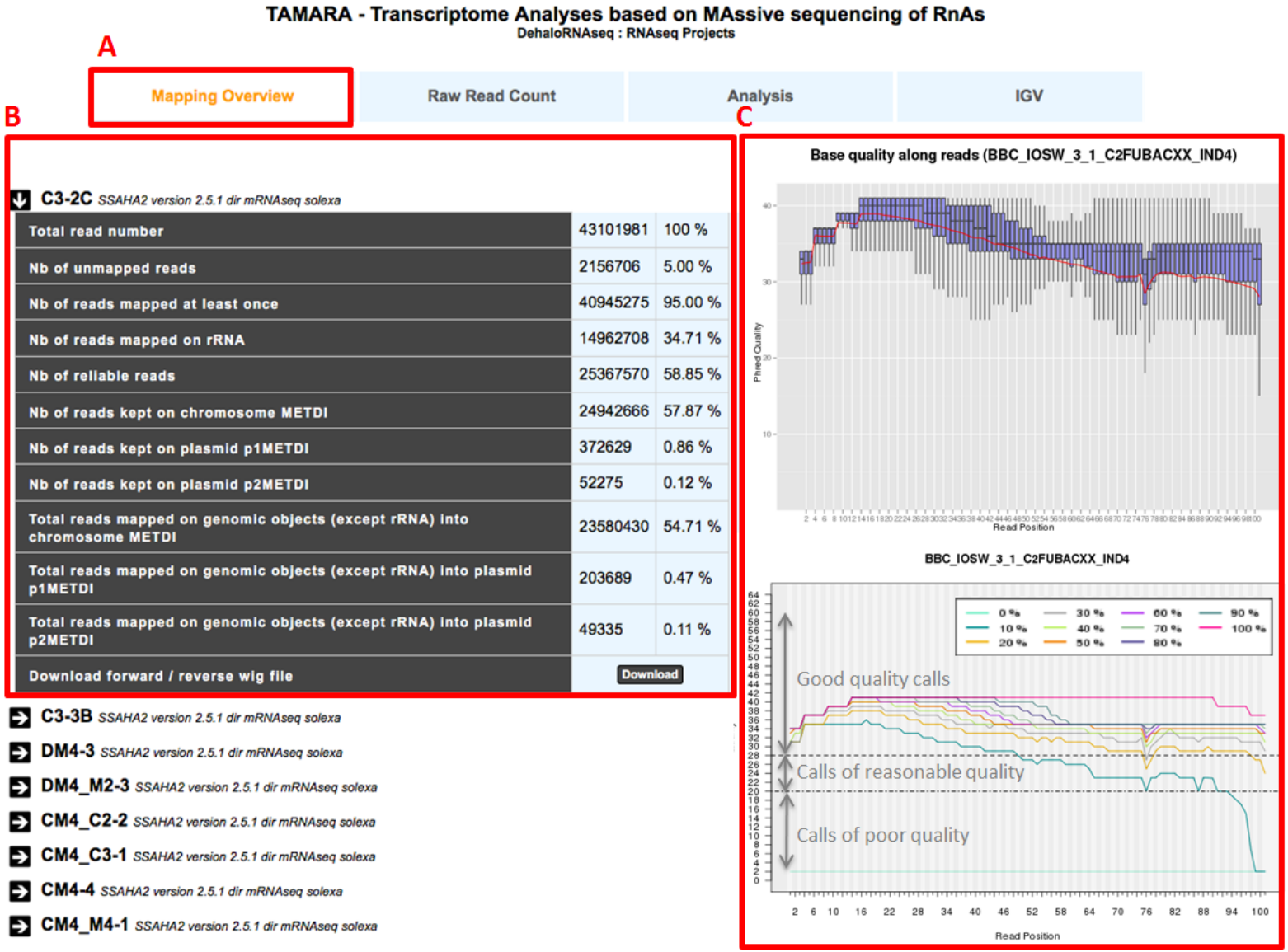


Figure 2.10. A screen capture of the "mapping overview" data available in "TAMARA" web platform

The screenshot was taken during analysis of the genome of *M. extorquens* CM4 in the "Transcriptomics" window. (A) "Mapping overview" gives access to information of mapped sequence data overview. (B) Summary table with total number of reads, unmapped reads, mapped reads on *M. extorquens* CM4 chromosome or on its two plasmids. Reads mapped at least once correspond to the number of reads used for the RNAseq analysis. (C) The quality of calls decreases as the run progresses. In the diagram called "Base quality along reads", the purple boxes represent the inter-quartile rage (25-75 %), the upper and lower whiskers represent the 10 % and 90 % points, the red line shows the mean value.

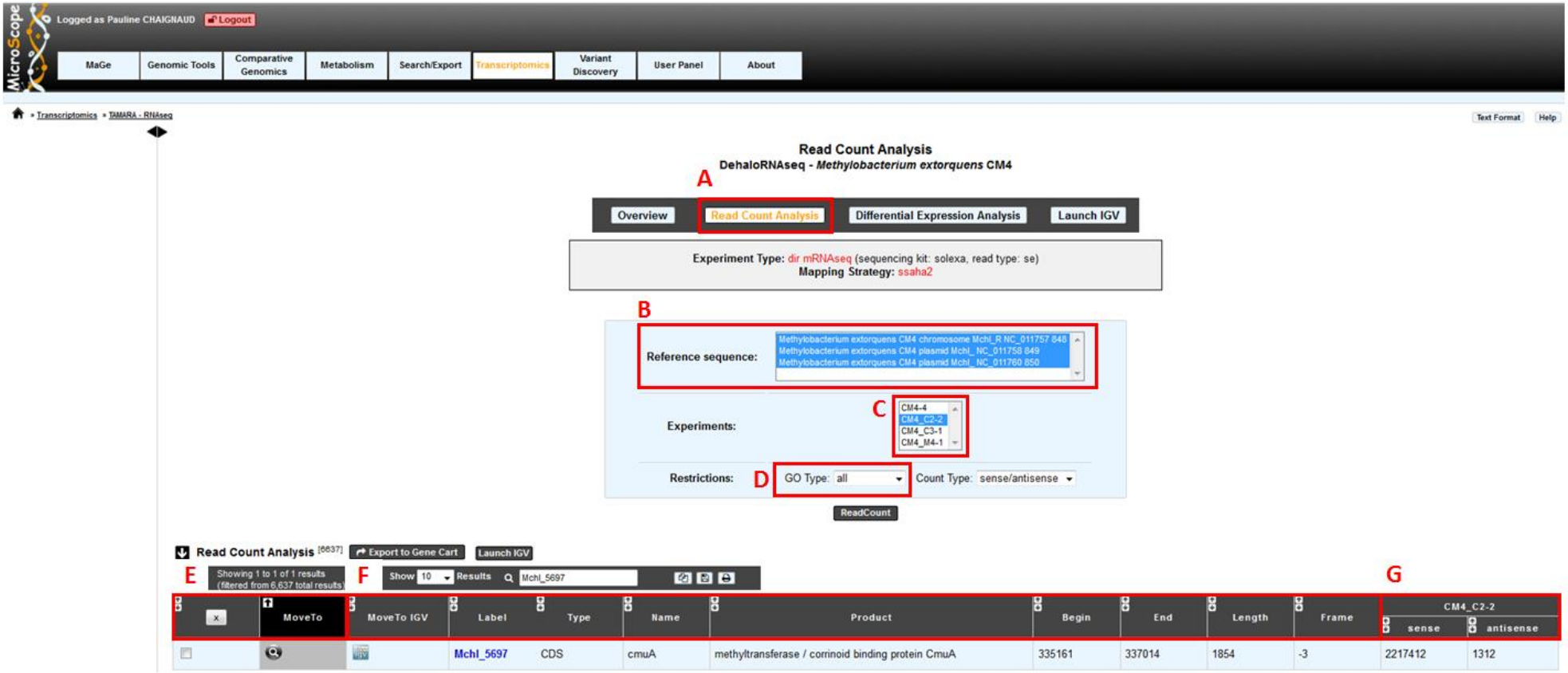


Figure 2.11. A screen capture of the "Raw read count" data available in TAMARA web platform

The "Raw read count" interface gives access to normalized read numbers matched in a given genomic object. Results are accessible following a 5 steps process which is described below. (B) Choose one or several reference sequences. (C) Select at least one experiment and compute the associated read count number per genomic object. (D) It is possible to restrict the query to one or several given classes of genomic objects (CDS, fCDS, rRNA, tRNA, miscRNA or all). (E) Export functions. This section allows users to make all genes (or subsets of genes) available for other analysis tools. (F) Information about genes (label, name, product, begin and end, length and frame). (G) Sense and antisens read numbers.

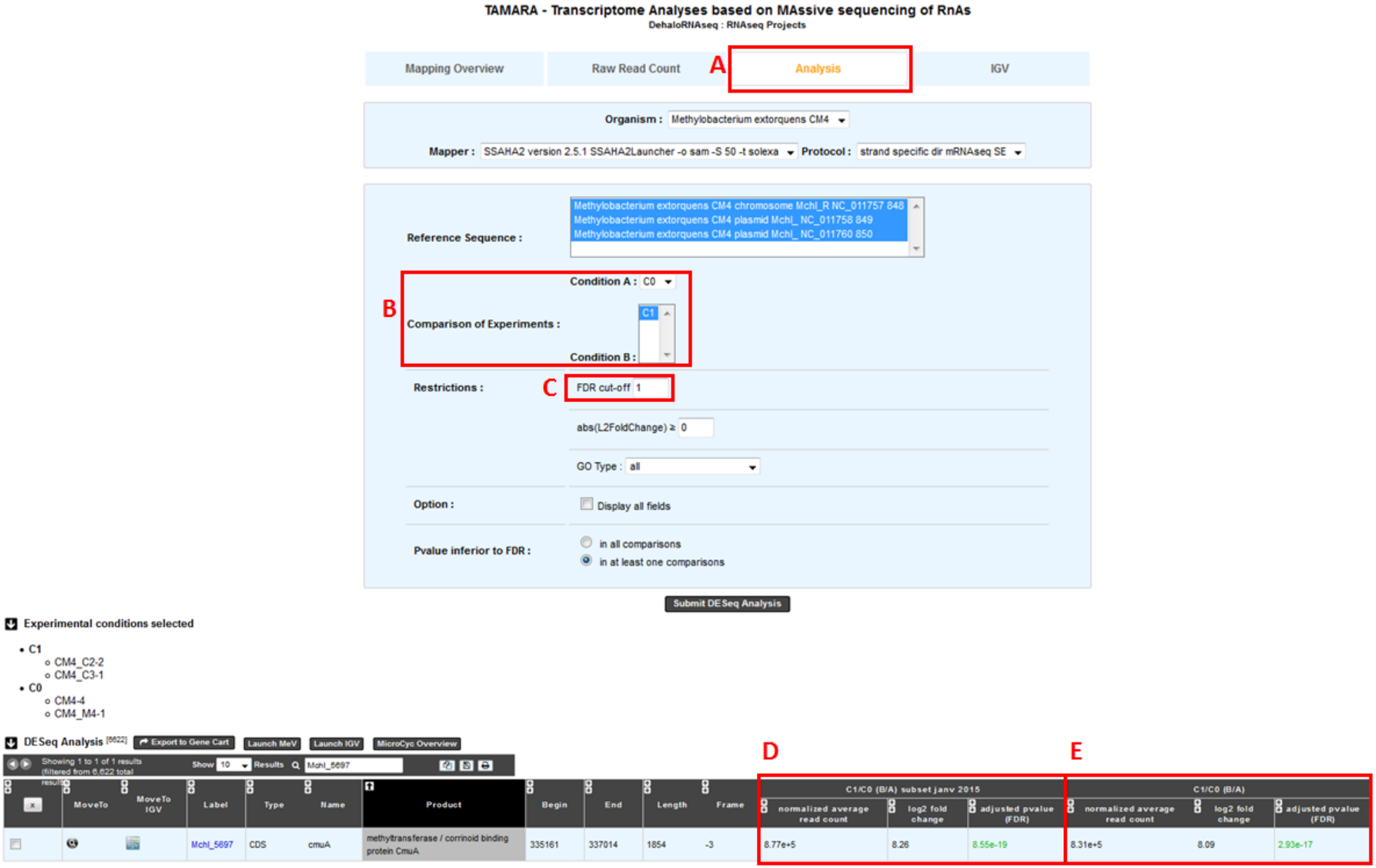


Figure 2.12. A screen capture of the "Analysis" interface available in TAMARA web platform

The "Analysis" interface give access to the differential gene expressions. (B) Selection of conditions for the differential analysis. (C) FDR (false discovery rate) threshold. Data normalized using (D) 55 reference genes or (E) the complete set of CDS.

Chapitre 3

RNAseq study of chlorinated compound utilization in *M. extorquens*

Chapter 3. RNAseq study of chlorinated compound utilization in *M. extorquens*

1. Introduction

Advanced sequencing technologies now provide the opportunity to improve our understanding of the structures and regulatory roles of functional elements involved in adaptation to chloromethane. This chapter describes the first chloromethane transcriptome. To better target the inventory of genes that are specifically expressed in growth with chloromethane, a differential RNAseq approach was performed with two *M. extorquens* reference strains cultivated in methylotrophic conditions. The chloromethane-degrading reference strain *M. extorquens* CM4 was grown with chloromethane versus methanol while the closely-related dichloromethane reference strain *M. extorquens* DM4 was grown with dichloromethane versus methanol. Chapter 3 consists with an introduction, a submitted article of the comparative dehalogenation transcriptomic study, additional data, and the Chapter ends with a brief conclusion.

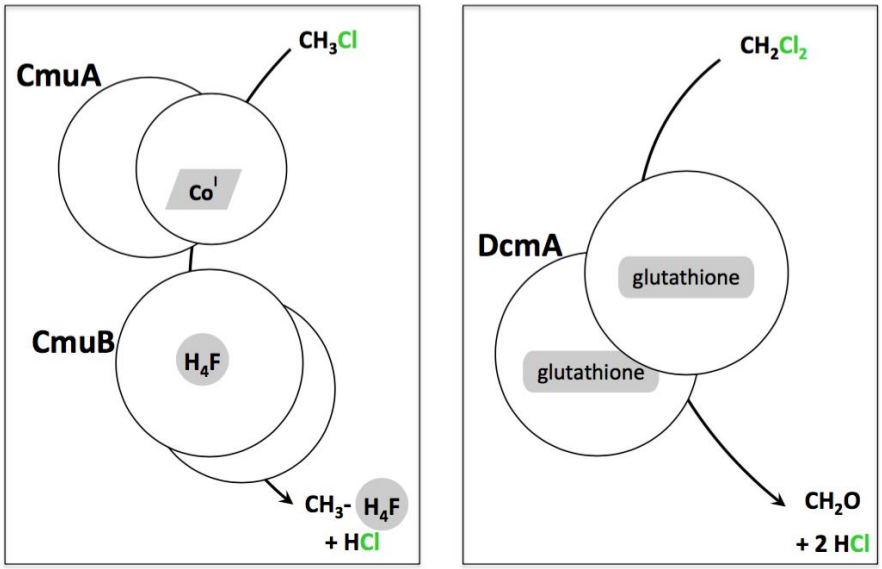


Figure 3.1. Aerobic C_1 dehalogenases
Numerous families of dehalogenase enzymes are known, totalling more than 200 annotation types in sequence databases (David *et al*., 2008). Diverse mechanistic types (hydrolytic or thiolytic, dehydrodehalogenative, involving methyl transfer; reductive, oxidative and/or respiratory; see e.g. (Janssen *et al*., 2005) are known, and many representatives of each mechanistic type have been investigated in detail in the laboratory. Aerobic microbial pathways for utilization of chlorinated methanes have been characterized in the chloromethane and the dichloromethane-degrading strains *M. extorquens* CM4 and DM4, respectively. In *M. extorquens* CM4, two methyltransferases encoded by *cmuA* and *cmuB* are

essential for growth with chloromethane, and require two cofactors, a cobalt-containing corrinoid compound and tetrahydrofolate (H_4F) (Vannelli *et al*., 1999). In *M. extorquens* DM4, the dichloromethane dehalogenase encoded by the *dcmA* gene is essential for dichloromethane utilization (Vuilleumier *et al*., 2009b). DcmA belongs to the glutathione-S-transferase family and uses glutathione as a cofactor. For each C_1 dehalogenase, cofactors are highlighted in grey colour. In both cases, the product of aerobic chlorinated methane dehalogenation leads to intracellular HCl production. A major difference between chloromethane and dichloromethane utilization is that the carbon of the chloromethane is funnelled into a specific metabolic pathway lacking formaldehyde (Kohler-Staub and Leisinger, 1985; Studer *et al*., 2002; Vannelli *et al*., 1999).

The aim of the intraspecies comparative transcriptomics was to explore RNA regulatory elements with the identification of transcript abundance patterns specific of:

- **The chloromethane response** that involves a corrinoid compound and tetrahydrofolate (H_4F)-dependent methylotrophic pathway (only detected in the chloromethane-grown CM4 strain). The chloromethane utilization (*cmu*) pathway is initiated by expression of the chloromethane dehalogenase that differs from the dichloromethane dehalogenase (Figure 3.1). In the chloromethane-degrading strain, some atypical, surnumerary H_4F-dependent enzymes not found in other *Methylobacterium* strains have been detected (gene *metF2*; *folC2*; *purU*; *folD* for instance), but their function or regulation has not been elucidated (Roselli *et al*., 2013). Among those genes, some genes, e.g. *metF2* and *folC2* may have functional redundancy with the “core genome” gene copy. Moreover, redundancy of genes coding for transport and biosynthesis of the chloromethane dehalogenase corrinoid cofactor have also been detected in strain CM4. Among those, expression of *cob* genes adjacent to *cmu* genes may be required for growth with chloromethane but whether these genes are transcribed has not been previously tested (Roselli *et al*., 2013). By comparing the chloromethane transcriptome to the methanol transcriptome in strain CM4, we expected to reveal RNA dependent regulatory mechanisms specifically involved in expression of the *cmu* H_4F-dependent methylotrophic pathway, as opposed to genes involved in standard formaldehyde and the tetrahydromethanopterin-dependent methyltrophic pathways (see Chapter 1, Figure 1.11).
- **The halogenated C_1 methylotrophy and dehalogenation-linked adaptation** such as intracellular chloride and acid stress (common to the chloromethane-grown CM4 strain and the dichloromethane-grown DM4 strain).

The transcriptomic data are discussed in view of our model for bacterial adaptation to the transformation of chlorinated methanes (Figure 3.2). By comparing transcriptomics of cultures grown with methanol to those grown with chlorinated methanes, we identified core and variable genome genes with transcription patterns modulated by only chloromethane, dichloromethane and methanol, or shared by both chlorinated methanes. To our knowledge, this is the first study that has experimentally identified the carbon source-dependent core genome contribution in methylotrophic conditions of chlorinated compounds in natural isolates of two closely related strains of the species *M. extorquens*. Those results will lead to a better understanding of key enzymes involved in bacterial degradation of chlorinated methanes and the regulation of their expression will provide molecular ecological tools to evaluate biodegradation in ecosystems contaminated by such compounds.

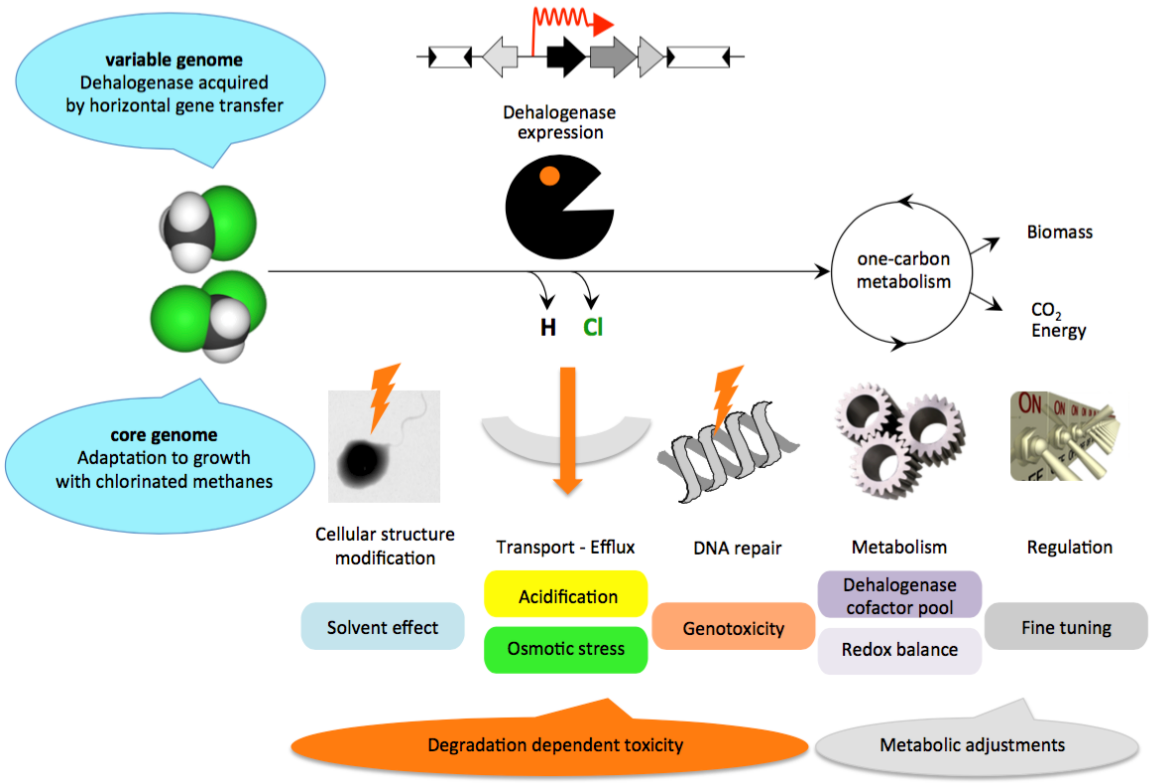


Figure 3.2. Adaptation to growth with chlorinated methanes as the sole carbon and energy source

This figure proposes a model for adaptation to growth with chlorinated methanes and has been adapted from Muller and collaborators (2011). Previous genetic, comparative genomics, proteomics and random mutagenesis studies have demonstrated that cultures grown with these compounds elicit a complex adaptive response encoded by the core genome, which is shared by both dechlorinating and non-dechlorinating *Methylobacterium* strains. It requires the expression of dedicated essential dehalogenase enzymes encoded by strain-specific dehalogenation gene modules. The variable genome includes strain-specific pathways of chloromethane and dichloromethane that are initiated by the expression of dehalogenation modules that respectively encode the *cmu* genes located on the 380 kb pCMU01 plasmid in strain CM4 and the *dcm* genes nested within a genomic island in strain DM4 (Marx *et al*, 2012; Vuilleumier *et al*., 2009). Both pathways produce intracellular HCl that lowers pH and increases inorganic chloride. Moreover, the microbial response to such pollutants can affect biological function at different levels (e.g. membrane structure, inactivation of key molecules, through genotoxic effects, and may involve metabolic adjustments) (Muller *et al*., 2011a; Roselli *et al*., 2013; Vuilleumier, 2002).

2. Article

Contribution of the Core and Variable Genome to the Transcriptomes of two *Methylobacterium extorquens* Strains Grown with Chloromethane, Dichloromethane or Methanol

Pauline Chaignaud[1,2], Marion Weiman[3], Adriana Alberti[3], Stéphane Vuilleumier[1], Steffen Kolb[4],[2] Stéphane Cruveiller[3], Françoise Bringel[1*]

Affiliations:

[1], CNRS Molecular Genetics, Genomics, Microbiology, Université de Strasbourg UMR 7156 UNISTRA CNRS, Strasbourg, France

[2], Department of Ecological Microbiology, University of Bayreuth, Bayreuth, Germany

[3], CNRS-UMR 8030 and Commissariat à l'Energie Atomique et aux Energies Alternatives CEA/DSV/IG/Genoscope LABGeM, Evry, France.

[4], Institute for Landscape Biogeochemistry - Leibniz Centre for Agricultural Landscape Research - ZALF, Müncheberg, Germany

*Corresponding author: francoise.bringel@unistra.fr; Phone: + 33 (0)3 68 85 18 15; Fax: + 33 (0)3 68 85 20 28

Address: 28 rue Goethe, 67083 Strasbourg CEDEX, France

Running title: Transcriptomics of chlorinated C_1 compound utilization

ABSTRACT. The transcriptomes for the utilization of methanol, chloromethane or dichloromethane as sole carbon and energy source were compared in two closely related *Methylobacterium extorquens* strains. Both strains grow with methanol, but strain CM4 grows on chloromethane while strain DM4 grows on dichloromethane. Both chlorinated methane utilization pathways produce intracellular HCl that lowers pH and increases inorganic chloride. Only 11 out of 4,620 common genes had higher transcript abundance in cultures grown with chlorinated methanes compared to methanol. These genes included stress response and related bioenergetics genes (*depP*, *hppA*). On the other hand, the variable genome (25 %) which includes the dehalogenation modules that encode the *cmu* genes in strain CM4 and the *dcm* genes in strain DM4 harbored the majority of the genes whose transcription was enhanced by chloromethane on plasmid pCMU01 such as the *cmu* and coenzyme-associated *cmu*-pathway genes for vitamin B_{12} and for H_4F. In the variable genome of strain DM4, the *dcm* genes nested within a genomic island (GEI) were enhanced in cultures grown with dichloromethane, whereas central metabolism genes (*ackA*, *adh*-like, *xfp*, cytochrome-associated genes) located on redundant GEIs had increased transcript abundance in cultures grown with methanol. A major finding is that core and variable genome-encoded genes involved in redox status adjustments were identified for growth with different C_1 compounds. Transcription responses to growth with chloromethane and dichloromethane were predominantly uncorrelated.

IMPORTANCE. Methylotrophic microorganisms often differ in their ability to use one-carbon compounds due to differential possession of gene modules. To grow with chlorinated one-carbon compounds, a methylotrophic microorganism must acquire a one-carbon dehalogenation module that integrates the existing methylotrophic metabolic and regulatory networks. This prompted us to investigate transcriptional adjustments during growth with different one-carbon growth substrates in the genomic background of *Methylobacterium extorquens*, the best-characterized microbial model for the study of aerobic methylotophy. *M. extorquens* has a broad ability to utilize one-carbon compounds, including the reference C_1 compound methanol. Using intraspecies comparative transcriptomics, we identified key functions encoded by the core genome or by genomic islands that are necessary for growth with different C_1 compounds, toxic natural (chloromethane) and industrial (dichloromethane) chlorinated compounds.

INTRODUCTION

One-carbon compounds such as methanol (CH_3OH), chloromethane (CM, CH_3Cl) and dichloromethane (DCM, CH_2Cl_2) are found in both pristine and polluted environments. In plants, methanol and chloromethane are both produced, although chloromethane emissions are at least twelve orders of magnitude lower than that of methanol (1,2). Chloromethane is released from terrestrial environments as a gas subsequently found in the atmosphere, where it is the most important halogenated trace gas contributing to ozone destruction (2). Dichloromethane, the most produced industrial worldwide-chlorinated solvent, is commonly detected in polluted sites. A few methylotrophic microorganisms, including bacteria and yeast, are able to utilize chlorinated methanes (such as chloromethane and dichloromethane) as their sole carbon and energy source. Methanol is utilized by a majority of methylotrophs, while only a restricted number of methylotrophs are able to degrade chloromethane or dichloromethane (3). Thus, methanol is the reference reduced one-carbon compound. A long history of biochemical and genetic utilization studies has set basis for genome-scale reconstruction of metabolic networks in methylotrophic bacteria (4,5). The genomes of methylotrophic bacteria contain a wide variety of gene-encoding methylotrophic metabolic pathway modules. These modules are unique, redundant, and sometimes interchangeable (6). Such methylotrophic modules were divided into three groups: for primary one-carbon oxidation, formaldehyde dissimilation, and assimilation pathways. In particular, dehalogenation modules initiate halogenated methane catabolism (Fig. 1, green box for chloromethane; purple box for dichloromethane).

FIG 1 Metabolism of chlorinated C_1 compounds chloromethane and dichloromethane and of methanol in *M. extorquens* CM4 and DM4. Chloromethane dehalogenation and methyl transfer to tetrahydrofolate (H_4F) metabolic steps are boxed in green. Dichloromethane dehalogenation and methanol oxidation into formaldehyde are boxed in purple and blue, respectively. Black boxes delimit common methylotrophic oxidation modules: H_4F and tetrahydromethanopterin (H_4MPT)-dependent C_1 transfer reactions, and formate oxidation for C_1 dissimulation. The serine-glyoxylate cycle for C_1 assimilation are schematized with two connected circles. Reactions related to essential co-factors for chlorinated methane utilization are highlighted with colored spots for H_4F (yellow), corrinoid cofactor (green), glutathione (purple) and pyrroquinoline quinone (blue). Genes located on genomic islands (GEI) in *M. extorquens* DM4 or on the plasmid pCMU01 in *M. extorquens* CM4 are shown in grey-shaded boxes. RNAseq data are summarized using color-coding for gene names when transcripts were detected as more abundant on chloromethane (green), dichloromethane (purple), both chloromethane and dichloromethane (dark red), or methanol (both *M. extorquens* CM4 and DM4 (dark blue), or methanol in *M. extorquens* DM4 (sky blue).

Compounds involved in energy balance reactions are highlighted with names written in orange when produced and red when consumed. The proposed figure is based on previous studies in *Methylobacterium* or in other Proteobacteria: cytochromes bd quinol oxidase, cytochrome d terminal oxidase, methanol dehydrogenase (Mxa, Mdh) and formate dehydrogenase (Fdh) complexes (4,7); acetone carboxylase complex (8), the membrane-bound transhydrogenase encoded by the *pnt* complex (9); the membrane-bound proton translocating pyrophosphatase transporter HppA (10); the ClcA H^+/Cl^- antiporter (11); the DctA multi-carbon compound/ H^+ symport-transporter (12,13); *phbC* encoding a poly-(R)-3-hydroxybutyrate (PHB) synthase (14); and *phaABC* encoding a β-ketothiolase, an acetoacetyl-CoA reductase and poly-β-hydroxybutyrate synthase respectively, involved in PHB biosynthesis (15). PHB metabolism has previously been shown to be linked to acetone production (16), to beta-oxidation by 3-hydroxy-butyryl CoA (17) and patatin-like phospholipase (18). Cbi, cobilamide; Cbl, cobalamin; CoA, co-enzyme A; PEP, phosphoenol pyruvate; OXA, oxaloacetate; TCA, tricarboxylic acid cycle.

The dynamics of acquisition, loss, cohabitation and regulation patterns of metabolic pathway-encoding modules are driven by the specific habitat's physico-chemical parameters and genetic repertoire differences. To grow with "alternative" one-carbon compounds, a methylotrophic microorganism must acquire a primary one-carbon oxidation or dehalogenation module that integrates the existing methylotrophic metabolic and regulatory networks. Thus, these bacteria need to adjust their overall genome expression during "alternative" chlorinated compounds utilization driven by the expression of variable genome modules and fine-tuning of functional methylotrophic modules.

While methanol can be assimilated by most methylotrophs, growth with "alternative" one-carbon sources such as chlorinated methanes can be quite challenging (11,19). Chloromethane and dichloromethane utilization starts with a dehalogenation step (Fig. 1), which induces physiological stress due to the production of intracellular HCl, lower pH and higher concentration of inorganic chloride, as well as formation of DNA adducts (19,20). Growth with dichloromethane in experimental evolution studies showed that the horizontal transfer of dehalogenase-encoding modules for dichloromethane dechlorination in naïve methylotrophic bacteria led to post-transfer refinement (11). In addition, previous studies of genome-wide random mutagenesis, differential proteomics studies have shown that adaptation to growth with chlorinated methanes requires genes located on genomic islands (GEI) or plasmids, as well as core genome-encoded housekeeping functions involved in stress response and central metabolic tuning (11,21,22).

The most extensively studied methylotroph is the alpha-proteobacteria *M. extorquens*, a species found in a wide variety of habitats including plants, soil, clouds, and wastewater

(23,24–27). The genome of five *M. extorquens* strains (AM1, BJ001, CM4, DM4 and PA1) has been completely sequenced (28,29). Isolates of *M. extorquens* are able to grow with methanol, but only *M. extorquens* strains CM4 and DM4 are able to utilize chlorinated compounds as their sole source of carbon and energy. *M. extorquens* CM4, isolated from petrochemical factory soil in Russia, grows on chloromethane (30) while *M. extorquens* DM4 isolated from contaminated industrial waste grows on dichloromethane (31). *M. extorquens* genomes are characterized by a high GC content (68 %), large chromosome size (between 5.5 and 5.9 Mb) with plasmid presence, except for strain PA1 (23). In *M. extorquens* CM4, plasmid pCMU01 (380 kb long) harbors two chloromethane dehalogenation modules that initiate the chloromethane-utilization pathway (*cmu*) and require expression of the essential genes *cmuA* and *cmuB* (22,32). The methyl group of chloromethane is transferred to the corrinoid methyltransferase protein CmuA encoded by the *cmuA* gene and then to tetrahydrofolate (H_4F) catalyzed by the methyltransferase CmuB (32). For each mole of chloromethane, one mole of HCl and one of methyl-H_4F are produced. CmuAB protein activity requires two essential cofactors, a corrinoid cofactor and H_4F. Thus, adjacent to the *cmuAB* genes it was not surprising to find the corrinoid compound vitamin B_{12}-associated aerobic biosynthesis and transport, and H_4F-dependent metabolism plasmid genes in addition to chromosomal copies conserved in *M. extorquens* genomes (22,33). In *M. extorquens* DM4, the dichloromethane utilization *dcm* genes are part of a genomic island of 126 kb located on the chromosome where *dcmA* encodes a dichloromethane dehalogenase DcmA of the gluthathione S-transferase family (22,29). The dichloromethane dehalogenase, converts dichloromethane into inorganic chloride and S-chloromethyl glutathion, which is genotoxic, unstable and decomposes spontaneously into glutathione, inorganic chloride and formaldehyde (HCHO) (19,34). For each mole of dichloromethane, two moles of HCl and one mole of formaldehyde are produced.

The *cmu* and *dcm* genes are the only known chlorinated methane degradation modules and have best been characterized in the *M. extorquens* strains CM4 and DM4, respectively (35,36). Several methanol primary oxidation modules with either *mxaFI*, *xoxF* or *mdh2* genes have been identified and may coexist simultaneously in genomes of methylotrophs (37,38). The first one discovered is the *mxa* gene cluster in *M. extorquens* AM1 containing the structural *mxaFI* genes coding for the methanol dehydrogenase (MDH) subunits and its associated cyctochrome c accepting the electrons (7). While gene *mxaF* has been proposed as a gene

marker for detection of methanol oxidation in the environment (39), *xoxF*-related genes are more widespread in metagenomic datasets of environmental samples. XoxF-related genes encode a MDH (methanol dehydrogenase) which shares approximately 50 % amino acid identity (aa Id) with MxaF, while the *mdh2* gene encodes a single-subunit MDH that belongs to a class of pyrroloquinoline quinone (PQQ)-linked dehydrogenases (38,40). For each mole of methanol, one mole of formaldehyde (HCHO) is produced by these dehydrogenases.

The major metabolic difference between chloromethane and dichloromethane utilization is the carbon entry in the methylotrophic metabolism and the link to formate production (Fig. 1). Chloromethane requires a reversal of the assimilatory H_4F pathway since methyl-H_4F must be oxidized into formate to yield energy. The formate produced via the H_4F-depending pathway essential for growth with chloromethane requires the bifunctional enzyme FolD (with methylene-H_4F dehydrogenase and a methenyl-H_4F cyclo-hydrolase activities) associated with *metF2* (coding a methyl-H_4F reductase) and *purU* (coding a formyl-H_4F hydrolase) (41). These enzymes are part of the *cmu* pathway first characterized in *M. extorquens* CM4 (41). In other *M. extorquens* strains and most methylotrophs, formaldehyde is the product of one-carbon compound primary oxidization when grown with methanol (42). *M. extorquens* DM4 grown with dichloromethane forms formaldehyde that is oxidized into formate via the tetrahydrometanopterine (H_4MPT) pathway. To produce biomass via the serine cycle, formate is reduced into methylene-tetrahydrofolate (CH_2=H_4F) (43) while to produce energy, formate is oxidized to CO_2 by formate dehydrogenase (FDH). *M. extorquens* core genomes harbor four characterized FDH (44,45).

In this study, we searched for phenotypic expression profiles of common or specific genes by comparative transcriptomics of phylogenetically very-closely related strains after growth under methylotrophic conditions with different one-carbon compounds. *M. extorquens* strains CM4 and DM4, isolated from soil on the basis of their ability to grow with chloromethane and dichloromethane respectively, acquired different chlorinated one-carbon dehalogenation modules (22,29,46). In this study, the core and variable genome content was analyzed using intraspecies comparative genomics of 5 sequenced genomes of *M. extorquens* strains with respect to their capacity to degrade chlorinated methanes. By comparing transcriptomics of cultures grown with methanol to those grown with chlorinated methanes, we identified core and variable genome genes with transcription patterns modulated by only methanol, chloromethane, dichloromethane or shared by both chlorinated methanes. To our

knowledge, this is the first study that has experimentally identified the carbon source-dependent core genome contribution in methylotrophic conditions of chlorinated compounds in natural isolates of two closely related strains of the species *M. extorquens*.

RESULTS

Deep sequencing was performed on cDNA prepared from samples of partially rRNA-depleted RNA extracts from methyltrophically-grown cultures with different one-carbon compounds (Table 1). Four transcriptomes were obtained, two for each of the two *M. extorquens* strains tested. For *M. extorquens* CM4, chloromethane or methanol was used as the sole source of carbon and energy. For *M. extorquens,* DM4, dichloromethane or methanol was used. Both *M. extorquens* CM4 and DM4 strains grew with a similar generation time (3.2 h) on methanol, which served as the reference growth condition. Longer generation times were needed for chloromethane (5.4 h) and even longer for dichloromethane (9.0 h), in agreement with earlier studies (21,22). Criteria for gene expression profiling included differential expression profiles calculated as log2 fold-change (log2fc) of normalized read numbers in cultures grown with either one of the two chlorinated methanes than with methanol, the reference non-chlorinated one carbon compound.

TABLE 1 Overview of *M. extorquens* CM4 and DM4 growth, genomic properties and transcriptomic data

Strain	Genome[a]	Carbon source (generation time in h)[b]	Total reads[c]	Mapped reads (%)	rRNA[d] (%)
CM4	chromosome 5.8 Mb, GC % = 68.2 pCMU01: 380.2 kb, GC % = 66.3 p2MCHL: 22.6 kb, GC % = 63.9	methanol (3.0 ± 0.2)	30,337,270	98.6	18.3
			37,414,003	98.2	14.3
		chloromethane (5.4 ± 0.4)	53,790,411	96.3	18.6
			36,665,352	97.5	25.4
DM4	chromosome 5.9 Mb, GC % = 68.1 p1METDI 141.5 kb, GC % = 65.3 p2METDI: 38.6 kb, GC % = 63.7	methanol (3.4 ± 0.4)	48,154,448	99.1	12.3
			38,757,418	97.9	23.0
		dichloromethane (9.0 ± 0.7)	43,101,981	95.0	34.7
			32,066,920	95.0	20.0

[a] GenBank accession n° for CM4, CP001298, CP001299, CP001300; for DM4, FP103042, FP103043, FP103044
[b] Aerobic growth in M3 media in presence of 10 mM one-carbon compound provided as the sole source of carbon and energy
[c] High-throughput sequencing with HiSeq2000 from Illumina. Coverage 600X
[d] Percentage of total read number (mapped and unmapped)

Extensive gene synteny between strains CM4 and DM4. *M. extorquens* CM4 and DM4 genomes share a large majority of their genes (75 %, 4,620 CDS) when defining common genes as encoding proteins with at least 80 % amino acid sequence identity on 80 % of their length. Their gene organization along the genome is conserved overall in large synteny segments (Fig. 2).

FIG 2 Overview of genomic islands and gene synteny in genomes of *M. extorquens* CM4 and DM4. Strand conservation (in purple) and strand inversion (in blue) lines representing a conserved group of at least 6 genes were drawn using the Conserved Synteny LinePlot on the MaGe plateform. One-carbon primary oxidation genomic islands (GEI) are described in Table 4: for dichloromethane (*dcm* genes on 126 kb genomic island (29), for methylamine (*mau* genes on 70 kb genomic island (29), and for chloromethane (*cmu* genes on the 380 kb plasmid pCMU01 (22), as well as other GEIs.

The genomes of three non-dechlorinating strains (AM1, BJ001, and PA1) affiliated to *M. extorquens* were then compared to those of the chloromethane-degrading strain CM4 and the dichloromethane-degrading strain DM4 (28,29). The assembly of the gene content found in five *M. extorquens* genomes resulted in a pan genome of 12,273 unique CDS for a total of 12.3 Mb. The core genome included 3,489-shared CDS (Fig. 3). The variable genome specific to strain CM4 and potentially specific to growth with chloromethane included 1,512 CDS. The variable genome specific to strain DM4 and potentially specific to growth with dichloromethane was smaller (952 CDS). The variable genome shared only by dechlorinating strains and potentially specific to growth with chlorinated methanes included 163 CDS. Of those, 85 genes were detected on two plasmids (pCMU01 in CM4; p1METDI in DM4) with no predicted function besides plasmid-related functions.

FIG 3 Core, variable and specific genome of five strains affiliated to the species *Methylobacterium extorquens*. Comparative analyses were performed using the fully sequenced genomes of *M. extorquens* strain AM1 (GenBank accession n°CP001510, CP001511, CP001512, CP
001513 and CP001514), strain BJ001 (CP001029, CP001030 and CP001031), strain CM4 (CP001298, CP001299 and CP001300), strain DM4 (FP103042, FP103043 and FP103044) and strain PA1 (CP000908) (28,29). The pan genome consists 12,274 genes. Homologous genes encode for proteins with at least 80 % of amino acid identity on 80 % of the CDS length. Genome size of the common (core), variable (found in two up to four genomes) and specific (only found in one strain) are indicated in the Venn diagram obtained with the MaGe platform interface pan/core-genome.

Methanol-enhanced expression of a core genome C_1/C_2 dehydrogenase *mdh2*-associated gene module. As expected, the overall transcription patterns are conserved in the two strain cultures grown with methanol as the sole carbon and energy source (Fig. S1A). Among those genes with unchanged transcript abundance patterns, we found genes encoding components of the well-characterized pyrroloquinoline quinone (PQQ)–dependent methanol-

dehydrogenase MxaFI (27) (Fig. S2). A minority of the core genome-encoded genes showed significantly higher transcript abundance in cultures grown with methanol compared to chlorinated methanes (9 genes group E in Table S2; log2fc ≤ -2.1, p-value <0.029). One of these, a highly transcribed gene, encodes a C_4-dicarboxylate transport (Dct) protein that displays 77 % aa identity with the characterized *Rhizobium* DctA transporter (47). Although Dct protein transport specificity has not been studied in *M. extorquens*, in other aerobic bacteria, Dct transports succinate, malate, fumarate, aspartate and a pyrimidine precursor orotate, from the periplasm across the inner membrane (12). We also found that a large majority of the methanol-enhanced transcript abundant genes (7 out of 9) were part of a single primary C_1/C_2 carbon compound oxidizing cluster. This cluster harbors the highly transcribed PQQ-dependent methanol/ethanol dehydrogenase-encoding gene *mdh2* which is involved in methanol utilization in other methylotrophs (40).

Chlorinated methane-enhanced expression of core genome genes were involved in stress response and bioenergetics. Utilization of either chloromethane or dichloromethane involves exposure to intracellular acidification and production of chloride, so we decided to focus on identifying a common adaptive response to the use of chlorinated methanes versus methanol. To do so, we searched for core genome-encoded genes with similar transcription patterns in cultures of strains CM4 and DM4. We found that 11 genes, representing 0.02 % of *M. extorquens* genome, had higher transcript abundance in cultures grown with chlorinated methanes compared to methanol (Fig. 4; Table S2 group B; log2fc from 2.0 up to 6.6).

FIG 4 Carbon source-dependent shared categories of gene transcript abundance between *M. extorquens* CM4 and DM4 genomes. The Venn diagram includes a total of 4,620 genes encoding CDS with at least 80 % aa identity on 80 % of the CDS length detected in both *M. extorquens* CM4 and DM4 genomes. Categories were defined on the basis of the log2 fold-change (log2fc) values of RNAseq reads for *M. extorquens* CM4 grown with chloromethane versus methanol, and for *M. extorquens* DM4 grown with dichloromethane versus methanol as follows: (A) constitutive expression (log2fc between 2 and -2); higher transcript abundance (log2fc > 2) in dichloromethane and chloromethane (B), in chloromethane only (C), in dichloromethane only (D); higher transcript abundance (log2fc < -2) in methanol (F), in strain CM4 only (except for the gene Mchl_1717/METDI2103 coding for a conserved protein of unknown function); and in strain DM4 only (G). Details found in Table S2.

Two genes had predicted functions. First, gene *depP* encodes a putative periplasmic serine protease (log2fc value of 6.6). Its corresponding *Escherichia coli* homolog HtrA (39 % aa Id) is a central housekeeping molecular chaperon protein in the bacterial envelope, which under

295 stress including membrane damage, controls the folded state of proteins and the biosynthesis
296 of partially folded outer-membrane proteins (48). Second, gene *hppA* encodes a membrane-
297 bound proton translocating pyrophosphatase. HppA in *Carboxydothermus hydrogenoformans*
298 (48 % aa Id) uses the energy of pyrophosphate hydrolysis as the driving force for proton
299 extrusion to build up a proton motive force (49). In *Rhodospirullum rubrum* (72 % aa Id), this
300 proton pump is involved in stress bioenergetics (10). Membrane-bound proton
301 translocating pyrophosphatases confer resistance to salt stress (50). Thus, *hppA* may also play
302 a role in the adaptive response and bioenergetics of chloromethane and dichloromethane
303 aerobic utilization in *M. extorquens*. The stress-related core genome genes *depP* and *hppA* are
304 located on different genomic regions.
305
306 **The chloromethane transcriptome**. Among 137 genes detected as more abundant with
307 chloromethane than methanol, 40 genes have been previously associated with growth with
308 chloromethane (22) (log2fc ≥ 2; Tables 2 and S3). These genes, classified in the chloromethane
309 utilization group, showed the highest chloromethane-dependent regulation levels (log2fc
310 values 7.7 to 10). These included essential genes for growth with chloromethane (*cmuABC*,
311 *metF2*, *purU*), genes previously found to be part of the *cmu* pathway (*folD*) (41), genes
312 systematically associated with *cmu* gene clusters in various chloromethane-degrading strains
313 (*fmdB*, *hutI*, *paaE*-like) (22,51), and corrinoid cofactor riboswitch regulatory elements (52,53)
314 (Fig. 1; misc _RNA_1 see Table 2).
315
316 **TABLE 2 M. extorquens CM4 transcripts differentially abundant in cultures grown with**
317 **chloromethane compared to methanol**

Label[a]	Name	Product	Occurr-ence[b]	RNAseq data	
				log2fc[c]	Adjusted p-value[d]
Chloromethane utilization					
Mchl_5694	*hutI*	putative imidazolonepropionase	pCMU01	7.7	< 0.001
Mchl_5696	-	putative regulatory protein FmdB	pCMU01	7.7	< 0.001
Mchl_5697	*cmuA*	methyltransferase/corrinoid binding protein CmuA	pCMU01[f]	8.3	< 0.001
Mchl_5698	*cmuC2*	putative methyltransferase, CmuC-like protein	pCMU01	8.4	< 0.001
Mchl_5699	*purU*	formyltetrahydrofolate hydrolase	pCMU01[f]	10.0	< 0.001
Mchl_5700	*folD*	bifunctional methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase	pCMU01	8.6	< 0.001
Mchl_5717	-	putative oxidoreductase FAD/NAD(P)-binding domain protein, PaaE-like	pCMU01[e]	8.1	< 0.001
Mchl_5726	*metF2*	methylenetetrahydrofolate reductase	pCMU01	9.6	< 0.001
Mchl_5727	*cmuB*	methylcobalamin:H4folate methyltransferase CmuB	pCMU01[e]	9.5	< 0.001
Mchl_5728	*cmuC*	putative methyltransferase CmuC	pCMU01	9.1	< 0.001

Cobalamin and precursor metabolism and transport					
Mchl_1718	-	putative cobalt transporter, subunit CbtA	core	2.5	< 0.001
Mchl_1719	-	putative cobalt transporter subunit CbtB	core	2.7	< 0.001
Mchl_2855	-	putative TonB-dependent siderophore receptor	core	3.4	< 0.001
Mchl_5676	-	putative cobalamin outer membrane transporter (BtuB)	pCMU01[e]	6.7	< 0.001
Mchl_5677	-	ABC transporter periplasmic binding component, putative vitamin B12 transporter (BtuF)	pCMU01[e]	6.0	< 0.001
Mchl_5678	-	ABC transporter membrane component, putative vitamin B12 transporter subunit (BtuC)	pCMU01[e]	5.3	< 0.001
Mchl_5679	-	ABC transporter ATP-binding component, putative vitamin B12 transport (BtuD)	pCMU01[e]	5.0	< 0.001
Mchl_5681	-	putative P-loop containing nucleoside triphosphate hydrolase	pCMU01[e]	4.8	< 0.001
Mchl_5682	-	putative TonB-dependent receptor	pCMU01[e]	4.4	< 0.001
Mchl_5685	*cobM2*	precorrin-4 C(11)-methyltransferase	pCMU01[e]	7.0	< 0.001
Mchl_5686	-	putative cobalamin biosynthesis protein CobE	pCMU01[e]	4.8	< 0.001
Mchl_5687	*cobL2*	precorrin-6Y C(5,15)-methyltransferase (decarboxylating)	pCMU01[e]	4.7	< 0.001
Mchl_5689	*cobJ2*	precorrin-3B C(17)-methyltransferase	pCMU01[e]	5.5	< 0.001
Mchl_5690	*cobI2*	precorrin-2 C(20)-methyltransferase	pCMU01[e]	5.7	< 0.001
Mchl_5691	*cobH2*	precorrin-8X methylmutase	pCMU01[e,f]	5.9	< 0.001
Mchl_5692	-	putative cobalamin biosynthesis protein, cobaltochelatase subunit CobN-like domain	pCMU01[e]	7.0	< 0.001
Mchl_5693	-	conserved exported protein of unknown function, CoxB-related protein	pCMU01[e]	8.9	< 0.001
Mchl_5702	*cobU2*	nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyltransferase	pCMU01	6.8	< 0.001
Mchl_5714	*czcB*	RND efflux transporter, putative membrane fusion protein	pCMU01[e]	2.5	< 0.001
Mchl_5715	*czcA2*	RND divalent metal cation efflux transporter membrane component, cobalt-zinc-cadmium resistance protein	pCMU01[e]	2.3	< 0.001
Mchl_5721	*cobP2*	bifunctional adenosylcobinamide kinase and adenosylcobinamide-phosphate guanylyltransferase	pCMU01	6.7	< 0.001
Mchl_5722	*cobO2*	cob(I)yrinic acid a,c-diamide adenosyltransferase	pCMU01	6.5	< 0.001
Mchl_5723	*cobQ2*	cobyric acid synthase	pCMU01	6.2	< 0.001
Mchl_5724	*cobD2*	cobalamin biosynthesis protein	pCMU01	2.3	< 0.001
Mchl_5729	*cbiD*	cobalamin biosynthesis protein, putative cobalt-precorrin-6A synthase	pCMU01	4.3	< 0.001
Mchl_5730	*cobC2*	L-threonine-O-3-phosphate decarboxylase domain	pCMU01	4.9	< 0.001
Mchl_5731	*cobA*	uroporphyrinogen-III C-methyltransferase	pCMU01	5.6	< 0.001
Mchl_5732	*bluB2*	5,6-dimethylbenzimidazole synthase (flavin destructase), putative cob(II)yrinic acid a,c-diamide reductase	pCMU01	5.2	< 0.001
misc_RNA_1	-	cobalamin riboswitch upstream of *btuB* (Mchl_5676)	pCMU01	5.9	< 0.001
Rmisc_RNA_4	-	cobalamin riboswitch upstream of the uncharacterized gene (Mchl_1720)	core	3.1	< 0.001
Tetrahydrofolate metabolism					
Mchl_0812	*gcvP*	glycine cleavage complex protein P, PLP-dependent glycine dehydrogenase	core	4.6	< 0.001
Mchl_0813	*gcvH*	glycine cleavage complex protein H	core	4.5	< 0.001
Mchl_0814	*gcvT*	glycine cleavage complex protein T, aminomethyltransferase tetrahydrofolate-dependent	core[f]	3.0	< 0.001
Mchl_5701	*folC2*	putative folylpolyglutamate synthase and dihydrofolate synthase	pCMU01	7.8	< 0.001
Rmisc_RNA_14 Rmisc_RNA_15	-	glycine riboswitch upstream of *gcvT*	core	2.2	< 0.001
C_1 metabolism					
Mchl_1538	-	c-cytochrome c550 protein, putative PQQ-dependent methanol/ethanol oxidation system	core[e]	-3.9	< 0.001
Mchl_1540	*mdh2*	PQQ-dependent methanol/ethanol dehydrogenase	core[e]	-3.4	< 0.001
Mchl_5683	*gck2*	glycerate kinase	pCMU01	5.1	< 0.001

MCv2_5148	*pqqA*	coenzyme pyrroloquinoline quinone (PQQ) biosynthesis protein A	core	2.2	< 0.001
MCv2_5149	*pqqA*	coenzyme PQQ biosynthesis protein A	core	2.1	< 0.001
Central metabolism					
Mchl_5057	-	putative seryl-tRNA synthetase	variable	2.2	< 0.001
Mchl_5058	*acd*	acyl-CoA dehydrogenase	variable	2.2	< 0.001
Mchl_5060	-	putative monooxygenase, small component, putative flavin: NADH reductase	variable	2.5	< 0.001
Mchl_5066	-	putative homoserine O-succinyltransferase (MetA)	variable	2.0	< 0.001
Mchl_5154	-	aldo/keto reductase family	core	2.1	< 0.001
Mchl_5521	*acxC*	acetone carboxylase gamma subunit	pCMU01	2.5	< 0.001
Mchl_5522	*acxB*	acetone carboxylase alpha subunit	pCMU01	2.1	< 0.001
Mchl_5523	*acxA*	acetone carboxylase beta subunit, putative N-methylhydantoinase A	pCMU01	3.7	< 0.001
Energy					
Mchl_2986	*pntAA*	NAD(P)$^+$ transhydrogenase, subunit alpha part 1	core[e]	2.7	< 0.001
Mchl_2987	*pntAB*	NAD(P)$^+$ transhydrogenase, subunit alpha part 2	core	3.6	< 0.001
Mchl_2988	*pntB*	NAD(P)$^+$ transhydrogenase, subunit beta	core	3.4	< 0.001
Mchl_3408	*hppa*	H($^+$) translocating pyrophosphate synthase	core	2.4	< 0.001
Adaptation to stress and transport					
Mchl_2781	-	putative L,D-transpeptidase catalytic domain (YkuD)	core	2.5	< 0.001
Mchl_3002	-	putative manganese catalase	core[f]	2.1	< 0.001
Mchl_3283	*dctA*	C_4-dicarboxylate transport protein	core	-2.9	< 0.001
Mchl_4687	*ada*	fused DNA-binding transcriptional dual regulator, O6-methylguanine-DNA methyltransferase	core	5.6	< 0.001
Mchl_5016	*ibpA*	small heat shock protein	core	2.8	< 0.001
Mchl_5157	*depP*	periplasmic serine protease (DegP)	core	4.6	< 0.001
Mchl_5405	*clcA2*	putative H($^+$)/Cl($^-$) exchange transporter	pCMU01	5.5	< 0.001
Mchl_5500	-	transporter, major facilitator superfamily	pCMU01[e]	4.4	< 0.001
Mchl_5718	-	putative membrane protein, putative transport protein	pCMU01	7.1	< 0.001
Mchl_5719	-	putative membrane protein, putative transport protein	pCMU01	7.7	< 0.001
Mchl_5720	-	putative protein with methyl-accepting signaling domain	pCMU01	7.2	< 0.001

[a] MaGe annotation (https://www.genoscope.cns.fr/agc/microscope). See complementary list Table S3

[b] Genes labeled as core or variable are respectively found in all or at least 2 genomes of *M. extorquens* strains AM1, BJ001, CM4, DM4 and PA1 (80 % of identity on at least 80 % of the protein length). Genes labeled as pCMU01 are located on the 380 kb plasmid and some have chromosomal gene copies described previously (22).

[c] log2 fold-change of normalized read numbers in cultures grown with chloromethane compared to methanol.

[d] False discovery rate, only adjusted p-value under 0.1 were considered.

[e] Part of gene clusters: Mchl_1533-1540; Mchl_5676-5693 gene; Mchl_5714-5716; Mchl_5499-5500..

[f] Corresponding encoded protein found more abundant in cell protein extracts of chloromethane- versus methanol-grown cultures (22)

In addition, 30 genes were involved in the metabolism and transport of the corrinoid cofactor, which is essential for CmuA dehalogenase and chloromethane utilization (32) (Table 2). Two transport systems related to corrinoid cofactors were identified: *cbtAB* genes encoding a putative cobalt transporter Cbt complex (log2fc around 2.6), and *btu* genes encoding the corrinoid transporter Btu complex formed by 4 proteins (BtuB, BtuC, BtuD, BtuF) associated with TonB (54) with corresponding genes all highly transcribed. A corrinoid cofactor riboswitch (53) was also found with higher transcript abundance in cultures grown with chloromethane compared to methanol upstream of gene *btuB* (misc_RNA_1, Table 2) (log2fc value of 5.9).

The gene *bluB2* had a high log2fc value of 5.2. It encodes for a protein highly similar to an enzyme that cleaves reduced flavin mononucleotide ($FMNH_2$) in an oxygen-dependent reaction to form erythrose 4-phosphate (E_4P) and dimethylbenzimidazole (DMB) (55). These products of the enzymatic reaction catalyzed by BluB2 are precursors of the corrinoid and H_4F cofactor biosynthesis, which are both essential cofactors for the *cmu* pathway (32). In the metabolism of the coenzyme H_4F, gene *folC2* encodes a predicted bifunctional dihydrofolate synthetase/H_4F synthase. Gene *folC2* displayed a tightly regulated chloromethane-dependent transcript abundance (high log2fc value of 7.8). The *gcvPHT* gene cluster, also involved in coenzyme H_4F metabolite interconversion, encodes the glycine cleavage complex which catalyzes the H_4F transformation to methylene-H_4F, and its upstream predicted glycine riboswitchs (53,56). The *gcvPHT* genes had transcripts with higher abundance with chloromethane than methanol (Fig. 4, group C; Table 2).

Other genes not previously linked to growth on chloromethane had higher transcript abundance in cultures grown with chloromethane compared to methanol. These included (i) the *pnt* gene cluster encoding the transhydrogenase PntAA, PntAB and PntB subunits needed in the transfer of reducing equivalent between NAD(H) and NADP(H) to the translocation of protons across the membrane, as demonstrated in *M. extorquens* AM1 (57) (log2fc values ≥ 2.7), (ii) two contiguous *ppqA* genes highly homologous to *M. extorquens* AM1 PqqA precursor of a low molecular weight redox active cofactor of dehydrogenases (23); (iii) a gene encoding an ubiquitous family of L,D-transpeptidase catalytic domain YkuD involved in cell wall composition adjustments (58) (log2fc value of 2.5); (iv) *ada* encoding a bifunctional transcriptional activator/DNA repair enzyme Ada in response to alkylation damage in DNA (51 % aa Id) (log2fc value of 5.6) (59); and (v) *ibpA* encoding a small heat shock protein (sHSP) displaying 57 % aa identity with the characterized *E. coli* IbpA that binds to misfolded proteins and protects enzymes from stressful oxidants (60). In conclusion, genes found with higher transcript abundance in cultures grown with chloromethane were involved in chloromethane dechlorination and associated corrinoid cofactor and H_4F-metabolism, as well as in adaptation, energy and central metabolisms.

Plasmid pCMU01 harbors a higher proportion of genes with chloromethane-enhanced transcription than the chromosome of strain CM4. The 380 kb pCMU01 plasmid harbors 62 genes out of the 137 genes detected with transcript reads more abundant in cultures grown

with chloromethane compared to methanol. Among these, unique genes with no chromosomal homologs were involved in carbon assimilation of the methylotrophic *cmu* pathway and related functions as described in the previous paragraph, and also in ATP-dependent carboxylation of C_2 compounds with the complete set of the *acx* genes highly similar to those characterized in aerobic bacterium *Xanthobacter autotrophicus* acetone carboxylase components AcxABC (>81 % aa Id) (61). A total of 31 duplicated genes displayed 38 to 87 % aa identity with chromosomal homologs (22) and had higher transcript abundance on chloromethane, unlike their corresponding chromosomal copy. A majority of the plasmid-borne homologs were involved in corrinoid cofactor metabolism such as the *cob* genes (Fig. 5A).

FIG 5 Involvement of plasmid pCMU01 in the transcriptional response to chloromethane Genes encoded by plasmid pCMU01 (in grey) or by the chromosome (in black) were grouped according to their log2fc values with log2fc ≥ 2.0 and ≤ -2.0 meaning reads more abundant on chloromethane and methanol, respectively. (A) Comparison of pCMU01- and chromosome-encoded paralogs with indication of their involvement in cofactor metabolisms essential for growth with chloromethane. (B) Ratio calculated by dividing the gene number in defined log2fc range, by the total number of genes located on the chromosome or on the plasmid multiplied by one hundred. *M. extorquens* CM4 chromosome and plasmid pCMU01 encode for a total number of 6261 and 361 genes, respectively.

Other plasmid homologs have been previously associated with the *cmu* pathway (*metF2* and *purU*) (32,35), the serine cycle for carbon assimilation (*gck2*), and for adaptation to the chloride stress (*clc2*) (Fig. 1; Table 2). For instance, the highly transcribed glycerate kinase encoding *gck2* gene seems to be more tightly regulated by chloromethane (log2fc value of 5.1) than the characterized chromosomal copy (log2fc value of - 0.5) (42), although they encode for closely related homologs with 85 % aa Id. Similarly, the H^+/Cl^- antiporter encoded by the plasmid *clcA* gene was highly transcribed on chloromethane (log2fc value of 5.5) unlike the chromosomal copy (68 % aa Id) that is required for adaptation to chloride stress of *M. extorquens* DM4 on dichloromethane (11). Among 98 plasmid-borne genes that had chromosomal homologs in *M. extorquens*, 70 genes were constitutively expressed under the tested conditions, including genes involved in one-carbon metabolism, such as the phosphoserine aminotransferase-encoding *serC2* gene and the PQQ-linked methanol dehydrogenase large subunit-encoding *xoxF2* gene (88 % aa identity with the corresponding chromosome homolog). Altogether, the plasmid pCMU01 had a higher proportion of genes

that are differentially transcribed in cultures grown with chloromethane versus methanol than the chromosomal homologues (Fig. 5B). This highlights the central role of plasmid pCMU01-borne genes in chloromethane utilization.

The dichloromethane transcriptome. The transcripts of 59 genes were found to be more abundant in cultures grown with dichloromethane compared to methanol (log2fc ≥ 2) (Tables 3 and S4). These included all four *dcm* genes of the dichloromethane utilization module (21). Genes *dcmA* and *dcmB* were among the most abundant transcripts (Fig. S2B). The majority of the genes that had increased reads on dichloromethane belonged to the core genome of *M. extorquens* (Fig 4., group D, 29 genes). Among these, the putative glutathione peroxidase (METDI0190, log2fc value of 2.5) is connected to the activity of the DcmA enzyme that catalyzes the glutathione-dependent hydrolysis of dichloromethane. The formate dehydrogenase homologs (*fdh4AB*) had lower read counts in cultures grown with methanol compared to dichloromethane. This concurs with a *fdh4A* mutant phenotype found in *M. extorquens* AM1 that was impaired for methanol utilization with formate accumulation in the medium (45), suggesting that the *fdh4A* homolog may play a role in methanol dissimilation in different *M. extorquens* strains. Sets of uncharacterized molybdenum cofactor-dependent enzyme-encoding genes had increased transcript abundance in cultures grown with dichloromethane compared to methanol. These include the molybdopterin oxidoreductase (METDI2693, log2fc value of 3.8) and three components of a molybdenum-containing oxidoreductase (METDI0091-METDI0093). Their role remains unknown, as is the case with a majority of the core genome encoding genes with higher transcript abundance on dichloromethane (Tables 3 and S4). We found in this study that within 126 kb *dcm* GEI (29) only the four *dcm* genes had enhanced expression on dichloromethane.

TABLE 3 *M. extorquens* DM4 transcripts differentially abundant in cultures grown with dichloromethane compared to methanol

Label[a]	Name	Product	Occurrence[b]	Directional RNAseq data	
				log2fc[c]	Adjusted p-value[d]
DCM utilization					
METDI2655	*dcmR*	transcriptional repressor of DCM dehalogenase	*dcm* GEI[g]	2.7	0.040
METDI2656	*dcmA*	dichloromethane dehalogenase (DCM dehalogenase)	*dcm* GEI[g]	4.3	< 0.001
METDI2657	-	conserved protein of unknown function DcmB	*dcm* GEI	4.0	< 0.001

METDI2658	-	conserved protein of unknown function DcmC	*dcm* GEI	3.5	< 0.001
C_1 metabolism					
METDI2873	*fdh4B*	formate dehydrogenase subunit B	core	-3.2	< 0.001
METDI2874	*fdh4A*	formate dehydrogenase subunit A	core	-4.0	< 0.001
METDI5146	*mxaW*	conserved hypothetical protein	core	2.1	0.002
Central metabolism					
METDI0091	-	oxidoreductase, 2Fe-2S subunit	core[e]	2.5	< 0.001
METDI0092	-	oxidoreductase, molybdenum cofactor binding subunit	core[e]	2.9	< 0.001
METDI0128	*ino*	inositol-3-phosphate synthase	core	2.1	0.003
METDI0129	-	putative dTDP-glucose 4,6-dehydratase	core	2.2	0.002
METDI0190	-	putative glutathione peroxidase	core	2.5	0.002
METDI0283	*adh-like*	putative alcohol dehydrogenase	GEI 160	-3.3	0.029
METDI0286	*xfp*	D-xylulose 5-phosphate/ D-fructose 6-phosphate phosphoketolase	GEI 160	-2.9	0.099
METDI2432	-	putative phosphotransferase	core	2.7	< 0.001
METDI2693	-	putative molybdopterin oxidoreductase	core	3.8	0.008
METDI3569	*arcB*	ornithine cyclodeaminase	core[f]	2.6	< 0.001
METDI4540	*adh-like*	putative alcohol dehydrogenase	GEI 197	-3.7	0.012
METDI4543	*xfp*	D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase	GEI 197	-3.1	0.085
METDI4670	-	putative monooxygenase with ATPase activity	core	2.0	0.003
Energy					
METDI0301	-	cation-transporting ATPase (P-type)	GEI 160	-3.3	0.071
METDI2011	*cydB*	cytochrome d terminal oxidase, polypeptide subunit II	core[e]	-3.4	0.036
METDI3863	*hppa*	H^+ translocating pyrophosphate synthase	core[f]	3.5	0.020
METDI4554	-	cation-transporting ATPase (P-type)	GEI 197	-3.5	0.059
METDI4816	*qxtB*	cytochrome bd-quinol oxidase subunit II	core[e]	-2.7	0.061
Lipid-related metabolism					
METDI0292	-	putative poly-beta-hydroxyalkanoate synthase (PhbC)	GEI 160	-3.1	0.085
METDI0293	*fabI*	enoyl-(acyl-carrier-protein) reductase (NADH)	GEI 160	-3.6	0.046
METDI0294	-	putative phosphate butyryltransferase	GEI 160	-3.6	0.045
METDI0295	-	putative acetate kinase (AckA)	GEI 160	-3.4	<0.001
METDI1770	*ohyA*	oleate hydratase	variable	-3.4	0.044
METDI4905	-	putative patatin-like phospholipase	core	2.5	< 0.001
Nitrogen-related metabolism					
METDI1697	-	putative formate/nitrate transporter	variable	-3.4	<0.001
METDI1754	-	putative cysteine desulfurase (SufS domain)	variable	-2.2	0.094
METDI2867	*glnK*	nitrogen regulatory protein P-II	core[e]	-2.5	0.029
METDI3157	*glnII*	glutamine synthetase, type II	core	-3.2	0.084
METDI5532	*nrtB*	nitrate transport permease protein	core[e]	-2.4	0.001
METDI5533	*nrtA*	nitrate transporter component	core[e]	-2.7	<0.001
Transport and adaptation					
METDI0093	-	putative voltage-dependent anion channel	core[e]	2.7	0.001
METDI0288	*clpB*	protein disaggregation chaperone	GEI 160	-3.6	0.014
METDI1223	-	putative cobalt/nickel resistance NcrA-like major facilitator superfamily permease	GEI 270	2.1	0.009
METDI1224	-	putative cobalt/nickel resistance NcrB-like regulator	GEI 270	2.8	< 0.001
METDI3043	-	conserved protein of unknown function with 2 CBS domains	core	4.4	0.002
METDI3067	-	putative methyl-accepting chemotaxis sensory transducer	core	2.5	0.009

METDI3303	-	putative extracellular protein with a calcium ion binding domain	core	2.8	< 0.001
METDI3835	*dctA*	C_4-dicarboxylate transport protein	core	-3.2	0.020
METDI4470	-	putative L,D-transpeptidase catalytic domain (YkuD)	GEI 197	-2.3	<0.001
METDI4699	-	transcriptional regulator, AraC family	core[e]	4.3	< 0.001
METDI5068	-	putative transcriptional regulator	variable[f]	3.2	0.011
METDI5746	*depP*	periplasmic serine protease (DegP)	core	6.6	0.010
METDI5875	-	putative photosynthesis gene regulator (*bchF-crtJ*)	variable	2.5	0.040
METDI5891	-	putative endonuclease	core	3.2	< 0.001

[a] MaGe annotation (https://www.genoscope.cns.fr/agc/microscope). See complete list Table S4

[b] Genes labeled as core, variable or specific are respectively found in all, at least 2 genomes of *M. extorquens* strains AM1, BJ001, CM4, DM4 and PA1 (80 % of identity on at least 80 % of the protein length), or only in the genome of strain DM4.

[c] log2 fold-change of normalized read numbers in cultures grown with dichloromethane compared to methanol.

[d] False discovery rate, only adjusted p-value under 0.1 were considered.

[e] Part of gene clusters: METDI1978-1985; METDI2004-2012; METDI2867-2874; METDI4813-4816; METDI5532-5533; METDI0091-0093; METDI4695-4699.

[l] Adjacent to a gene of unknown function with transcript abundance higher in dichloromethane versus methanol

[m] Corresponding encoded protein found more abundant in cell protein extracts of dichloromethane versus methanol-grown cultures (21).

Methanol-enhanced transcript abundance of gene clusters located on redundant genomic islands in *M. extorquens* DM4. Most of the genes with higher transcript abundance in cultures grown with methanol compared to dichloromethane (Tables 3 and S4, 121 genes) were localized on three GEI with sizes ranging from 160 to 270 kb (Table 4, 76 genes). The proportion of genes with differential transcript abundance represented 19 and 31 % of the gene content of two partially redundant GEI of 160 and 197 kb, called GEI 160 and GEI 197, respectively (Table 4). Amongst these, an *ackA*-like gene displayed the strongest differential transcript abundance for methanol in this study (log2fc value of -5.0 for METDI4544; Table S4). A comparative genetic map of the two redundant GEI regions revealed transcription pattern similarities between homologous genes sharing at least 74 % aa identity between their encoded proteins (Fig. S3). Differentially expressed genes have predicted functions involved in: i) central metabolism (two copies of putative acetate kinase-encoding *ackA*-like genes, one copy of an alcohol dehydrogenase-encoding *adh*-like gene, two copies of a D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase-encoding *xfp* gene), ii) energy metabolism (two copies of a cation-transporting ATPase, fourteen copies of cytochrome c and cytochrome c oxidase SU), iii) lipid metabolism (one copy of an enoyl-(acyl-carrier-protein) reductase-encoding *fabI* gene and a putative poly-beta-hydroxyalkanoate synthase-encoding *phbC* gene), iv) stress response (a copy of the heat shock protein encoding *dnaJ* gene and of a protein disaggregation chaperone-encoding *clpB* gene), and v) uncharacterized functions (five copies of genes coding for members of conserved proteins containing domain ANAH (adenine nucleotide alpha hydrolase protein superfamily Interpro SSF52402) (Fig. S3; S4).

Among the differentially transcribed homologous genes more abundant in methanol, a few were localized both on the variable and the core genomes, such as genes coding a cation-transporting ATPase, a cytochrome c oxidase, a Xfp enzyme and two members of the ANAH domain containing protein superfamily (Table 4). In conclusion, unique and accessory duplicated genes located on GEI may be involved in central metabolism, including carbon, energy and lipid metabolisms and stress for *M. extorquens* DM4 grown with methanol.

TABLE 4 Genomic island (GEI)-encoded transcripts differentially abundant in cultures of *M. extorquens* strains CM4 or DM4 grown with chlorinated methanes compared to methanol

Genomic island (GEI) and plasmid features [a]					Higher read counts[c] (CDS METDI)		Number of regulated-genes (ratio) & relevant functions (genes)
Name	Length (kb)	Start-End (CDS Label)[b]	Start-End (nt)	(left border) inside (right border)	Chlorinated methane	Methanol	
Within *M. extorquens* DM4							
dcm GEI	126	2551-2682	2459291-2585910	(none) tRNA-int-mob-misc_RNA-SIGI-AH (none)	2655-2658	none	4 (3) DCM utilization (*dcmRABC)*
GEI 160	160	0225-0425	232397-392215	(tRNA)-int-SIGI-AH (int)	0344	0259>0386	35 (18) central metabolism (*ackA*-like; *adh*-like; *xfp*), energy (ATPase; cyt. c; cyt. c oxidase SU), lipid metabolism (*fabI*; *phbC*), stress (*clpB*; *DnaJ*),
GEI 197	197	4320-4570	4260147-4456880	(none) tRNA-int-mob-SIGI-AH (IS)	None	4470; 4503>4565	33 (13) central metabolism (*ackA*-like; *adh*-like; *xfp*), energy (ATPase; cyt. c; cyt. c oxidase SU), members of ANAH-like superfamily
GEI 270	270	1158-1475	1065633-1335359	(tRNA) int-mob-misc_RNA-SIGI-AH (tRNA)	1223-1224	1414>1446	8 (3) metabolism (uncharacterized oxidoredcutase)
Within *M. extorquens* CM4							
GEI 84	84	1228-1308	1250078-1333760	SIGI-IVOM-Specific_Region	1258 ; 1287-1288	none	3 (4) transcriptional regulator
GEI 56	56	1470-1536	1531700-1587516	tRNA-int-SIGI-IVOM-Specific_Region	1525	1533>1536	5 (8) metabolism (NosX precursor)
GEI 12	12	1537-1558	1597506-1609551	(none)	None	1537>1540	4 (19) energy (cyt. c), metabolism (*mdh2*)
GEI 107	107	4756-4848	5048681-5155763	tRNA-int-SIGI-IVOM-Specific_Region	4776-4777; 4779	none	3 (3) proteins of unknown functions
GEI 28	28	5040-5066	5394282-5422511	(none)	5057-5058; 5060; 5066	none	4 (15) central metabolism (*metA* monooxygenase, tRNA synthetase) lipid metabolism (*acd*)
pCMU01	380	5386-5736	0-380207	nd[d]	5499; 5676-5704; 5714-5717; 5721-5732	none	62 (18) chloromethane utilization (*cmu*), corrinoïd cofactor biosynthesis (*cob*, *czc*, *bluB2*), transposase

[a] Basing on "Regions of Genomic Plasticity" finder (MaGe plateform) using *M. extorquens* CM4, DM4, PA1 and BJ001 genomes. Int, integrase; mob, mobility determinant; AH, region rich in genes detected by Alien Hunter based on variable-length k-mers bias (Vernikos and Parkhill, 2006) and SIGI, region with biased codon usage (62). Plasmid pCMU01 detected so far only in *M. extorquens* CM4 (22).

[b] Genes are labelled as Mchl and METDI in *M. extorquens* CM4 and DM4, respectively.

[c] Chlorinated methanes (log2fc > 2), methanol (log2fc < -2) as defined in material and methods.

[d] Not determined

DISCUSSION

The ratio of carbon source-dependent transcript abundance of *M. extorquens* genes was significantly lower for genes in the core genome than in the variable genome (Table 5).

TABLE 5 Contribution of core and variable genome genes to carbon source-dependent differential transcriptional expression

Strain	Genome	CDS number	Differential transcript abundance		
			Ratio (%)	Gene number[a]	
				Up	Down
M. extorquens CM4	core	3,489	1.6	45	11
	variable[b] *(CM4 only)*	2,783 *1,512*	3.4 *2.0*	92 *29*	2 *1*
M. extorquens DM4	core	3,489	1.9	31	34
	variable[b] *(DM4 only)*	2,112 *952*	5.4 *5.6*	28 *10*	87 *43*

[a] Transcript abundance higher (log2fc > 2) or lower (log2fc < -2) in cultures grown with chlorinated methane compared to with methanol.

[b] Variable genome includes genes not found in all *M. extorquens* genomes. Italics refer to strain-specific genes.

Strain-specific genes were among the most transcribed genes, especially gene clusters involved in chlorinated methane dehalogenation. The *dcm* genes were more expressed in growth with dichloromethane compared to methanol: *dcmA* coding for the dichloromethane dehalogenase DcmA essential for growth with dichloromethane, *dcmR* coding for a regulator of its own gene and of *dcmA* transcription (34,46), *dcmB* coding for a protein of unknown function previously demonstrated to be dichloromethane-induced (21), and the uncharacterized *dcmC*. Similarly, the *cmuAB* genes coding for essential chloromethane dehalogenation enzymes underwent massive induction in cultures grown with chloromethane, as previously described (63). Adjacent to the *cmuAB* genes on plasmid pCMU01, other *cmu* pathway-associated genes may also be massively induced (*cmuC*, *cmuC2*, *folD*, *hutI*, *metF2*, paaE, *purU*) as suggested in this study (22,51). In addition, plasmid-borne genes with chromosomal homologs had higher transcript abundance in cultures with chloromethane compared to methanol, unlike the chromosomal copies. This includes 14 *cob* genes of the *cmu* pathway-associated corrinoid coenzyme biosynthesis (22), as well as *gck2,* which encodes for a putative glycerate kinase displaying 85 % of aa identity with the chromosomal homolog (Mchl_2974) involved in carbon assimilation by the serine pathway (log2fc value of 5.1) (Table 2; *clcA* homologs are discussed in the next paragraph). This

suggests that a yet uncharacterized chloromethane-dependent regulation mechanism favors the expression of genes of the variable genome over homologous genes of the core genome. In the chloromethane transcriptome, the *acx* genes of the acetone carboxylase complex had higher transcripts in cultures grown with chloromethane compared to methanol (log2fc values of 2.1 to 3.7; Table 2). The *acx* genes, part of the variable genome located on plasmid pCMU01 in *M. extorquens* CM4, were conserved in the few sequenced genomes of isolates that degrade chloromethane by the *cmu* pathway (51). Acetone has been described as the substrate that regulates expression of the *acx* genes in the acetone-degrading *Helicobacter pylori* (8), but the compound that induces *acx* gene expression during growth with chloromethane has yet to be identified.

The transcriptional response common to *M. extorquens* growing with chloromethane or dichloromethane is limited to a few genes only. Bacteria that grow with chlorinated methanes have to cope with intracellular production of chloride ions and protons that are formed during the first catabolic reactions (Fig. 1). Thus, we expected to find higher expression of housekeeping genes involved in chloride and proton associated stress and bioenergetics as a response to growth with chloromethane or dichloromethane. We identified such genes in this study but only 11 among 4,620 genes were shared between the chloromethane- and dichloromethane-degrading strains (Fig. 4 group B; Table S2). Of these genes, *depD* and *hppA* had characterized homologs in other bacteria and were highly expressed in cultures grown with chlorinated methanes (Table S2, group B). The central housekeeping molecular chaperon protein DepD (48) repairs misfolded proteins damaged by reactive oxygen produced during aerobic chlorinated methane utilization (21,22). As a proton pump, the membrane-bound proton translocating pyrophosphatase HppA most likely plays a role in proton extrusion and stress bioenergenetics, as demonstrated for a highly similar protein characterized in *Rhodospirullum rubrum* (72 % aa identity with *M. extorquens* homolog) (10). For chloride extrusion, the H^+/Cl^- exchange transporter ClcA was encoded by a constitutively highly expressed core genome gene in *M. extorquens* DM4, which confirms previous RT-qPCR data in wild-type dichloromethane-degrading isolates (11). A different situation occurred in *M. extorquens* CM4 that harbors two *clcA* paralogs (69 % aa Id). The *clcA* chromosomal core genome copy displayed moderate constitutive transcript abundance whereas the *clcA2* plasmid-borne gene significantly increased in transcript abundance with chloromethane

(log2fc value of 5.5), suggesting that *clcA*-driven chloride extrusion is also required during growth with chloromethane. The occurrence of *clcA* paralogs with specific expression patterns suggests that growth with chloromethane in *M. extorquens* CM4 and growth with dichloromethane in *M. extorquens* DM4 both generate an adaptive response to chloride stress that is similar but not identical. Previous studies have shown that when gene *clcA* is transcribed from its native promoter present in a synthetic expression module containing *dcmA*, it confers higher fitness to dichloromethane in *Methylobacterium* strains that had not been previously exposed to dichloromethane (11). This is not the case when using a synthetic expression module containing the *cmu* pathway for growth with chloromethane (64) (J. K. Michener, S. Vuilleumier, F. Bringel, C. J. Marx, personal communications; submitted article). In addition, in laboratory evolution experiments, the genetic modules for heterologous utilization of chloromethane and dichloromethane have shown that phylogeny poorly predicts the utility of a challenging horizontally transferred gene in *Methylobacterium* strains (65), which would suggest that strain-specific rather than core genome-encoded genes play a crucial role in adaptation to growth with chlorinated methanes.

In both *M. extorquens* strains, 9 genes with lower transcript abundance were identified in cultures grown with chlorinated methanes compared to methanol (Fig. 4 group E; Table S2). Among these, *dctA* encodes a (Na^+/H^+)-C_4-dicarboxylate transporter (Dct) component highly similar to characterized DtcA proteins (77 % and 63 % aa identity with *Rhizobium meliloti* and *E. coli* DctA, respectively) (47). For each C_4-dicarboxylate molecule imported, the Dct transporter imports two protons (12). Therefore, during growth with chlorinated methanes, an additional increase in intracellular proton concentration would be detrimental and may explain the higher *dctA* transcript levels in cultures grown with methanol compared to those grown with chloromethane or dichloromethane. In a previous study in *M. extorquens* AM1, higher *dctA* transcript levels were found in cultures grown with succinate compared to methanol while other two core genome *dctA* copies were not differentially expressed (66). Of the three DctA homologs found in *M. extorquens* genomes that share around 50 % aa Id, only one *dctA* copy has been identified as differentially expressed in this study (Table S2). Thus, the expression of the same *dctA* homolog was modulated by carbon source-dependent growth conditions in the three *M. extorquens* strains. In *E. coli*, DctA is a component of the Dct transporter, as well as a co-regulator of the membrane-bond carboxylate sensor kinase DctS

(67). Therefore the differentially transcribed *dctA* gene may also play a potential role in carboxylate compound sensing in *M. extorquens*.

This study did not find any differentially expressed genes amongst the 163 genes shared exclusively between the genomes of the two chlorinated methane-degrading strains and absent in other non-degrading closely-related *M. extorquens* strains (Fig. 3; data not shown). Conversely, a larger number of genes only present in either *M. extorquens* CM4 (6 genes) or *M. extorquens* DM4 (10 genes) were differentially expressed in cultures grown with the corresponding chlorinated methane compared to methanol (Tables 2, 3, S3 and S4). Our study found that core genome encoded housekeeping proteins were differentially expressed in cultures grown with dichloromethane compared to methanol (two uncharacterized oxidoreductase with molybdenum cofactors METDI0092 and METDI2693; a putative glutathione peroxidase; an ornithine cyclodeaminase ArcB). Other genes were only detected in cultures grown with chloromethane (components of the $NAD(P)^+$ transhydrogenase Pnt and of the glycine cleavage system Gcv; fused transcriptional regulator and O6-methylguanine-DNA methyltransferase Ada) (Fig. 4). We found that the transcriptional responses to growth with chloromethane or dichloromethane are predominantly uncorrelated, except a restricted number of core-encoded genes involved in stress and bioenergetic intracellular HCl production (*clcA* paralogs, *depP*, *ibpA* and *hppA*).

Core genome and strain-specific transcription patterns of methanol dehydrogenase and alternative genes for alcohol/aldehyde utilization. The *M. extorquens* core genome harbors several sets of methanol dehydrogenase (MDH)-encoding genes including the canonical MxaFI, the environmental-widespread XoxF-type, and a Mdh2-type MDH encoded gene cluster closely related to alcohol dehydrogenase ADH. The *mxaFI* and *xoxF* genes were not significantly more expressed during growth with methanol compared to chloromethane or dichloromethane (core genome-encoded genes *mxa* and *xox* and pCMU01 plasmid-borne gene *xoxF2*). The unchanged expression profile with *xoxF* was expected since transcription of *xoxF*-type genes are often undetected in cultures grown in the laboratory under conditions different from environmental samples (68). Earlier microarray mRNA quantification in *M. extorquens* strain AM1 found that *mxaFI* transcript abundance in cultures grown with methanol was higher than with succinate, suggesting that *mxaFI* genes were induced during methylotophic growth, whereas no gene belonging to the *mdh2* gene cluster was more

abundant (66). In *M. extorquens* AM1, *mdh2* alone cannot sustain growth with methanol since a *mxaFI* mutant was unable to utilize methanol (69). In this study, surprisingly, all 8 genes of the *mdh2* gene cluster were more expressed in methanol-grown strains CM4 and DM4 (Fig. 4). The Mdh2 protein typically has low affinity for methanol except in *Betaproteobacteria* species (40). In *M. extorquens*, Mdh2 enzyme affinity for methanol and other alcohol or aldehyde compounds still needs to be tested whereas MxaFI catalyzes the oxidation of methanol as well as formaldehyde (69). In this study, very high normalized read numbers of *mxaFI* genes were detected in all the tested methylotrophic cultures although formaldehyde is a byproduct of methanol and dichloromethane degradation but not of chloromethane dehalogenation catalyzed by CmuAB (Fig. 1) (34,41,69). The equilibrium between free formaldehyde and methylene-tetrahydrofolate (CH_2=H_4F) lies overwhelmingly on the side of methylene-H_4F formation, so significant reverse flux from CH_2=H_4F to formaldehyde and then to formate seems unlikely during chloromethane utilization (70). The high level of normalized reads of *mxaFI* mRNA during growth with chloromethane (Fig. S2A) suggests a yet unknown role of MxaFI during growth with chloromethane or a loss of regulation in strain CM4.

Two *adh*-like genes of the family of zinc-dependent alcohol dehydrogenases (ADH) encode similar proteins in strain DM4 (METDI0283 and METDI4540 display 90 % aa Id). They catalyze the NAD(P)(H)-dependent interconversion of alcohols into aldehydes or ketones (71) and may be involved in methanol utilization or C_2 metabolism (Fig. 1). These two *adh* genes, like a large majority of the 121 transcripts found to be more abundant in cultures grown with methanol compared to dichloromethane, are encoded by the variable genome and in particular two partially-redundant GEI (68 genes; GEI 160 and GEI 197, Fig. S3; Tables 3 and S4). This shared GEI region is also found as a single copy on the genome of *M. extorquens* AM1 nested within a 64 kb-long GEI (Fig. S3). However, no differential transcript abundance has been detected in previous microarray mRNA quantification in cultures grown with methanol compared to succinate as growth substrate (66). This would suggest a dichloromethane-dependent repression of the GEI genes or alternatively, a strain-specific methanol-dependent regulation process. The *adh*-like genes and adjacent genes located on two redundant GEI found in *M. extorquens* DM4 may be involved in central metabolism readjustments during methylotrophic growth with methanol compared to dichloromethane.

Redox status adjustment as a possible key adaptation to growth with different C_1 compounds. Formate dehydrogenases are central methylotrophic modules for C_1 dissimulation (Fig. 1) and are encoded by four core genome *fdh* gene clusters (44,45). In previous studies, the protein Fhd2B was detected as more abundant in strain DM4 grown with dichloromethane compared to methanol (21), while *fdh1* and *fdh2* genes had higher transcript abundance in strain AM1 grown with methanol compared to succinate (66). In this study, the only difference in transcript abundance for *fdh4AB* was detected in *M. extorquens* DM4 (Tables 3 and S4). Taken together, expression of different copies of NAD-dependent formate dehydrogenase components would suggest that methanol versus dichloromethane utilization may affect specific redox reactions. This hypothesis is confirmed by the observation that several types of cytochrome oxidase-encoded gene clusters, which belong to the core and the variable genome, had higher transcript counts in methanol-grown cultures (Fig. 1; Table S3, for instance *cbaB*, *qxtB*, *cydB* genes for an ATP transporter and an adjacent cytochrome d terminal oxidase (METDI2004-METDI2012). In particular, within the variable genome of *M. extorquens* DM4, a large set of GEI-encoded genes involved in cytochrome electron transport and the formation of H^+ gradient had enhanced transcript abundance in cultures grown with methanol, and thus may be involved in ATP production (Fig. 1).

Other energy-associated adjustments in *M. extorquens* DM4 grown with methanol compared to dichloromethane include very high transcript reads of genes involved in lipid metabolism (Fig. 1; Tables 3 and S4.). Poly-beta-hydroxyalkanoate (PHB) constitutes a carbon and energy reserve material during growth with both multicarbon- and single-carbon substrates (72). It can represent 2.3 % of the cell dry weight of *M. extorquens* growing with methanol (4). PHB metabolism controls carbon and energy spillage and its effectiveness depends on nitrogen availability (9,73). In this study we identified five genes of the core genome with higher transcript abundance in cultures grown with methanol. These five genes encode nitrogen-related functions (regulatory protein (*glnK*), transporters (*atmB*; *ntrAB*; putative formate/nitrate transporter METDI1697) and enzymes (glutamine synthase glnII; cysteine desulfurase; Table 3). In summary, the differential RNAseq (methanol versus dichloromethane) study revealed the differential expression of core and variable genome-encoded genes with functions associated with energy metabolism.

For each molecule of chloromethane transformed to formate by the *cmu* pathway, two molecules of NADPH are generated from $NADP^+$ in two successive reactions catalyzed by

MetF2 and FolD (Fig. 1). The corresponding genes *metF2* and *folD* are two of the most differentially expressed genes in the chloromethane transcriptome (Table 2). Thus, lowering the intracellular pool of $NADP^+$ should limit growth with chloromethane unless a metabolic reaction could regenerate $NADP^+$. A good candidate for this is the membrane-bound transhydrogenase encoded the *pnt* gene cluster, which was significantly more expressed on chloromethane both at the level of transcription and protein (Table 2) (22). It catalyzes the reversible reaction $NADPH + NAD^+ + H^+_{in} <=> NADP^+ + NADH + H^+_{out}$ (74). This membrane-bound transhydrogenase plays a predominant role in laboratory evolution of *M. extorquens* and *E. coli* deficient in NADPH production (49,66), which would be the case in cells utilizing chloromethane with the *cmuA* pathway. In addition to $NADP^+$ regeneration via the membrane-bound transhydrogenase, concomitant proton efflux would also lower internal cellular pH occurring during chloromethane dehalogenation. Whether this core genome membrane-bound transhydrogenase is essential for $NADP^+$ regeneration in cultures grown with chloromethane still needs to be tested.

Finally, we observed that the difference in transcript abundance was the highest for genes of the variable genome (*cmu* genes) during growth with chloromethane compared to methanol (log2fc of 10 instead of 4 for *dcm* genes). On the other hand, similar transcript abundances were detected for core genome-encoded genes when both strains grew with methanol (Fig. S2). Thus, chloromethane-dependent transcription regulation seems very effective and confirms earlier studies (63). Here, we observed a looser dichloromethane-dependent control at the transcription level with high *dcm* transcript abundance even in the non-induced condition (Table 3). At the protein level, nevertheless, no DcmA protein has been detected in the methanol proteome while it accounts for about 25 % of total proteins in the dichloromethane proteome (21). These observations strongly suggest that an effective post-transcriptional regulation of the variable genome of the *dcm* islet of *M. extorquens* DM4 may take place. Future transcriptional start site determination will be conducted to characterize core and variable genome transcription regulatory elements and mechanisms responsible for growth with different C_1 compounds.

MATERIALS AND METHODS

Bacterial growth conditions and RNA purification. *M. extorquens* strains CM4 and DM4 grew aerobically at 30°C on a rotary shaker (120 rpm) in 1.2-liter Erlenmeyer flasks closed with gas-tight screw caps with mininert valves (Supelco) in 220 mL *Methylobacterium* mineral medium (M3) adapted from (75) ($ZnSO_4$ replaced $ZnCl_2$ and $(NH_4)_2SO_4$ was at final concentration 0.2 $g.L^{-1}$). The one-carbon compounds added at 10 mM final concentration were either methanol (2.75 mL of a methanol filter-sterilized stock solution at 800 mM), dichloromethane (141 µL of dichloromethane 100 %) or chloromethane (40 mL of CH_3Cl 100 % assuming a Henry constant of 0.0106 $m^3.atm.\text{-}mol^{-1}$ at 30°C as previously described) (76). The 220-mL cultures were stopped during the mid-exponential phase ($OD_{600nm} \approx 0.15$) by the addition of 27.5 mL of stabilization buffer. The stabilization buffer was prepared as follows: 5 mL of phenol was mixed with 5 mL of 1 M sodium acetate pH 5.5 and 90 mL of DEPC-treated water. After centrifugation at 1800 g for 3 min, 5 mL of the inferior phase was mixed with 95 mL of ethanol 100 %. Cell suspensions were centrifuged at 5000 rpm for 5 min at 4°C and resuspended in 10 mL of TE containing 2 mg. mL^{-1} of lysozyme (Euromedex). After an incubation 15 min at 37°C, the suspension was centrifuged 10 min at 4°C, then the pellet was resuspended with 10 mL of Trizol (Invitrogen), then 2.5 mL of chloroform was added. RNAs were precipitated with isopropanol and washed with ethanol (70 %), then resuspended in DEPC-water and treated with DNAse (TurboDNAse, Invitrogen). DNA depletion was checked by PCR on *rrs* (27f, 5'-GTTTGATCCTGGCTCAG-3' and 927r, 5'-CCGTCAATTCCTTTRAGTTT-3') and on specific genes: *cli* (cli1R, 5'-ATCGTCCTGATCGCACCT-3' and cli1F, 5'-GCCTCGAAACGTTGAACACC-3'), *dcmA* (dcmAiR, 5'-TCGATCTAGATAGCGCGGTCTAAGGGAC-3' and dcmAiF, 5'-GACTTCTAGACGTGAGCCCGAATCAAC-3') or *cmuA* (cmuA802F, 5'-TTCAACGGCGAYATGTATCCYGG-3' and cmuA1244*R*, 5'-TABTCCATDATGGCYTCGA-3'). RNA quality was checked on a Bioanalyzer 2100 using RNA 6000 Nano ship (Agilent) and quantified with Qubit RNA kit (Invitrogen). Then, rRNA were removed by treating 5 µg of total RNA with Gram negative RiboZero Magnetic kit (Tebu-Bio) according to the manufacturer's protocol.

cDNA library preparation, sequencing and data normalization. Fifty to sixty ng of rRNA-depleted RNA were used to construct cDNA banks with the TruSeq stranded mRNA LT kit (Illumina). This construction started with the step "Purify and fragment mRNA" from Illumina protocol, by adding 13 µL of the solution "Fragment, Prime, Finish Mix" to the 5 µL of RNA containing 50-60 ng of RNA. Then the manufacture's protocol was followed. Bank

constructions were validated by Bioanalyzer with DNA 1000 kit (Agilent). Sequencing was performed using HiSeq2000. Paired-end sequence mapping was used to limit mapping artifacts and remove any ambiguity at mapping level between homologs. For data normalization (Fig. S1), the log2fc value of normalized read numbers in chlorinated methanes versus methanol was normalized with a set of 55 reference genes listed in Table S1, and a cut-off value = 1. The reference genes were unique and constitutively expressed in the tested conditions. Genes were defined as differentially abundant in cultures grown with chlorinated methane compared to methanol when log2fc values were ≥ 2.0 and differentially abundant in cultures grown with methanol compared to chlorinated methanes when log2fc values were ≤ -2.0 (77). Each condition was performed in duplicate; the number of reads was based on the average of read sense and antisense for each condition. Data were only considered significant when the p-value was ≤ 0.1. The complete RNAseq dataset is accessible online (https://www.genoscope.cns.fr/agc/microscope/transcriptomic/NGSProjectRNAseq.php?projType=RNAseq).

Comparative genomics and phylogenetic analysis. The core genome of *M. extorquens* was done using the sequenced genomes of five strains of the *M. extorquens* species (GenBank accession n°): strain AM1 (CP001511, CP001512, CP001513 and CP001514); strain BJ001 (CP001029, CP001030 and CP001031); strain CM4 (CP001298, CP001299, CP001300); strain DM4 (FP103042 and FP103043 and FP103044) and strain PA1 (CP000908) (28,29). The core genome was established via the MaGe platform (https://www.genoscope.cns.fr/agc/microscope/compgenomics/pancoreTool.php?). A gene was defined as belonging to the core genome when it encodes a protein sharing at least 80 % of identity on 80 % of CDS length in all the tested genomes. Gene was defined as belonging to the core genome when it is shared by the 5 strains, or belonging to the variable genome when it is shared at least by 2 strains and maximum by 4 strains. Sequences were aligned via ClustalX (78). The phylogenetic tree was constructed using the maximum parsimony method with a bootstrap of 500 replicates with MEGA4 software (79).

SUPPLEMENTAL MATERIAL

FIGURE S1 Comparison of genome wide gene transcription between *M. extorquens* strains CM4 and DM4 grown with methanol. The percentage of reads represents the number of reads for each gene divided by the cumulated number of reads of the 4,620 genes shared (at

least 80 % of identity on the protein level on 80 % of the CDS length) in a given strain. Name of highly transcribed genes are indicated. (A) Normalization using 55 reference genes. (B) Normalization with the standard method using the complete set of CDS.

FIGURE S2 Genes with highest differential transcription in cultures grown with chlorinated methanes compared to methanol. Black rectangles indicate values of log2fc (values higher than 2.0 or lower than -2.0 mean normalized reads more abundant in cultures grown with chlorinated methanes or with methanol, respectively). Grey rectangles represent normalized read numbers divided by gene length in kb. (A) RNAseq data of *M. extorquens* CM4 cultures grown with chloromethane compared to methanol. (B) RNAseq data of *M. extorquens* DM4 cultures grown with dichloromethane compared to methanol.

FIGURE S3 Transcription profiles and gene redundancy of key gene clusters more highly transcribed in methanol- than dichloromethane-grown cultures of *M. extorquens* DM4. Conserved stretches between two GEI-borne regions (Table 4) are delimited with grey rectangles. Numbers indicate the percentage of amino acid identity between encoded METDI protein labels defined in MicroScope (https://www.genoscope.cns.fr/agc/microscope/mage/viewer. php?S_id=240). Genes are indicated as colored arrows: in red when higher normalized read numbers were detected in cultures grown with methanol compared to dichloromethane (log2fc ≤ -2), in blue when no significant difference was observed (-2 < log2fc values >2), and in purple for the METDI0344-encoding gene detected with higher reads in dichloromethane (log2fc value of 2.2). Interrupted GEI segments are indicated with zigzagged arrows. IS stands for insertion sequences. The ANAH conserved domain is found in proteins of the “adenine nucleotide alpha hydrolases-like superfamily” including N type ATP PPases, ATP sulphurylases, universal stress response proteins and electron transfer flavoproteins (http://supfam.cs.bris.ac.uk/SUPERFAMILY/cgi-bin/scop.cgi?sunid=52402).

FIGURE S4 Phylogenetic tree of proteins with adenine nucleotide alpha hydrolase (ANAH) domains in *M. extorquens* strains CM4 and DM4. Alphaproteobacteria members of the aligned ANAH family proteins. Protein containing only one ANAH domain (Interpro SSF52402) are boxed. Alignments were performed using ClustalX (78) for proteins mean sizes between 270-280 aa. The phylogenetic tree was constructed using software MEGA (79) with the neighbor joining method (Saitou and Nei, 1987). Color-coded protein names correspond to genes with higher (in orange) or unchanged (in black) transcript abundance in *M. extorquens* cells grown with methanol compared to chlorinated methanes. In parenthesis, protein-encoded gene occurrence in the variable genome (genomic island) or in *M. extorquens* shared genes (core). Symbols indicate when an ANAH-encoding gene is located in the vicinity (less than 2 kb with) of a gene coding for an ATPase (star), a cytochrome c (disk), an alcohol dehydrogenase-like (ADH-like) (triangle) or *fixJL* (square) encoding a putative two component low oxygen sensor histidine kinase. Sequence accession numbers for *M. extorquens* CM4: Mchl_0097 (WP_012605265), Mchl_1555 (WP_015950242), Mchl_3807 (WP_015951826), Mchl_3811 (WP_015951830), Mchl_3815 (WP_015951832), Mchl_4817 (WP_015952568); for *M. extorquens* DM4: METDI0258 (WP_0127790266), METDI0272 (WP_012779075), METDI0274 (WP_012779077), METDI0276 (WP_012779079), METDI0302 (WP_012779104), METDI0320 (WP_012779122), METDI0323 (WP_012779125), METDI2004 (WP_015821999), METDI2442 (WP_01582230), METDI4304 (WP_015823643), METDI4309 (WP_015823647),

METDI4314 (WP_015823651), METDI4521 (WP_012753369), METDI4535 (WP_015823830), METDI4537 (WP_012753383), METDI4539 (WP_012753385), METDI4555 (WP_0158238333), METDI4563 (WP_012753411), METDI4565 (WP_01275344); for *Bradyrhizobium japonicum* (WP_011085546*), Sinorhizobium meliloti* (WP_015242313) and *Xanthobacter autotrophicus* (WP_012115156).

TABLE S1 Set of 55 genes used for RNAseq data normalization

TABLE S2 Groups of genes with similar transcription shared between *M. extorquens* strains CM4 and DM4

TABLE S3 Differentially abundant transcripts in cultures of *M. extorquens* CM4 grown with chloromethane compared to methanol (Table 2 complementary list)

TABLE S4 Differentially abundant transcripts in cultures of *M. extorquens* DM4 grown with dichloromethane compared to methanol

ACKNOWLEDGMENTS

PhD grant to CP (funded by Région Alsace (France) and partially by DFG grant ko2912/5-1 (Germany); the German Academic Exchange Service (DAAD); University of Bayreuth) and Franco-German mobility funds (Deutscher Akademischer Austausch Dienst (PROCOPE)) are gratefully acknowledged. Thanks to Beatrice Segurens (Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA) CNRS-UMR8030, France) for RNA extraction protocol optimization, and David A. C. Beck (University of Washington, Seattle, USA) and Anne Francez-Charlot (ETH Zurich, Swiss) for checking proposed RNAseq reference genes in unpublished *M. extorquens* transcriptomic libraries.

REFERENCES

1. **Galbally IE, Kirstine W**. 2002. The production of methanol by flowering plants and the global cycle of methanol. J Atmospheric Chem **43**: 195–229. 10.1023/A:1020684815474.
2. **Clerbaux C, Cunnold DM, Anderson J**. 2006. Halogenated long-lived compounds, scientific assessment of ozone depletion. Global Ozone Research and Monitoring Project 10 Report No 50. World Meteorological Organization, Geneva.
3. **Trotsenko YA, Doronina NV.** 2003. The biology of methylobacteria capable of degrading halomethanes. Microbiology **72**: 121–131. doi:10.1023/A:1023218326407

4. **Peyraud R, Schneider K, Kiefer P, Massou S, Vorholt JA, Portais J-C.** 2011. Genome-scale reconstruction and system level investigation of the metabolic network of *Methylobacterium extorquens* AM1. BMC Syst Biol **5**: 189. doi:10.1186/1752-0509-5-189
5. **Šmejkalová H, Erb TJ, Fuchs G.** 2010. Methanol assimilation in *Methylobacterium extorquens* AM1: demonstration of all enzymes and their regulation. PLoS ONE **5**: e13001. doi:10.1371/journal.pone.0013001
6. **Chistoserdova L.** 2011. Modularity of methylotrophy, revisited: modularity of methylotrophy, revisited. Environ Microbiol **13**: 2603–2622. doi:10.1111/j.1462-2920.2011.02464.x
7. **Schmidt S, Christen P, Kiefer P, Vorholt JA.** 2010. Functional investigation of methanol dehydrogenase-like protein XoxF in *Methylobacterium extorquens* AM1. Microbiology **156**: 2575–2586. doi:10.1099/mic.0.038570-0
8. **Brahmachary P, Wang G, Benoit SL, Weinberg MV, Maier RJ, Hoover TR.** 2008. The human gastric pathogen *Helicobacter pylori* has a potential acetone carboxylase that enhances its ability to colonize mice. BMC Microbiol **8**: 14. doi:10.1186/1471-2180-8-14
9. **Sauer U.** 2003. The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. J Biol Chem **279**: 6613–6619. doi:10.1074/jbc.M311657200
10. **Lopez-Marques RL, Perez-Castineira JR, Losada M, Serrano A.** 2004. Differential regulation of soluble and membrane-bound inorganic pyrophosphatases in the photosynthetic bacterium *Rhodospirillum rubrum* provides insights into pyrophosphate-based stress bioenergetics. J Bacteriol **186**: 5418–5426. doi:10.1128/JB.186.16.5418-5426.2004
11. **Michener JK, Neves AAC, Vuilleumier S, Bringel F, Marx CJ.** 2014. Effective use of a horizontally-transferred pathway for dichloromethane catabolism requires post–transfer refinement. eLife **3**: e04279. doi:10.7554/eLife.04279.001
12. **Janausch IG, Zientz E, Tran QH, Kröger A, Unden G**. 2002. C_4-dicarboxylate carriers and sensors in bacteria. Biochim Biophys Acta BBA-Bioenerg **1553**: 39–56. doi:10.1016/S0005-2728(01)00233-x
13. **Davies SJ, Golby P, Omrani D, Broad SA, Harrington VL, Guest JR, *et al*.** 1999. Inactivation and regulation of the aerobic C_4-dicarboxylate transport *dctA* gene of *Escherichia coli*. J Bacteriol **181**: 5624–5635. doi:10/1999; 181(18):5624-35.

14. **Naik S, Venu Gopal SK, Somal P.** 2008. Bioproduction of polyhydroxyalkanoates from bacteria: a metabolic approach. World J Microbiol Biotechnol **24**: 2307–2314. doi:10.1007/s11274-008-9745-z

15. **Korotkova N, Lidstrom ME.** 2001. Connection between poly-hydroxybutyrate biosynthesis and hrowth on C_1 and C_2 compounds in the methylotroph *Methylobacterium extorquens* AM1. J Bacteriol **183**: 1038–1046. doi:10.1128/JB.183.3.1038-1046.2001

16. **Thomson AW, O'Neill JG, Wilkinson JF.** 1976. Acetone production by methylobacteria. Arch Microbiol **109**: 243–246. doi:10.1007/BF00446635

17. **Eggers J, Steinbuchel A.** 2013. Poly-3-Hydroxybutyrate degradation in *Ralstonia eutropha* H16 is mediated stereoselectively to (S)-3-hydroxybutyryl coenzyme A (CoA) via crotonyl-CoA. J Bacteriol **195**: 3213–3223. doi:10.1128/JB.00358-13

18. **Sznajder A, Jendrossek D.** 2014. To be or not to be a poly-3-hydroxybutyrate (PHB) depolymerase: PhaZd1 (PhaZ6) and PhaZd2 (PhaZ7) of *Ralstonia eutropha*, highly active PHB depolymerases with no detectable role in mobilization of accumulated PHB. Appl Environ Microbiol **80**: 4936–4946. doi:10.1128/AEM.01056-14

19. **Kayser MF, Vuilleumier S.** 2001. Dehalogenation of dichloromethane by dichloromethane dehalogenase/glutathione S-transferase leads to formation of DNA adducts. J Bacteriol **183**: 5209–5212. doi:10.1128/JB.183.17.5209-5212.2001

20. **Torgonskaya ML, Doronina NV, Hourcade E, Trotsenko YA, Vuilleumier S.** 2011. Chloride-associated adaptive response in aerobic methylotrophic dichloromethane-utilising bacteria. J Basic Microbiol **51**: 296–303. doi:10.1002/jobm.201000280

21. **Muller EEL, Hourcade E, Louhichi-Jelail Y, Hammann P, Vuilleumier S, Bringel F.** 2011. Functional genomics of dichloromethane utilization in *Methylobacterium extorquens* DM4: functional genomics of *Methylobacterium* dichloromethane utilization. Environ Microbiol **13**: 2518–2535. doi:10.1111/j.1462-2920.2011.02524.x

22. **Roselli S, Nadalig T, Vuilleumier S, Bringel F.** 2013. The 380 kb pCMU01 plasmid encodes chloromethane utilization genes and redundant genes for vitamin B_{12}- and tetrahydrofolate-dependent chloromethane metabolism in *Methylobacterium extorquens* CM4: a proteomic and bioinformatics study. PLoS ONE **8**: e56598. doi:10.1371/journal.pone.0056598

23. **Ochsner AM, Sonntag F, Buchhaupt M, Schrader J, Vorholt JA.** 2015. *Methylobacterium extorquens*: methylotrophy and biotechnological applications. Appl Microbiol Biotechnol **99**: 517–534. doi:10.1007/s00253-014-6240-3

24. **Bringel F, Couée I.** 2015. Pivotal roles of phyllosphere microorganisms at the interface between plant functioning and atmospheric trace gas dynamics. Front Microbiol **06**. doi:10.3389/fmicb.2015.00486

25. **Temkiv TŠ, Finster K, Hansen BM, Nielsen NW, Karlson UG.** 2012. The microbial diversity of a storm cloud as assessed by hailstones. FEMS Microbiol Ecol **81**: 684–695. doi:10.1111/j.1574-6941.2012.01402.x

26. **Kolb S.** 2009. Aerobic methanol-oxidizing in soil. FEMS Microbiol Lett **300**: 1–10. doi:10.1111/j.1574-6968.2009.01681.x

27. **Mitsui R, Omori M, Kitazawa H, Tanaka M.** 2005. Formaldehyde-limited cultivation of a newly isolated methylotrophic bacterium, *Methylobacterium* sp. MF1: enzymatic analysis related to C_1 metabolism. J Biosci Bioeng **99**: 18–22. doi:10.1263/jbb.99.18

28. **Marx CJ, Bringel F, Chistoserdova L, Moulin L, Farhan Ul Haque M, Fleischman DE, *et al*.** 2012. Complete genome sequences of six strains of the genus *Methylobacterium*. J Bacteriol **194**: 4746–4748. doi:10.1128/JB.01009-12

29. **Vuilleumier S, Chistoserdova L, Lee M-C, Bringel F, Lajus A, Zhou Y, *et al*.** 2009. *Methylobacterium* genome sequences: a reference blueprint to investigate microbial metabolism of C_1 compounds from natural and industrial sources. Ahmed N, editor. PLoS ONE **4**: e5584. doi:10.1371/journal.pone.0005584

30. **Doronina NV, Sokolov AP, Trotsenko YA.** 1996. Isolation and initial characterization of aerobic chloromethane-utilizing bacteria. FEMS Microbiol Lett **142**: 179–183. doi:10.1016/0378-1097(96)00262-5

31. **Gälli R, Leisinger T.** 1985. Specialized bacterial strains for the removal of dichloromethane from industrial waste. Conservation & Recycling **8**: 91–100. 10.1016/0361-3658(85)90028-1

32. **Studer A, Stupperich E, Vuilleumier S, Leisinger T.** 2001. Chloromethane: tetrahydrofolate methyl transfer by two proteins from *Methylobacterium chloromethanicum* strain CM4. Eur J Biochem **268**: 2931–2938. doi:10.1046/j.1432-1327.2001.02182.x

33. **Nadalig T, Farhan Ul Haque M, Roselli S, Schaller H, Bringel F, Vuilleumier S.** 2011. Detection and isolation of chloromethane-degrading bacteria from the *Arabidopsis*

936 *thaliana* phyllosphere, and characterization of chloromethane utilization genes:
937 chloromethane-degrading bacteria from *A. thaliana*. FEMS Microbiol Ecol **77**: 438–448.
938 doi:10.1111/j.1574-6941.2011.01125.x
939 34. **Leisinger T, Bader R, Hermann R, Schmid-Appert M, Vuilleumier S.** 1994. Microbes,
940 enzymes and genes involved in dichloromethane utilization. Biodegradation **5**: 237–248.
941 doi:10.1007/BF00696462
942 35. **Studer A, McAnulla C, Buchele R, Leisinger T, Vuilleumier S.** 2002. Chloromethane-
943 induced genes define a third C_1 utilization pathway in *Methylobacterium*
944 *chloromethanicum* CM4. J Bacteriol **184**: 3476–3484. doi:10.1128/JB.184.13.3476-
945 3484.2002
946 36. **Schmid-Appert M, Zoller K, Traber H, Vuilleumier S, Leisinger T.** 1997. Association of
947 newly discovered IS elements with the dichloromethane utilization genes of
948 methylotrophic bacteria. Microbiology **143**: 2557–2567. doi:10.1099/00221287-143-8-
949 2557
950 37. **Nakagawa T, Mitsui R, Tani A, Sasa K, Tashiro S, Iwama T, *et al*.** 2012. A catalytic role of
951 XoxF1 as La^{3+}-dependent dethanol dehydrogenase in *Methylobacterium extorquens* strain
952 AM1. Battista JR, editor. PLoS ONE **7**: e50480. doi:10.1371/journal.pone.0050480
953 38. **Skovran E, Palmer AD, Rountree AM, Good NM, Lidstrom ME.** 2011. XoxF is required for
954 expression of methanol dehydrogenase in *Methylobacterium extorquens* AM1. J Bacteriol
955 **193**: 6032–6038. doi:10.1128/JB.05367-11
956 39. **Neufeld JD, Chen Y, Dumont MG, Murrell JC.** 2008. Marine methylotrophs revealed by
957 stable-isotope probing, multiple displacement amplification and metagenomics:
958 metagenomics of active marine methyotrophs. Environ Microbiol **10**: 1526–1535.
959 doi:10.1111/j.1462-2920.2008.01568.x
960 40. **Kalyuzhnaya MG, Hristova KR, Lidstrom ME, Chistoserdova L.** 2008. Characterization of a
961 novel methanol dehydrogenase in representatives of *Burkholderiales*: implications for
962 environmental detection of methylotrophy and evidence for convergent evolution. J
963 Bacteriol **190**: 3817–3823. doi:10.1128/JB.00180-08
964 41. **Vannelli T, Messmer M, Studer A, Vuilleumier S, Leisinger T.** 1999. A corrinoid-dependent
965 catabolic pathway for growth of a *Methylobacterium* strain with chloromethane. Proc Natl
966 Acad Sci **96**: 4615–4620. doi:10.1073/pnas.96.8.4615

967 42. **Šmejkalová H, Erb TJ, Fuchs G.** 2010. Methanol assimilation in *Methylobacterium*
968 *extorquens* AM1: demonstration of all enzymes and their regulation. PLoS ONE **5**: e13001.
969 doi:10.1371/journal.pone.0013001
970 43. **Marx CJ, Van Dien SJ, Lidstrom ME.** 2005. Flux analysis uncovers key role of functional
971 redundancy in formaldehyde metabolism. PLoS Biol.3: e16.
972 doi:10.1371/journal.pbio.0030016
973 44. **Chistoserdova L, Laukel M, Portais J-C, Vorholt JA, Lidstrom ME.** 2004. Multiple formate
974 dehydrogenase enzymes in the facultative methylotroph *Methylobacterium extorquens*
975 AM1 are dispensable for growth on methanol. J Bacteriol **186**: 22–28.
976 doi:10.1128/JB.186.1.22-28.2004
977 45. **Chistoserdova L, Crowther GJ, Vorholt JA, Skovran E, Portais J-C, Lidstrom ME.** 2007.
978 Identification of a fourth formate dehydrogenase in *Methylobacterium extorquens* AM1
979 and confirmation of the essential role of formate oxidation in methylotrophy. J Bacteriol
980 **189**: 9076–9081. doi:10.1128/JB.01229-07
981 46. **Kayser MF, Ucurum Z, Vuilleumier S.** 2002. Dichloromethane metabolism and C_1
982 utilization genes in *Methylobacterium* strains. Microbiology **148**: 1915–1922.
983 doi:10.1099/00221287-148-6-1915
984 47. **Engelke T, Jording D, Kapp D, Pühler A.** 1989. Identification and sequence analysis of the
985 *Rhizobium meliloti dctA* gene encoding the C_4-dicarboxylate carrier. J Bacteriol **171**: 5551–
986 5560. doi:10.1007/BF00284690
987 48. **Ge X, Wang R, Ma J, Liu Y, Ezemaduka AN, Chen PR, *et al*.** 2014. DegP primarily functions
988 as a protease for the biogenesis of β-barrel outer membrane proteins in the Gram-negative
989 bacterium *Escherichia coli*. FEBS J **281**: 1226–1240. doi:10.1111/febs.12701
990 49. **Belogurov GA.** 2002. A Lysine Substitute for K^+. J Biol Chem **277**: 49651–49654.
991 doi:10.1074/jbc.M210341200
992 50. **Mirete S, Mora-Ruiz MR, Lamprecht-Grandío M, de Figueras CG, Rosselló-Móra R,**
993 **González-Pastor JE.** 2015. Salt resistance genes revealed by functional metagenomics from
994 brines and moderate-salinity rhizosphere within a hypersaline environment. Front
995 Microbiol **6**. doi:10.3389/fmicb.2015.01121
996 51. **Nadalig T, Greule M, Bringel F, Keppler F, Vuilleumier S.** 2014. Probing the diversity of
997 chloromethane-degrading bacteria by comparative genomics and isotopic fractionation.
998 Front Microbiol **5**. doi:10.3389/fmicb.2014.00523

52. **Trausch JJ, Ceres P, Reyes FE, Batey RT.** 2011. The structure of a tetrahydrofolate-sensing riboswitch reveals two ligand binding sites in a single aptamer. Structure **19**: 1413–1423. doi:10.1016/j.str.2011.06.019

53. **Vitreschak AG.** 2003. Regulation of the vitamin B_{12} metabolism and transport in bacteria by a conserved RNA structural element. RNA **9**: 1084–1097. doi:10.1261/rna.5710303

54. **Bassford PJ, Kadner RJ.** 1977. Genetic analysis of components involved in vitamin B_{12} uptake in *Escherichia coli*. J Bacteriol **132**: 796–805. 10.1371/journal.pone.0062382

55. **Yu T-Y, Mok KC, Kennedy KJ, Valton J, Anderson KS, Walker GC, *et al*.** 2012. Active site residues critical for flavin binding and 5,6-dimethylbenzimidazole biosynthesis in the flavin destructase enzyme BluB. Protein Sci **21**: 839–849. doi:10.1002/pro.2068

56. **Mandal M.** 2004. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. Science. 306: 275–279. doi:10.1126/science.1100829

57. **Chou H-H, Marx CJ, Sauer U.** 2015. Transhydrogenase promotes the robustness and evolvability of *Escherchia coli* deficient in NADPH production. PLoS Genet **11**: e1005007–e1005007. doi:10.1371/journal.pgen.1005007

58. **Bielnicki J, Devedjiev Y, Derewenda U, Dauter Z, Joachimiak A.** 2005. Derewenda ZS. *Bascillus subtilis* ykuD protein at 2.0 Å resolution: insights into the structure and function of a novel, ubiquitous family of bacterial enzymes. Proteins Struct Funct Bioinforma **62**: 144–151. doi:10.1002/prot.20702

59. **Mc Carthy T V, Lindahl T.** 1985. Methyl pbosphotnesters in alkylated DNA are repaired by the Ada regulatory protein of *Escherichia coli*. Nucleic Acids Research **13**: 2683–2698.

60. **Kitagawa M, Miyakawa M, Matsumura Y, Tsuchido T.** 2002. *Escherichia coli* small heat shock proteins, IbpA and IbpB, protect enzymes from inactivation by heat and oxidants. Eur J Biochem **269**: 2907–2917. doi:10.1046/j.1432-1033.2002.02958.x

61. **Sluis MK, Larsen RA, Krum JG, Anderson R, Metcalf WW, Ensign SA.** 2002. Biochemical, molecular, and genetic analyses of the acetone carboxylases from *Xanthobacter autotrophicus* strain Py2 and *Rhodobacter capsulatus* strain B10. J Bacteriol **184**: 2969–2977. doi:10.1128/JB.184.11.2969-2977.2002

62. **Waack S, Keller O, Asper R, Brodag T, Damm C, Fricke WF, *et al*.** 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC Bioinformatics **7**: 142.

63. **Farhan Ul Haque M, Nadalig T, Bringel F, Schaller H, Vuilleumier S.** 2013. Fluorescence-based bacterial bioreporter for specific detection of methyl halide emissions in the environment. Appl Environ Microbiol. 79: 6561–6567. doi:10.1128/AEM.01738-13

64. **Michener, J.K., Vuilleumier, S., Bringel, F., Marx, C.J.** submitted. Effectiveness of heterologous catabolism of chloromethane and dichloromethane are uncorrelated in *Methylobacterium* strain. Front Microbiol

65. **Michener JK, Vuilleumier S, Bringel F, Marx CJ.** 2014. Phylogeny poorly predicts the utility of a challenging horizontally transferred gene in *Methylobacterium* strains. J Bacteriol **196**: 2101–2107. doi:10.1128/JB.00034-14

66. **Okubo Y, Skovran E, Guo X, Sivam D, Lidstrom ME.** 2007. Implementation of microarrays for *Methylobacterium extorquens* AM1. OMICS J Integr Biol **11**: 325–340. doi:10.1089/omi.2007.0027

67. **Steinmetz PA, Wörner S, Unden G.** 2014. Differentiation of DctA and DcuS function in the DctA/DcuS sensor complex of *E scherichia coli* : function of DctA as an activity switch and of DcuS as the C_4 -dicarboxylate sensor: DctA and DcuS functional differentiation in DctA/DcuS. Mol Microbiol **94**: 218–229. doi:10.1111/mmi.12759

68. **Delmotte N, Knief C, Chaffron S, Innerebner G, Roschitzki B, Schlapbach R, *et al*.** 2009. Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. Proc Natl Acad Sci **106**: 16428–16433. doi:10.1073/pnas.0905240106

69. **Nunn DN, Lidstrom ME.** 1986. Phenotypic characterization of 10 methanol oxidation mutant classes in *Methylobacterium* sp. strain AM1. J Bacteriol **166**: 591–597. doi:0021-9193/86/050591-07$02.00/0

70. **Kallen RG, Jencks WP.** 1966. The mechanism of the condensation of formaldehyde with tetrahydrofolic acid. J Biol Chem **241**: 5851–5863.

71. **Persson B, Hedlund J, Jörnvall H.** 2008. Medium- and short-chain dehydrogenase/reductase gene and protein families: The MDR superfamily. Cell Mol Life Sci **65**: 3879–3894. doi:10.1007/s00018-008-8587-z

72. **Escapa IF, García JL, Bühler B, Blank LM, Prieto MA.** 2012. The polyhydroxyalkanoate metabolism controls carbon and energy spillage in *Pseudomonas putida*. Environ Microbiol **14**: 1049–1063. doi:10.1111/j.1462-2920.2011.02684.x

73. **Hervas AB, Canosa I, Santero E.** 2008. Transcriptome analysis of *Pseudomonas putida* in response to nitrogen availability. J Bacteriol **190**: 416–420. doi:10.1128/JB.01230-07

74. **Carroll SM, Marx CJ.** 2013. Evolution after introduction of a novel metabolic pathway consistently leads to restoration of wild-type physiology. Schneider D, editor. PLoS Genet **9**: e1003427. doi:10.1371/journal.pgen.1003427

75. **Vannelli T, Studer A, Kertesz M, Leisinger T.** 1998. Chloromethane metabolism by *Methylobacterium* sp. strain CM4. Appl Environ Microbiol **64**: 1933–1936. doi:05/1998; 64(5):1933-6.

76. **Chen F, Freedman DL, Falta RW, Murdoch LC.** 2012. Henry's law constants of chlorinated solvents at elevated temperatures. Chemosphere **86**: 156–165. doi:10.1016/j.chemosphere.2011.10.004

77. **Yang Y-X, Wang M-M, Yin Y-L, Onac E, Zhou G-F, Peng S, *et al*.** 2015. RNA-seq analysis reveals the role of red light in resistance against *Pseudomonas syringae* pv. tomato DC3000 in tomato plants. BMC Genomics **16**: 120. doi:10.1186/s12864-015-1228-7

78. **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, *et al*.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23**: 2947–2948. doi:10.1093/bioinformatics/btm404

79. **Tamura K, Dudley J, Nei M, Kumar S.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol **24**: 1596–1599

FIG 1 Metabolism of chlorinated C_1 compounds chloromethane and dichloromethane and of methanol in *M. extorquens* CM4 and DM4

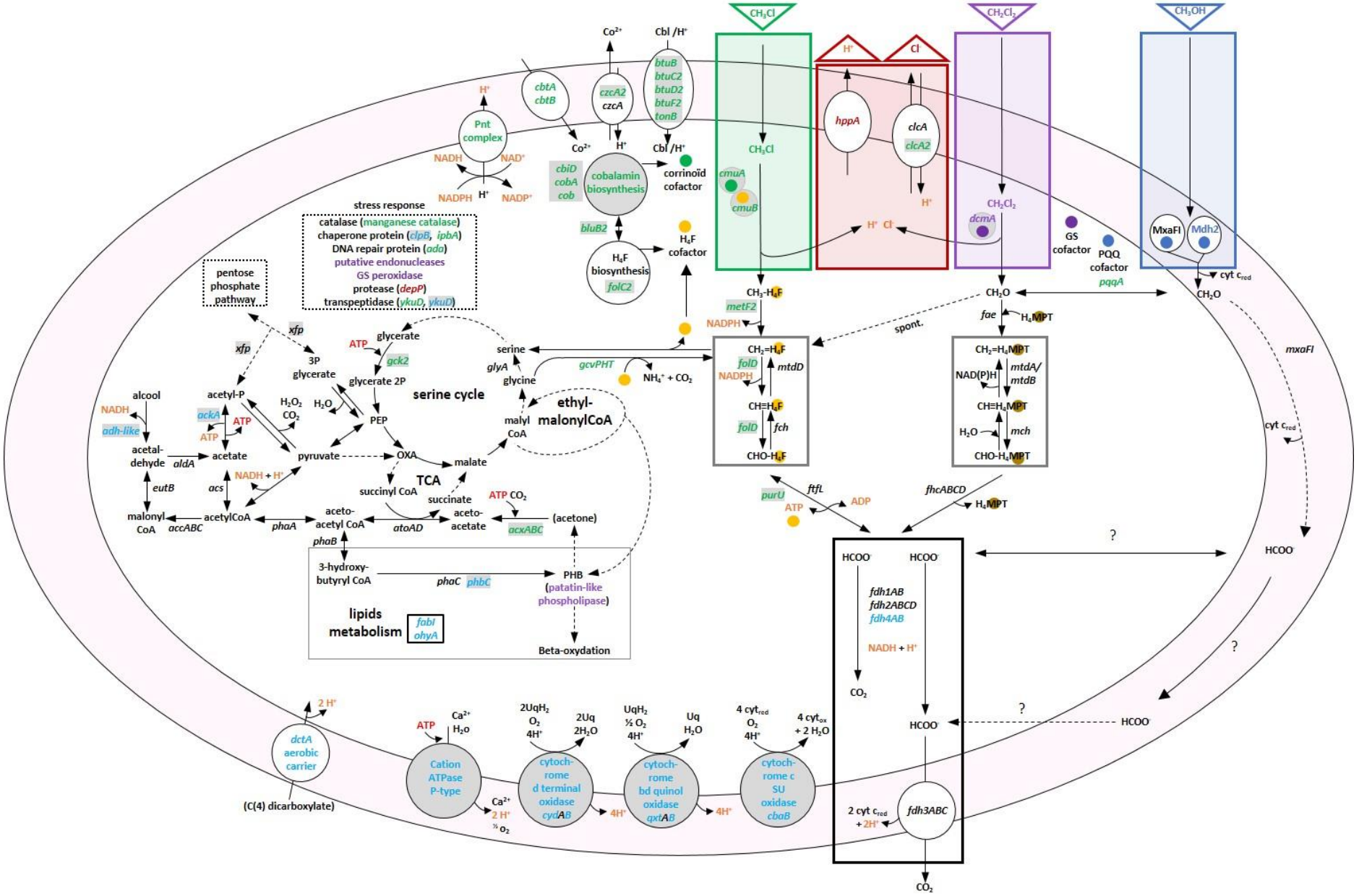


FIG 2 Overview of genomic islands and gene synteny in genomes of *M. extorquens* CM4 and DM4

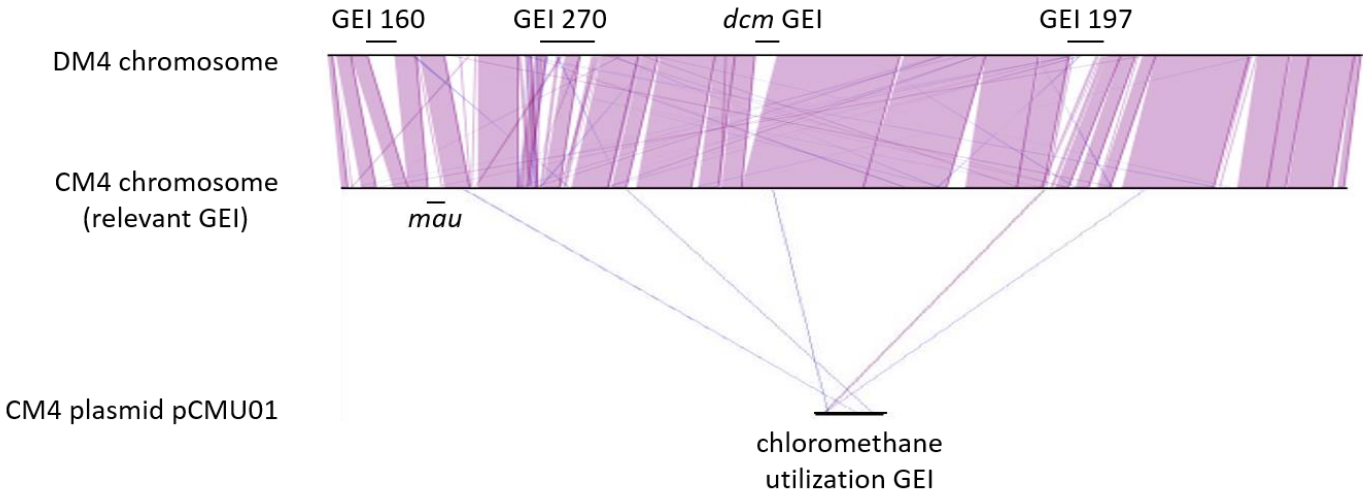


FIG 3 Core, variable and specific genome of five strains affiliated to the species *Methylobacterium extorquens*

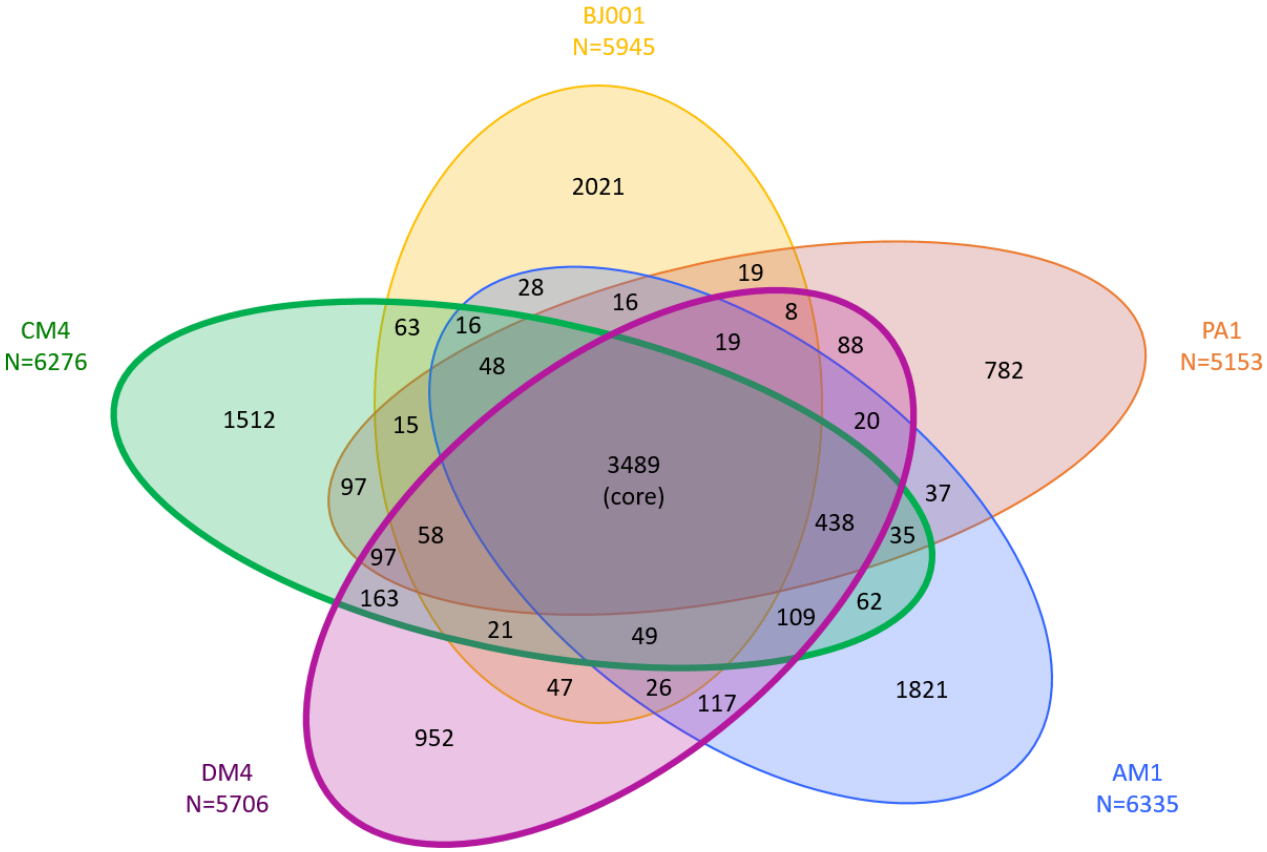


FIG 4 Carbon source-dependent shared categories of gene transcript abundance between *M. extorquens* CM4 and DM4 genomes

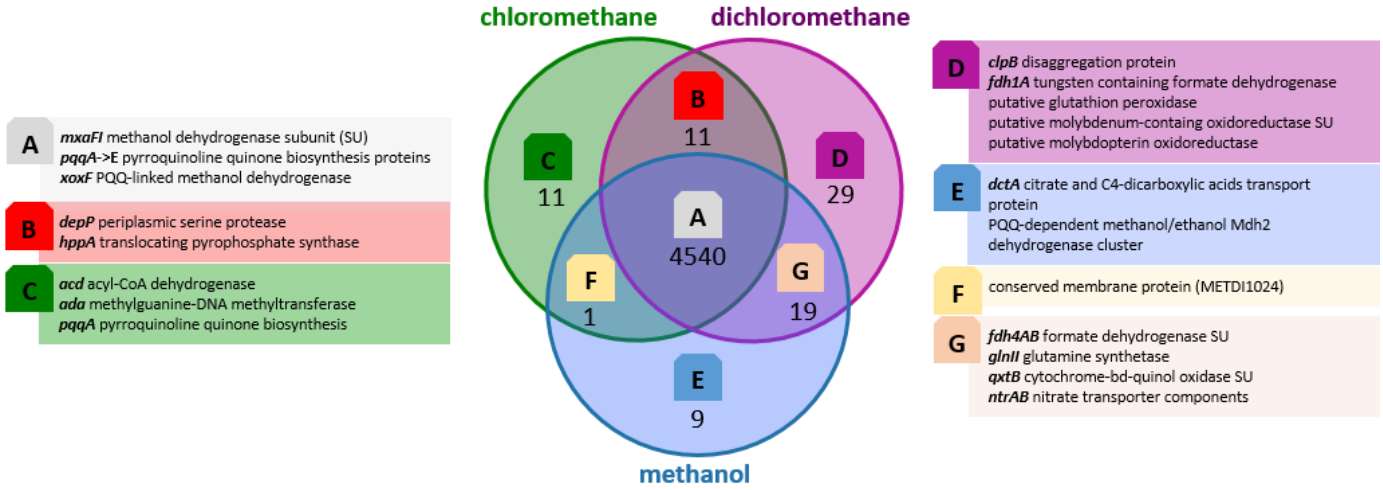


FIG 5 Involvement of plasmid pCMU01 in the transcriptional response to chloromethane

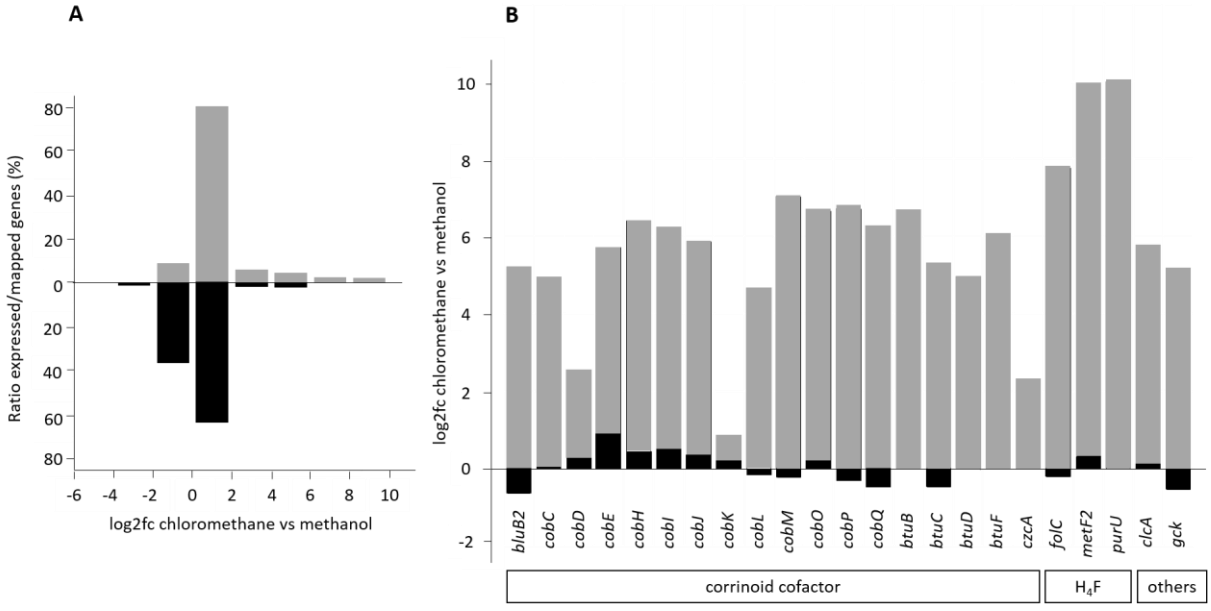


FIGURE S1 Comparison of genome wide gene transcription between *M. extorquens* strains CM4 and DM4 grown with methanol

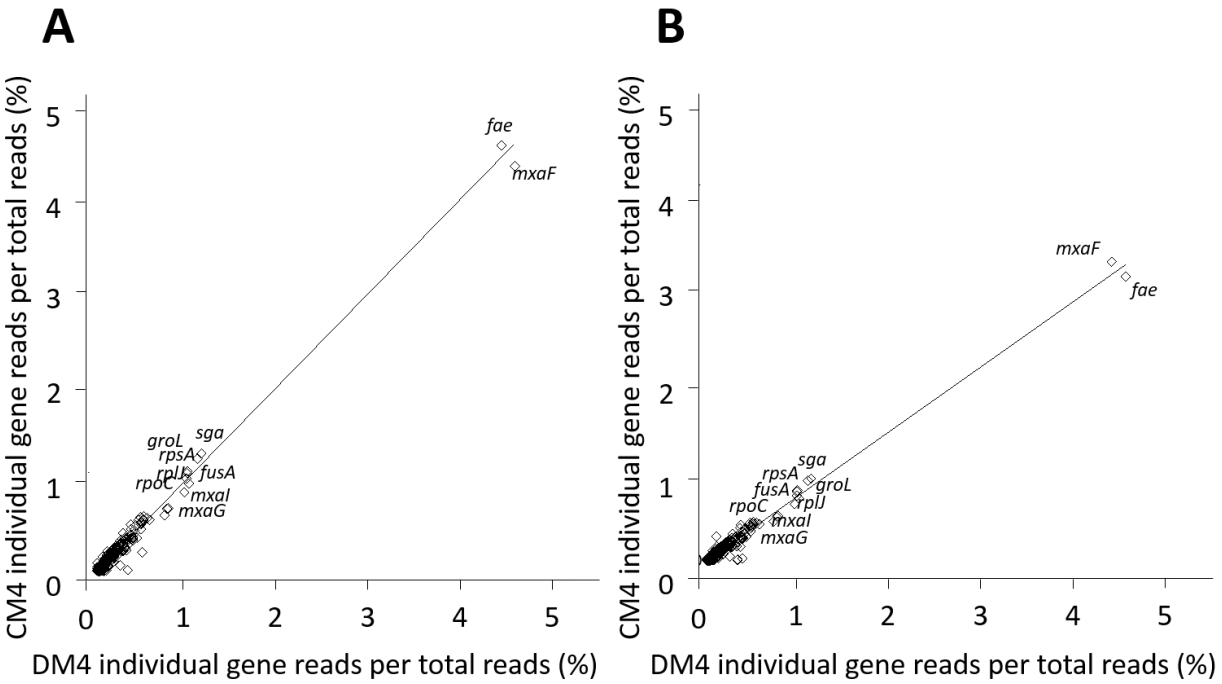


FIGURE S2 Genes with highest differential transcription in cultures grown with chlorinated methanes compared to methanol

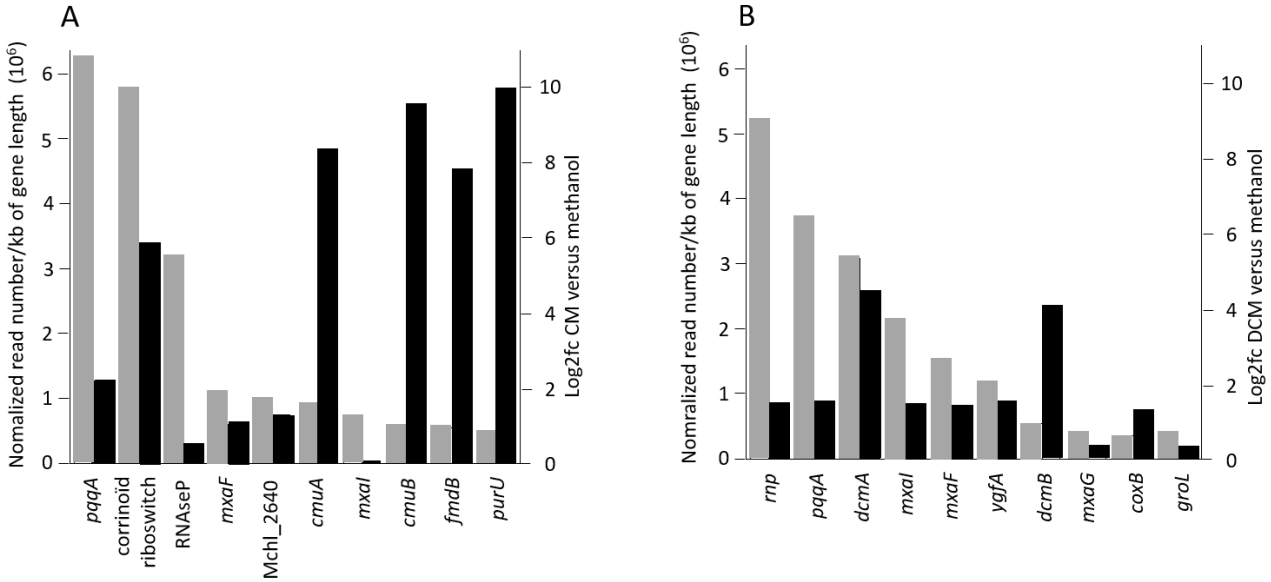


FIGURE S3 Transcription profiles and gene redundancy of key gene clusters more highly transcribed in methanol- than dichloromethane-grown cultures of *M. extorquens* DM4

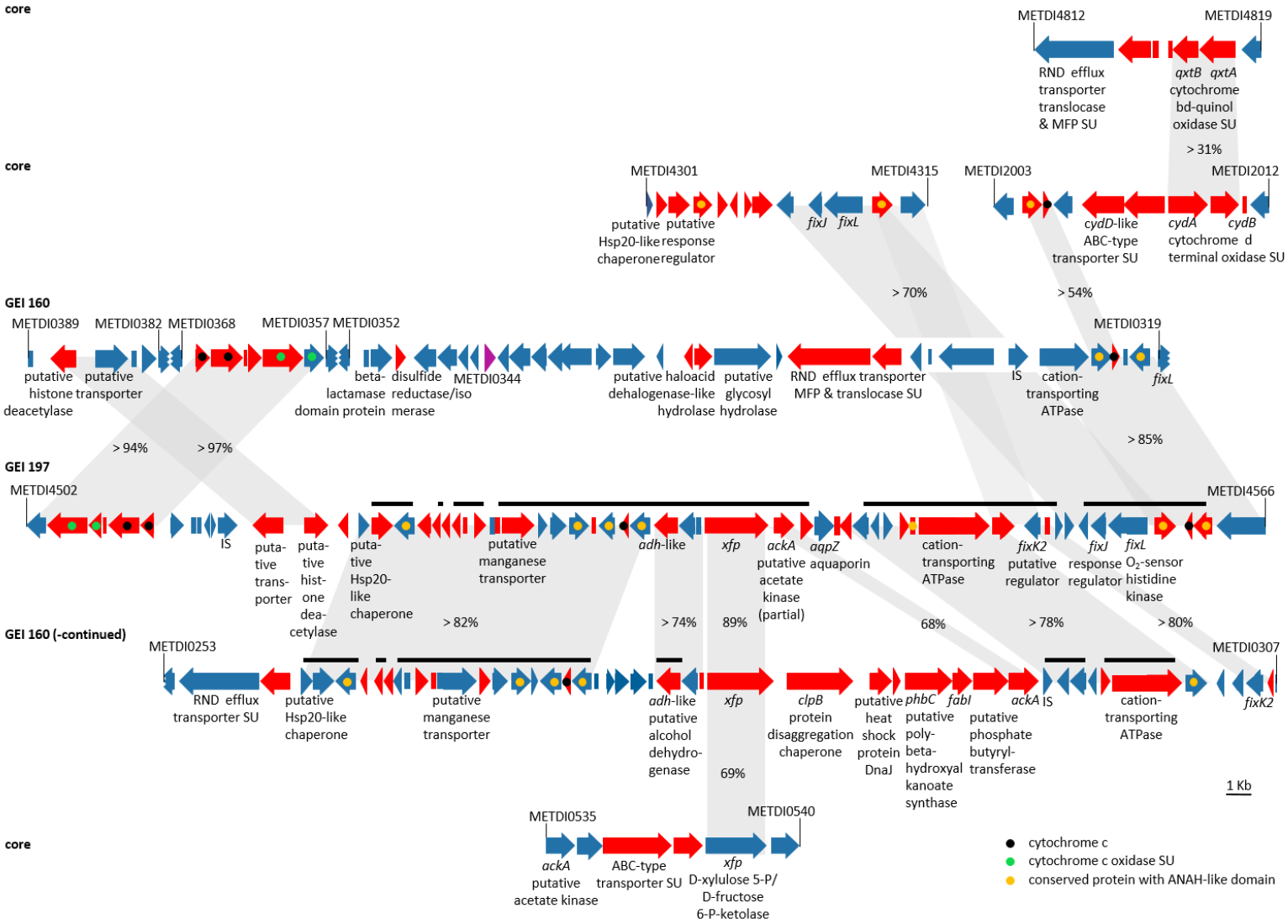
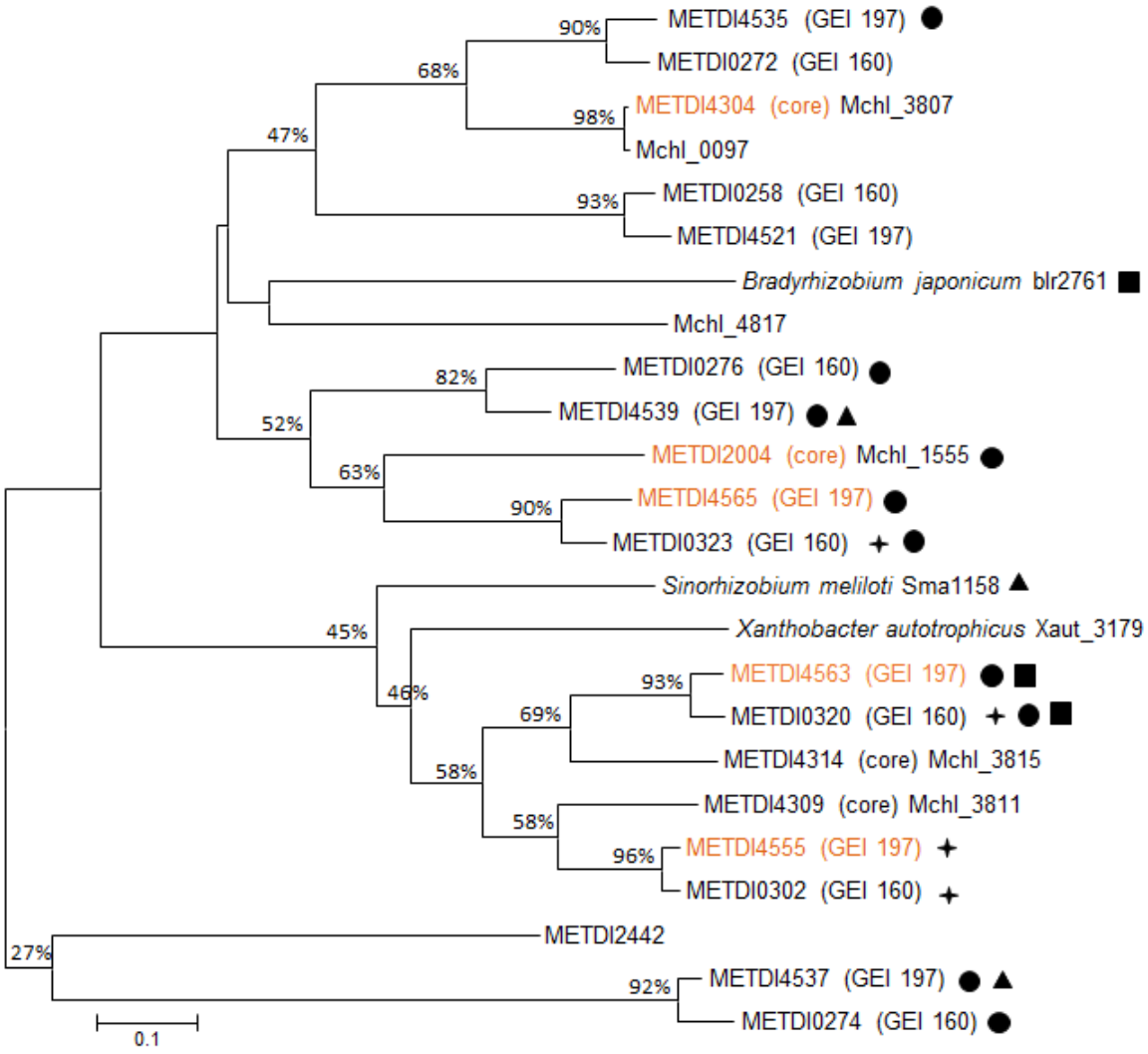


FIGURE S4 Phylogenetic tree of proteins with adenine nucleotide alpha hydrolase (ANAH) domains in *M. extorquens* strains CM4 and DM4



S1 Table. Set of 55 genes used for RNAseq data normalization

Label[a]		Gene	Product
Strain CM4	Strain DM4		
Mchl_0002	METDI0001	*dnaA*	DNA replication initiator protein
Mchl_0220	METDI4779	-	putative phosphodiesterase
Mchl_0372	METDI0202	*glnE*	glutamate-ammonia-ligase adenylyltransferase
Mchl_0480	METDI0517	*zwf*	glucose-6-phosphate 1-dehydrogenase
Mchl_0641	METDI0609	-	conserved protein of unknown function
Mchl_0710	METDI0680	-	conserved protein of unknown function
Mchl_0713	METDI0683	*qxtB*	cytochrome bd ubiquinol oxidase, subunit II
Mchl_0917	METDI1100	*prmC*	protein-(glutamine-N5) methyltransferase, release factor-specific
Mchl_1568	METDI2018	-	putative CelB-like protein
Mchl_1668	METDI2051	-	conserved protein of unknown function precursor
Mchl_1859	METDI2245	*trmE*	tRNA modification GTPase trmE
Mchl_2123	METDI2469	*lysA*	diaminopimelate decarboxylase
Mchl_2201	METDI2686	-	conserved protein of unknown function
Mchl_2204	METDI2689	-	transglutaminase family protein cysteine peptidase
Mchl_2251	METDI2738	-	putative DNA topoisomerase I
Mchl_2277	METDI2766	*gmk*	guanylate kinase
Mchl_2458	METDI2953	*adk*	adenylate kinase
Mchl_2482	METDI2981	*map*	methionine aminopeptidase
Mchl_2549	METDI3056	*rnc*	ribonuclease III
Mchl_2552	METDI3059	*rsuA*	16S rRNA pseudouridylate 516 synthase
Mchl_2581	METDI3088	*ung*	uracil-DNA glycosylase superfamily protein
Mchl_2691	METDI3197	-	ribonuclease BN family protein
Mchl_2750	METDI3257	-	putative decarboxylase, thiamine pyrophosphate requiring
Mchl_2817	METDI3327	*thrB*	homoserine kinase
Mchl_2824	METDI3334	*recA*	RecA protein
Mchl_3177	METDI3720	*lpxC*	UDP-3-O-acyl N-acetylglucosamine deacetylase
Mchl_3288	METDI3840	*gatB*	aspartyl/glutamyl-tRNA amidotransferase subunit B
Mchl_3478	METDI3942	-	ATP-dependent DNA helicase
Mchl_3479	METDI3943	-	conserved protein of unknown function
Mchl_3486	METDI3950	*ribE*	6,7-dimethyl-8-ribityllumazine synthase
Mchl_3504	METDI3967	*mutY*	A/G-specific adenine glycosylase
Mchl_3512	METDI3975	-	putative adenylate cyclase protein
Mchl_3524	METDI3989	*rimI*	ribosomal-protein-alanine N-acetyltransferase
Mchl_3541	METDI4005	*murI*	glutamate racemase
Mchl_3578	METDI4045	*thrC*	threonine synthase
Mchl_3610	METDI4078	*nhaA*	sodium:proton antiporter
Mchl_3635	METDI4104	*pyrB*	aspartate carbamoyltransferase
Mchl_3693	METDI4186	-	conserved protein of unknown function
Mchl_3946	METDI4648	*fadD*	acyl-CoA synthetase
Mchl_4052	METDI4721	-	tonB-dependent siderophore receptor protein
Mchl_4185	METDI4861	*prs*	phosphoribosylpyrophosphate synthetase
Mchl_4253	METDI4871	*gid*	tRNA uridine 5-carboxymethylaminomethyl modification enzyme
Mchl_4306	METDI4930	*dnaB*	replicative DNA helicase DnaB
Mchl_4344	METDI4968	-	putative cysteine desulfurase pyridoxal P-dependent
Mchl_4404	METDI5028	*ftsK*	cell division protein
Mchl_4487	METDI5112	-	conserved protein of unknown function
Mchl_4851	METDI5422	*prfC*	peptide chain release factor RF-3,GTP-binding factor
Mchl_4855	METDI5427	*tuf*	elongation factor Tu (EF-Tu)
Mchl_4914	METDI5486	-	putative ATP-independent RNA helicase
Mchl_5105	METDI5692	*dnlJ*	DNA ligase, NAD(+)-dependent
Mchl_5143	METDI5731	*lpxB*	lipid-A-disaccharide synthase
Mchl_5159	METDI5748	*valS*	valine tRNA synthetase
Mchl_5243	METDI5833	*pbpC*	penicillin-binding protein
Mchl_5271	METDI5861	*hemA*	5-aminolevulinic acid synthase
Mchl_5315	METDI5905	-	conserved protein of unknown function

[a] MaGe annotation (https://www.genoscope.cns.fr/agc/microscope) of genes with constitutive transcript abundance (log2fc value between -0.9 and 1.2 of normalized read numbers in cultures grown with chlorinated methanes compared to methanol.

S2 Table. Groups of genes with similar transcription shared between *M. extorquens* strains CM4 and DM4

Group and label[a]		Name	Product	log2fc[b]		Adjusted p-value[c]	
CM4	DM4			CM4	DM4	CM4	DM4
Group B: more abundant on both chlorinated methanes							
Mchl_0651	METDI0620	-	conserved protein of unknown function	2.3	3.8	< 0.001	< 0.001
Mchl_1347	METDI1765	-	conserved protein of unknown function	2.1	3.3	< 0.001	0.004
Mchl_2116	METDI2461	-	conserved protein of unknown function	2.6	2.5	< 0.001	0.016
Mchl_2560	METDI3067	-	putative methyl-accepting chemotaxis sensory transducer	2.0	2.5	< 0.001	0.009
Mchl_2793	METDI3303	-	protein of unknown function	2.7	2.8	< 0.001	< 0.001
Mchl_3129	METDI3671	-	conserved protein of unknown function, DUF1328	2.5	2.7	< 0.001	< 0.001
Mchl_3408	METDI3863	*hppa*	H^+ translocating pyrophosphate synthase	2.4	3.5	< 0.001	0.019
Mchl_3885	METDI4584	-	protein of unknown function	2.2	2.5	< 0.001	0.006
Mchl_3919	METDI4618	-	conserved protein of unknown function	2.4	2.9	< 0.001	< 0.001
Mchl_4560	METDI5190	-	protein of unknown function	2.1	3.8	< 0.001	< 0.001
Mchl_5157	METDI5746	*depP*	periplasmic serine protease (DegP)	4.6	6.6	< 0.001	< 0.001
Group C: more abundant on chloromethane							
Mchl_1416	METDI1841	-	putative membrane protein	2.2	1.4	< 0.001	0.071
Mchl_1525	METDI1969	-	conserved protein of unknown function	2.0	1.9	< 0.001	0.006
Mchl_2287	METDI2777	-	conserved protein of unknown function	2.2	1.9	< 0.001	0.083
Mchl_3132	METDI3674	-	protein of unknown function	2.2	1.9	< 0.001	0.021
Mchl_4484	METDI5109	-	protein of unknown function	4.2	1.7	< 0.001	0.059
Mchl_5057	METDI5648	-	putative seryl-tRNA synthetase	2.2	1.6	< 0.001	0.072
Mchl_5058	METDI5649	*acd*	acyl-CoA dehydrogenase	2.2	1.7	< 0.001	0.036
MCv2_5149	METDI2503	*pqqA*	coenzyme PQQ biosynthesis protein A	2.1	1.5	< 0.001	0.063
Group D: more abundant on dichloromethane							
Mchl_0009	METDI0008	-	protein of unknown function	1.9	2.4	< 0.001	< 0.001
Mchl_0154	METDI0091	*iorA*	isoquinoline 1-oxidoreductase, alpha subunit	0.9	2.5	0.014	< 0.001
Mchl_0155	METDI0092	-	oxidoreductase, molybdenum cofactor binding subunit	1.5	2.9	< 0.001	< 0.001
Mchl_0156	METDI0093	-	voltage dependent anion channel protein	1.9	2.7	< 0.001	0.001
Mchl_0171	METDI0108	-	protein of unknown function	1.0	2.0	0.017	0.011
Mchl_0189	METDI0128	-	inositol-1-phosphate synthase	1.7	2.1	< 0.001	0.003
Mchl_0190	METDI0129	-	putative dTDP-glucose 4,6-dehydratase	1.8	2.2	< 0.001	0.002
Mchl_0194	METDI0133	-	putative glycosyl transferase	1.3	2.6	0.048	0.006
Mchl_0360	METDI0190	-	putative glutathione peroxidase	1.1	2.5	0.002	0.009
Mchl_1112	METDI1593	-	transposase of IS*Mex11*, IS*3* family	0.7	2.1	0.410	0.260
Mchl_1492	METDI1934	-	protein of unknown function	1.7	2.2	< 0.001	0.014
Mchl_1516	METDI1959	-	conserved protein of unknown function precursor	0.7	2.5	0.047	0.003
Mchl_1729	METDI2115	-	conserved protein of unknown function	0.1	3.0	0.920	< 0.001
Mchl_2091	METDI2432	-	putative aminoglycoside phosphotransferase	1.7	2.7	< 0.001	< 0.001
Mchl_2208	METDI2693	-	putative molybdopterin oxidoreductase	1.9	3.8	< 0.001	0.008
Mchl_2537	METDI3043	-	conserved protein of unknown function with 2 CBS domains	1.0	4.4	0.003	0.002
Mchl_2848	METDI3359	-	protein of unknown function	1.1	2.1	0.029	0.002
Mchl_2894	METDI3426	-	conserved protein of unknown function	1.3	2.1	< 0.001	0.006
Mchl_3030	METDI3569	*arcB*	ornithine cyclodeaminase	1.6	2.6	< 0.001	< 0.001
Mchl_3921	METDI4620	-	protein of unknown function	1.2	2.0	0.026	0.002
Mchl_3964	METDI4670	-	putative monooxygenase with ATPase activity	1.8	2.0	< 0.001	0.003
Mchl_4030	METDI4699	-	transcriptional regulator, AraC family	-0.6	3.2	< 0.001	< 0.001
Mchl_4335	METDI4959	-	conserved protein of unknown function	1.5	2.2	0.001	< 0.001
Mchl_4440	METDI5067	-	conserved protein of unknown function	1.9	4.3	0.098	0.002
Mchl_4491	METDI5117	-	protein of unknown function	1.2	3.5	0.007	< 0.001
Mchl_4727	METDI1224	-	putative NcrB-like regulator, DUF156	0.8	2.8	0.059	< 0.001
Mchl_4728	METDI1223	-	putative NcrA-like major facilitator superfamily permease	0.9	2.1	0.034	0.009
Mchl_5285	METDI5875	-	putative photosynthesis gene regulator (*bchF-crtJ*)	1.1	2.5	0.001	0.037
Mchl_5301	METDI5891	-	putative endonuclease	1.5	3.2	< 0.001	< 0.001
Group E: more abundant on methanol							
Mchl_1165	METDI1517	-	conserved membran protein of unknown function	-3.1	-2.3	< 0.001	< 0.001
Mchl_1533	METDI1978	-	conserved protein of unknown function	-2.9	-2.1	< 0.001	0.037
Mchl_1534	METDI1979	-	putative NosX precursor (nitrous oxide reductase)	-3.3	-3.2	< 0.001	0.013
Mchl_1535	METDI1980	-	putative regulatory protein NosR	-3.4	-2.6	< 0.001	0.019

Mchl_1536	METDI1981	-	conserved protein of unknown function	-3.7	-2.5	< 0.001	< 0.001
Mchl_1537	METDI1982	-	conserved exported protein of unknown function	-3.3	-3.3	< 0.001	< 0.001
Mchl_1538	METDI1983	-	cytochrome c550 protein, putative PQQ-dependent methanol/ethanol oxidation system	-3.9	-3.8	< 0.001	< 0.001
Mchl_1539	METDI1984	-	conserved protein of unknown function, periplasmic binding protein precursor, putative periplasmic binding protein-like II	-4.5	-3.8	< 0.001	< 0.001
Mchl_3283	METDI3835	*dctA*	C_4-dicarboxylate transport protein	-2.9	-3.2	< 0.001	0.029
Group F: more abundant on methanol in strain CM4 only							
Mchl_0844	METDI1024	-	conserved membran protein of unknown function	-2.0	-1.8	< 0.001	0.031
Group G: more abundant on methanol in strain DM4 only							
Mchl_0500	METDI0537	-	ABC transporter, membrane protein	1.6	-2.5	< 0.001	0.059
Mchl_0501	METDI0538	-	ABC transporter, permease	1.1	-2.0	0.015	0.073
Mchl_1243	METDI1697	-	putative formate/nitrate transporter	-1.6	-3.4	< 0.001	< 0.001
Mchl_1337	METDI1753	*eshA*	eshA	-1.1	-3.3	< 0.001	< 0.001
Mchl_1338	METDI1754	-	putative cysteine desulfurase (SufS domain)	-1.1	-2.2	< 0.001	0.094
Mchl_1341	METDI1761	-	conserved exported protein of unknown function	-1.4	-2.6	< 0.001	< 0.001
Mchl_1717	METDI2103	-	conserved protein of unknown function	2.5	-2.4	< 0.001	0.003
Mchl_2380	METDI2873	*fdh4B*	formate dehydrogenase subunit B	-1.2	-3.2	0.001	< 0.001
Mchl_2381	METDI2874	*fdh4A*	formate dehydrogenase subunit A	-0.7	-4.0	0.048	< 0.001
Mchl_2653	METDI3157	*glnII*	glutamine synthetase II	-0.8	-3.2	0.081	0.065
Mchl_3806	METDI4303	-	putative transcriptional regulator	1.5	-3.6	< 0.001	0.048
Mchl_3807	METDI4304	-	conserved protein with ANAH-like domain	1.1	-3.0	0.004	0.054
Mchl_3808	METDI4305	-	conserved protein of unknown function	0.8	-2.4	0.069	0.039
Mchl_3810	METDI4307	-	protein of unknown function	0.9	-2.7	0.028	0.034
Mchl_4141	METDI4814	-	protein of unknown function, putative exported protein	0.9	-4.0	0.036	0.001
Mchl_4142	METDI4816	*qxtB*	cytochrome bd-quinol oxidase subunit II	1.2	-2.7	0.008	0.061
Mchl_4650	METDI5284	-	putative membrane of unknown function	-1.6	-3.1	< 0.001	< 0.001
Mchl_4958	METDI5532	*nrtB*	nitrate transport permease protein	-1.1	-2.4	0.017	0.001
Mchl_4959	METDI5533	*nrtA*	nitrate transporter component	-1.1	-2.7	0.007	< 0.001

[a] MaGe annotation (https://www.genoscope.cns.fr/agc/microscope). Group as defined in Fig 3.
[b] log2fc of normalized read numbers in cultures grown with chlorinated methanes compared to methanol.
[c] False discovery rate, only adjusted p-value under 0.1 were considered.

S3 Table. Differentially abundant transcripts in cultures of *M. extorquens* CM4 grown with chloromethane compared to methanol (Table 2 complementary list)

Label [a]	Name	Product	Occurr-ence[b]	Directional RNAseq data	
				log2 fc[c]	Adjusted p-value[d]
Mchl_0472	-	putative PYP-like sensor domain	core	4.8	< 0.001
Mchl_1288	-	putative transcriptional regulator, AraC family	variable	2.2	< 0.001
Mchl_1416	-	putative membrane protein	core	2.2	< 0.001
Mchl_1525	-	protein of unknown function	variable	2.0	< 0.001
Mchl_1533	-	protein of unknown function	core[e]	-2.9	< 0.001
Mchl_1534	-	putative NosX precursor (nitrous oxide reductase)	core[e]	- 3.3	< 0.001
Mchl_1535	-	conserved membrane protein, FMN-binding and 4Fe-4S binding domains	core[e]	-3.4	< 0.001
Mchl_1536	-	conserved protein of unknown function, putative YVTN beta-propeller repeat family protein	core[e]	-3.7	< 0.001
Mchl_1537	-	conserved exported protein of unknown function	core[e]	-3.3	< 0.001
Mchl_1539	-	conserved protein of unknown function, putative periplasmic binding protein-like II	core[e]	-4.5	< 0.001
Mchl_2560	-	putative exported sensory transducer protein	core	2.0	< 0.001
Mchl_2920	-	putative exported sensory transducer protein	variable	-2.4	
Mchl_3129	-	conserved protein of unknown function, DUF1328	core	2.5	< 0.001
Mchl_4776	-	conserved protein of unknown function, putative esterase	specific	2.1	< 0.001
Mchl_4777	-	putative outer-membrane protein	variable	2.1	< 0.001
Mchl_4779	-	conserved protein of unknown function	specific	2.4	< 0.001
Mchl_5060	-	putative monooxygenase, small component; putative flavin:NADH reductase	variable	2.5	< 0.001
Mchl_5347	-	conserved protein of unknown function, DUF892	core	2.1	< 0.001
Mchl_5465	-	transposase (fragment)	pCMU01	2.3	< 0.001
Mchl_5499	-	response regulator in two-component regulatory system, sigma-54 interaction domain	pCMU01	2.4	< 0.001
Mchl_5505	-	integrase catalytic region	pCMU01	2.9	0.018
Mchl_5526	-	transposase of IS*Mex11*, IS*3* family (ORF 2)	pCMU01	2.3	< 0.001
Mchl_5581	-	transposase (fragment)	pCMU01	2.3	0.011
Mchl_5618	-	conserved protein of unknown function (fragment)	pCMU01	2.2	< 0.001
Mchl_5663	-	conserved protein of unknown function	pCMU01	2.1	< 0.001
Mchl_5680	-	protein of unknown function	pCMU01	4.8	< 0.001
Mchl_5684	-	putative transposase (fragment)	pCMU01	4.2	< 0.001
Mchl_5703	-	putative transposase (fragment)	pCMU01	2.4	0.015
Mchl_5704	-	putative transposon (fragment)	pCMU01	3.2	< 0.001
Mchl_5716	-	conserved protein of unknown function, CobS-like domain	pCMU01	7.1	< 0.001
Mchl_R0001	-	tRNA-Thr	core	2.1	< 0.001
Mchl_R0005	-	tRNA-Glu	core	2.3	< 0.001
Mchl_R0006	-	tRNA-Ser	core	2.0	< 0.001
Mchl_R0057	-	tRNA-Thr	core	2.3	< 0.001
Mchl_R0078	-	tRNA-Asn	core	2.1	< 0.001
Mchl_RncRNA 2297962R	-	SRP RNA, RNA component of signal recognition particule	core	2.1	< 0.001
p1MCv2_0094	-	protein of unknown function	pCMU01	2.3	< 0.001
p1MCv2_0095	-	conserved protein of unknown function	pCMU01	2.3	< 0.001
p1MCv2_0379	-	protein of unknown function	pCMU01	2.4	< 0.001
p1MCv2_0380	-	protein of unknown function	pCMU01	2.1	< 0.001

[a] MaGe annotation (https://www.genoscope.cns.fr/agc/microscope).

[b] Genes labeled as core, variable or specific are respectively found in all, at least 2 genomes of *M. extorquens* genomes of strains AM1, BJ001, CM4, DM4 and PA1 (80 % of identity on at least 80 % of the protein length), or only in the chromosome of strain CM4. Genes labeled as pCMU01 are located on the 380 kb plasmid (Roselli *et al*., 2013).

[c] log2 fold-change of normalized read numbers in cultures grown with chloromethane compared to methanol.

[d] False discovery rate, only adjusted p-value under 0.1 were considered.

[e] Methanol-more expressed gene cluster Mchl_1533 to Mchl_1540.

S4 Table. Differentially abundant transcripts in cultures of *M. extorquens* DM4 grown with dichloromethane compared to methanol

Label[a]	Name	Product	Occurr-ence[b]	Directional RNAseq data	
				log2fc[c]	Adjusted p-value [d]
List of genes more abundant on dichloromethane					
METDI0008	-	protein of unknown function	core	2.4	< 0.001
METDI0108	-	protein of unknown function	variable	2.0	0.012
METDI0190	-	putative glutathione peroxidase	core	2.5	0.001
METDI0620	-	conserved exported protein of unknown function	core	3.8	< 0.001
METDI1593	-	transposase of IS*Mex11*, IS*3* family (ORF 2)	variable	2.1	0.009
METDI1597	-	exported protein of unknown function	specific	2.3	0.001
METDI1598	-	protein of unknown function	specific	2.1	0.001
METDI1755	-	conserved protein of unknown function	specific	2.1	0.004
METDI1765	-	protein of unknown function	core	3.3	0.004
METDI1934	-	protein of unknown function	variable	2.2	0.014
METDI1959	-	conserved protein of unknown function precursor	variable	2.5	0.036
METDI2115	-	conserved exported protein of unknown function	variable	3.0	< 0.001
METDI2432	-	putative phosphotransferase	core	2.7	< 0.001
METDI2461	-	protein of unknown function	specific	2.5	0.016
METDI2693	-	putative molybdopterin oxidoreductase	core	3.8	0.008
METDI3043	-	conserved protein with 2 CBS domains	core	4.4	0.002
METDI3067	-	putative methyl-accepting chemotaxis sensory transducer	core	2.5	0.009
METDI3303	-	putative extracellular protein with a Ca binding domain	core	2.8	0.001
METDI3359	-	protein of unknown function	core	2.1	0.002
METDI3426	-	conserved protein of unknown function	core	2.1	0.006
METDI4190	-	exported protein of unknown function	core	2.1	0.048
METDI4320		putative alcohol dehydrogenase (fragment)	specific	2.6	0.064
METDI4584	-	protein of unknown function	core	2.5	0.006
METDI4608	-	conserved membrane protein of unknown function	variable	2.2	0.006
METDI4618	-	conserved protein of unknown function	core	2.9	0.004
METDI4620	-	protein of unknown function	variable	2.2	0.002
METDI4670	-	putative monooxygenase with ATPase activity	core	2.0	0.004
METDI4741	-	putative response regulator, (CheY-like protein)	core	2.0	0.009
METDI4905	-	putative patatin-like phospholipase	core	2.5	< 0.001
METDI4959	-	conserved protein of unknown function	core	2.2	< 0.001
METDI5117	-	protein of unknown function	variable	3.5	< 0.001
METDI5146	*mxaW*	conserved hypothetical protein	core	2.1	0.001
METDI5190	-	protein of unknown function	variable	3.8	0.001
METDI5265	-	protein of unknown function	variable	2.4	0.002
METDI5746	*depP*	periplasmic serine protease	core	6.6	< 0.001
METDI5875	-	putative photosynthesis gene regulator (*bchF-crtJ*)	variable	2.5	0.004
METDI5891	-	putative endonuclease	core	3.2	< 0.001
METDI0091-METDI0093 gene cluster					
METDI0091	-	oxidoreductase, 2Fe-2S subunit	core	2.5	< 0.001
METDI0092	-	oxidoreductase, molybdenum cofactor binding subunit	core	2.9	< 0.001
METDI0093	-	putative voltage-dependent anion channel	core	2.7	0.001
METDI0128-METDI0133 gene cluster					
METDI0128	*ino*	inositol-3-phosphate synthase	core	2.1	0.003
METDI0129	-	putative dTDP-glucose 4,6-dehydratase	core	2.2	0.002
METDI0133	-	putative glycosyl transferase	variable	2.6	0.006
GEI 160					
METDI0344	-	conserved protein of unknown function	variable	2.2	0.012
GEI 270					
METDI1223	-	putative cobalt/nickel resistance NcrA-like major facilitator superfamily permease	variable	2.1	0.009
METDI1224	-	putative cobalt/nickel resistance NcrB-like regulator	variable	2.8	< 0.001
***dcm* GEI**					
METDI2655	*dcmR*	transcriptional repressor of DCM dehalogenase	specific[f]	2.7	0.046
METDI2656	*dcmA*	dichloromethane dehalogenase (DCM dehalogenase)	specific[f]	4.3	< 0.001
METDI2657	-	conserved protein of unknown function DcmB	specific	4.0	< 0.001
METDI2658	-	conserved protein of unknown function DcmC	specific	3.5	< 0.001
METDI3569-METDI3571 gene cluster					

METDI3569	*arcB*	ornithine cyclodeaminase	core	2.6	< 0.001
METDI3671	-	conserved membrane protein of unknown function	core	2.7	< 0.001
		METDI3863-METDI3864 gene cluster			
METDI3863	*hppa*	H^+ translocating pyrophosphate synthase	core	3.5	0.019
METDI3864	-	protein of unknown function	specific	2.4	0.004
		METDI4695-METDI4699 gene cluster			
METDI4695	*proV*	glycine/betaine/proline ABC transporter, ATP-binding component	variable	2.3	0.001
METDI4696	*opuAB*	glycine/betaine/proline ABC transporter, membrane component	variable	2.0	0.016
METDI4699	-	transcriptional regulator, AraC family	core	4.3	< 0.001
		METDI5067-METDI5068 gene cluster			
METDI5067	-	conserved protein of unknown function	core	3.0	0.002
METDI5068	-	putative transcriptional regulator	variable[e]	3.2	0.001
List of genes more abundant on methanol					
METDI1517	-	conserved membrane protein of unknown function	core	-2.3	< 0.001
METDI1697	-	putative formate/nitrate transporter	variable	-3.4	< 0.001
METDI1842	-	conserved protein of unknown function	core	-2.1	0.005
METDI2103	-	conserved protein of unknown function	core	-2.4	0.003
METDI3157	*glnII*	glutamine synthetase, type II	core	-3.2	0.008
METDI3835	*dctA*	C_4-dicarboxylate transport protein	core	-3.2	0.003
METDI4261	-	conserved protein of unknown function	core	-2.6	0.001
METDI5284	-	membrane protein of unknown function	core	-3.1	< 0.001
p2METDI0004	-	conserved protein of unknown function	specific	-3.7	0.002
p2METDI0005	-	protein of unknown function	specific	-3.1	0.005
		METDI1978-METDI1985 gene cluster			
METDI1978	-	conserved protein of unknown function	core	-2.1	0.004
METDI1979	-	conserved protein of unknown function, NosX-related protein	core	-3.2	0.012
METDI1980	-	conserved membrane protein, FMN-binding and 4Fe-4S binding domains	core	-2.6	0.002
METDI1981	-	conserved exported protein of unknown function	core	-2.5	< 0.001
METDI1982	-	conserved exported protein of unknown function	core	-3.3	< 0.001
METDI1983	-	cytochrome c550	core	-3.8	< 0.001
METDI1984	-	conserved periplasmic protein of unknown function	core	-3.8	< 0.001
METDI1985	*mdh2*	PQQ-dependent methanol/ethanol dehydrogenase	core	-1.9	0.301
		GEI 160			
METDI0255	-	RND efflux transporter, MFP subunit	specific	-2.1	0.084
METDI0260	-	glucokinase (C-terminal fragment)	variable	-4.3	< 0.001
METDI0262	-	glucokinase (N-terminal fragment)	variable	-2.6	0.016
METDI0263	-	protein of unknown function	specific	-2.7	0.003
METDI0266	-	putative ion transport domain	variable	-3.1	< 0.001
METDI0267	-	conserved protein of unknown function	variable	-2.8	0.007
METDI0270	-	protein of unknown function	variable	-2.6	0.094
METDI0275	-	putative cytochrome c	variable	-2.9	0.045
METDI0283	*adh-like*	putative alcohol dehydrogenase	variable	-3.3	0.029
METDI0285	-	protein of unknown function	specific	-3.3	0.048
METDI0286	*xfp*	D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase	variable	-2.9	0.098
METDI0288	*clpB*	protein disaggregation chaperone	core	-3.6	0.014
METDI0289	-	putative heat shock protein DnaJ	specific	-3.9	0.008
METDI0290	-	DnaJ-associated protein of unknown function	specific	-4.0	0.001
METDI0292	*phbC*	putative poly-beta-hydroxyalkanoate synthase	specific	-3.1	0.085
METDI0293	*fabI*	enoyl-[acyl-carrier-protein] reductase (NADH)	specific	-3.6	0.046
METDI0294	-	putative phosphate butyryltransferase	specific	-3.6	0.046
METDI0295	-	putative acetate kinase (AckA)	specific	-3.4	< 0.001
METDI0300	-	protein of unknown function	specific	-2.6	0.029
METDI0301	-	cation-transporting ATPase	variable	-3.3	0.077
METDI0306	-	protein of unknown function	specific	-2.7	< 0.001
METDI0322	-	putative cytochrome c, class I	specific	-2.2	0.034
METDI0329	-	RND efflux transporter, MFP subunit	specific	-3.0	0.062
METDI0330	-	RND efflux transporter, HME family, translocase subunit	variable	-3.0	< 0.001
METDI0333	-	putative haloacid dehalogenase-like hydrolase	specific	-2.7	< 0.001
METDI0334	-	exported protein of unknown function	specific	-2.2	0.017
METDI0349	-	putative protein disulfide reductase/isomerase, putative thioredoxin	specific	-2.4	0.057
METDI0363	-	putative cytochrome c oxidase, subunit I	specific	-2.6	0.011
METDI0364	-	putative cytochrome c oxidase, subunit II (CbaB)	specific	-3.4	< 0.001
METDI0365	-	conserved membane protein of unknown function	specific	-3.3	< 0.001

METDI0366	-	putative cytochrome c	specific	-2.9	0.001
METDI0367	-	putative cytochrome c	specific	-2.7	< 0.001
METDI0384	-	protein of unknown function	specific	-2.5	0.004
METDI0386	-	histone deacetylase family protein	variable	-2.9	0.029
METDI0537-METDI0538 gene cluster[e]					
METDI0537	-	ABC transporter, fused tandem ATPase and permease domains	core	-2.5	0.059
METDI0538	-	ABC-2 transporter, permease	core	-2.0	0.073
GEI 270					
METDI1414	*ardC*	antirestriction protein (ArdC)	specific	-2.2	0.026
METDI1416	-	transposase of IS*Mdi10*, IS*110* family	specific	-3.7	0.008
METDI1418	-	exported protein of unknown function	specific	-3.5	0.059
METDI1419	-	protein of unknown function	specific	-3.6	0.046
METDI1420	-	protein of unknown function	specific	-2.9	0.084
METDI1446	-	protein of unknown function	specific	-2.8	0.003
METDI1753-METDI1770 gene cluster					
METDI1753	*eshA*	eshA	core	-3.3	< 0.001
METDI1754	-	putative cysteine desulfurase (SufS domain)	variable	-2.2	0.094
METDI1761	-	conserved exported protein of unknown function	variable	-2.6	< 0.001
METDI1762	-	conserved protein of unknown function, DUF1236	variable	-2.1	0.007
METDI1769	-	conserved protein of unknown function	core	-3.7	0.034
METDI1770	*ohyA*	oleate hydratase	variable	-3.4	0.043
METDI2004-METDI2012 gene cluster[e]					
METDI2004	-	conserved protein with ANAH-like domain	core	-3.3	0.046
METDI2005	-	putative cytochrome c, class I	core	-3.2	0.042
METDI2007	-	ABC transporter, fused ATPase and permease domains (CydD-like)	core	-3.4	< 0.001
METDI2009	-	ABC transporter related, fused ATPase and permease domains (CydD-like)	core	-3.3	< 0.001
METDI2011	*cydB*	cytochrome d terminal oxidase, subunit II	core	-3.4	0.036
METDI2012	-	putative exported *cyd* operon protein	specific	-3.7	< 0.001
METDI2867-METDI2874 gene cluster					
METDI2867	*glnK*	nitrogen regulatory protein P-II	core	-2.5	0.029
METDI2873	*fdh4B*	formate dehydrogenase subunit B	core	-3.2	< 0.001
METDI2874	*fdh4A*	formate dehydrogenase subunit A	core	-4.0	< 0.001
METDI4303-METDI4313 gene cluster[e]					
METDI4303	-	putative transcriptional regulator, Crp/Fnr family	variable	-3.6	0.048
METDI4304	-	conserved protein with ANAH-like domain	core	-3.1	0.054
METDI4305	-	conserved protein of unknown function	variable	-2.4	0.008
METDI4307	-	protein of unknown function	variable	-2.7	0.034
METDI4308	-	protein of unknown function	variable	-3.4	< 0.001
METDI4310	-	protein of unknown function	specific	-2.7	0.003
GEI 197					
METDI4470	-	putative L,D-transpeptidase catalytic domain (YkuD)	specific	-2.3	< 0.001
METDI4503	-	putative cytochrome c oxidase, subunit I	specific	-2.4	0.008
METDI4504	-	putative cytochrome c oxidase, subunit II	specific	-3.0	< 0.001
METDI4505	-	conserved membrane protein of unknown function	specific	-2.5	0.034
METDI4506	-	putative cytochrome c, putative exported protein	specific	-2.5	0.006
METDI4507	-	putative cytochrome c	specific	-3.3	< 0.001
METDI4514	-	putative transporter, SulP family	specific	-3.3	0.076
METDI4515	-	histone deacetylase family protein	variable	-3.8	0.033
METDI4517	-	protein of unknown function	specific	-2.5	0.007
METDI4520	-	putative transcriptional regulator, Crp/Fnr family	variable	-3.2	0.072
METDI4522	-	putative glucokinase (C-terminal fragment)	specific	-3.5	< 0.001
METDI4524	-	putative glucokinase (N-terminal fragment)	variable	-2.6	0.035
METDI4525	-	protein of unknown function	specific	-2.8	0.046
METDI4526	-	conserved protein of unknown function	variable	-3.0	0.075
METDI4527	-	putative biotin/lipoyl attachment domain	variable	-2.8	0.054
METDI4528	-	putative ion transport domain	variable	-3.1	< 0.001
METDI4530	-	conserved protein of unknown function	variable	-3.0	0.034
METDI4531	-	putative manganese transporter	variable	-3.3	0.051
METDI4536	-	protein of unknown function	variable	-2.4	0.011
METDI4538	-	putative cytochrome c, class I	variable	-2.8	0.034
METDI4540	*adh-like*	putative alcohol dehydrogenase	variable	-3.7	0.011
METDI4543	*xfp*	D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase	variable	-3.1	0.085
METDI4544	-	putative acetate kinase (partial)	variable	-5.0	< 0.001
METDI4545	-	protein of unknown function	variable	-3.6	0.044

METDI4547	-	protein of unknown function	specific	-3.3	0.003
METDI4548	-	conserved protein of unknown function	variable	-3.2	0.083
METDI4552	-	protein of unknown function	variable	-3.5	0.036
METDI4553	-	exported protein of unknown function	variable	-2.9	0.003
METDI4554	-	putative cation-transporting ATPase (P-type)	variable	-3.5	0.059
METDI4555	-	conserved protein with ANAH-like domain	variable	-3.3	0.001
METDI4557	-	exported protein of unknown function	variable	-3.4	< 0.001
METDI4563	-	conserved protein with ANAH-like domain	variable	-3.3	0.051
METDI4564	-	putative cytochrome c, class I	variable	-2.6	0.003
METDI4565	-	conserved protein with ANAH-like domain	variable	-2.9	0.051
METDI4813-METDI4816 gene cluster					
METDI4813	-	RND efflux transporter, MFP subunit	core	-3.6	< 0.001
METDI4814	-	exported protein of unknown function	variable	-4.0	0.001
METDI4815	-	protein of unknown function	variable	-2.7	0.003
METDI4816	*qxtB*	cytochrome bd-quinol oxidase subunit II	core	-2.7	0.061
METDI5532-METDI5533 gene cluster					
METDI5532	*nrtB*	nitrate transport permease protein	core	-2.4	0.001
METDI5533	*nrtA*	nitrate transporter component	core	-2.7	< 0.001

[a] MaGe annotation (https://www.genoscope.cns.fr/agc/microscope)

[b] Genes labeled as core, variable or specific are respectively found in all, at least 2 genomes of *M. extorquens* genomes of strains AM1, BJ001, CM4, DM4 and PA1 (80 % of identity on at least 80 % of the protein length), or only in the genome of strain DM4.

[c] log2 fold-change of normalized read numbers in cultures grown with dichloromethane compared to methanol.

[d] False discovery rate, only adjusted p-value under 0.1 were considered.

[e] Adjacent to dichloromethane-enhanced gene encoding a protein of unknown function

[f] Using 2D gel proteomics of cell extracts of dichloromethane- versus methanol-grown *M. extorquens* DM4, proteins detected as more abundant on dichloromethane (METDI2655 and METDI2656) and less abundant on methanol (METDI0320) (Muller *et al.*, 2011)

3. Supplemental data

3.1. Defining the core genome

M. extorquens CM4 and DM4 are the only known *M. extorquens* strains that have been described in the literature to be able to utilize chlorinated compounds. Conversely, all *M. extorquens* strains are able to grow with methanol as sole carbon and energy source (Doronina *et al*., 1996; Vuilleumier *et al*., 2009). Thus, the two dechlorinated methanes may share specific gene sets involved in adaptation to growth with either chloromethane or dichloromethane but not with methanol. Comparative genomics of assembled reference genome sequences of five *M. extorquens* strains AM1, BJ001, CM4, DM4 and PA1 (Vuilleumier *et al*., 2009; Marx *et al*., 2012) aimed to identify the genes of the core genome. The core genome is defined as the set of genes (orthologous genes) shared by all the strains of the same bacterial species. A set of 3,489-shared CDS were defined for genes encoding for proteins with at least 80 % of amino acid identity on 80 % of the CDS length. The two dechlorinating strains, the chloromethane-degrading strain CM4 and the dichloromethane-degrading strain DM4, share a total of 4,620 genes as found using the MaGe plateform "Phyloprofile Exploration" tool (https://www.genoscope.cns.fr/agc/microscope/home/). Among those, 99 % are encoding for proteins with at least 90 % of amino acid identity on 80 % of the CDS length (data not shown). Within the MaGe plateform, "**Mchl**" label refer to genetic element of strain CM4 basing on its historical first taxonomical name: *Methylobacterium chl*oromethanicum (Doronina *et al*., 1996). Similarity, genes in the genome of strain DM4 are labelled as "**METDI**" for *Methylobacterium extorquens* dichloromethane-degrading strain. The correspondence of "Mchl"/"METDI" names (except for transposable elements) is listed in Table 3.1 (https://seafile.unistra.fr/f/a924b1b359/).

3.2. RNAseq data of the variable genome only shared by dechlorinating strains

The variable genome exclusively shared by the genomes of the dechlorinating strains included 163 genes (listed in Table 3.2) with 67 plasmid-borne genes with no predicted functions besides those related to plasmids (replication, integrase, transfert). They were located on two plasmids: pCMU01 in *M. extorquens* CM4 and p1METDI in *M. extorquens* DM4. The extent of relatedness between these plasmids is remarkable since 47 % of p1METDI-borne CDS have a homolog on plasmid pCMU01 in three synteny regions (39 genes; (Roselli, 2009)). A few

chromosomal genes had predicted function in central metabolism (oxidoreductase, ATP-hydrolyzing enzyme, succinyl CoA synthetase), and stress response (beta-lactamase, esterase) (Table 3.2). Using intraspecies comparative RNAseq analysis, no gene shared only by *M. extorquens* CM4 and DM4 were detected as differentially abundant in cultures grown with chlorinated methanes compared to methanol.

Table 3.1. RNAseq data of genes shared exclusively by *M. extorquens* CM4 and DM4

CM4 label	DM4 label	Gene	Product	*M. extorquens* CM4		*M. extorquens* DM4	
				log2fc[a]	Adjusted p-value[b]	log2fc	Adjusted p-value
Mchl_0111	METDI4323	-	protein of unknown function	0.7	0.140	-0.9	0.530
Mchl_0185	METDI0124	-	protein of unknown function	1.2	0.014	1.4	1.000
Mchl_0225	METDI4784	-	protein of unknown function	1.3	<0.001	1.6	0.130
Mchl_0318	METDI4325	-	protein of unknown function	0.4	0.450	-0.7	0.700
Mchl_0688	METDI0657	-	fragment of putative flagellin (FlaA-like; FlaB-like)	-1.0	<0.001	-0.9	0.830
Mchl_0906	METDI1089	-	putative OsmB-like lipoprotein	0.4	0.340	1.0	0.770
Mchl_0969	METDI1157	-	phage integrase	0.1	0.920	0.8	0.580
Mchl_0970	METDI1158	-	phage integrase	-1.0	<0.001	-0.2	0.990
Mchl_0971	METDI1159	-	putative phage-associated protein	-0.8	0.047	-1.0	0.550
Mchl_0974	METDI1162	-	putative ferrichrome-iron receptor protein	-0.1	0.940	-0.3	0.970
Mchl_0975	METDI1163	-	putative exported esterase (IroE-like)	0.4	0.670	-0.2	1.000
Mchl_0976	METDI1164	-	protein of unknown function	0.2	0.730	0.8	0.590
Mchl_0977	METDI1165	-	protein of unknown function	-0.2	0.740	-0.6	0.840
Mchl_0979	METDI1168	-	putative resolvase/recombinase	0.5	0.280	0.5	0.900
Mchl_0980	METDI1170	-	putative resolvase/recombinase	0.5	0.200	0.5	0.900
Mchl_0982	METDI1174	-	putative phosphohydrolase/phosphoesterase	0.5	0.470	-0.1	1.000
Mchl_0983	METDI1175	-	putative transcriptional repressor	0.5	0.430	-1.3	0.160
Mchl_0984	METDI1180	-	protein of unknown function	-0.1	0.990	0.4	0.850
Mchl_0985	METDI1181	-	putative ATP-hydrolyzing enzyme	0.5	0.380	-0.1	1.000
Mchl_0988	METDI1314	-	fragment of protein of unknown function	0.3	0.720	0.4	0.860
Mchl_0990	METDI1315	-	protein of unknown function	0.1	0.980	0.4	0.890
Mchl_1005	METDI1336	-	protein of unknown function	0.6	0.480	1.4	0.240
Mchl_1022	METDI1354	-	protein of unknown function	0.9	0.011	1.2	0.770
Mchl_1037	METDI1369	-	protein of unknown function	1.1	<0.001	0.7	0.910
Mchl_1080	METDI1216	-	protein of unknown function	0.6	0.590	-0.3	0.970
Mchl_1114	METDI1282	-	conserved exported protein of unknown function	0.4	0.620	0.9	0.570
Mchl_1136	METDI1461	*opdE*	transcription regulatory protein	0.1	0.960	-0.4	0.930
Mchl_1137	METDI1462	-	putative short-chain dehydrogenase/reductase SDR	0.1	1.000	-0.3	0.970
Mchl_1138	METDI1463	-	conserved exported protein of unknown function	0.1	0.970	-0.1	1.000
Mchl_1141	METDI1467	-	putative short-chain dehydrogenase/reductase SDR	-0.4	0.690	-0.7	0.850
Mchl_1142	METDI1468	-	putative transcriptional regulator, LysR family	-0.1	0.940	0.3	0.930
Mchl_1143	METDI1469	-	oxidoreductase, 2Fe-2S subunit	-0.2	0.810	0.7	0.610
Mchl_1144	METDI1470	-	oxidoreductase, FAD-binding subunit	0.1	0.960	0.8	0.500
Mchl_1145	METDI1471	-	oxidoreductase, molybdopterin-binding subunit	0.2	0.800	0.6	0.750
Mchl_1147	METDI1473	-	protein of unknown function	0.4	0.530	0.5	0.820
Mchl_1259	METDI1666	-	putative transcriptional regulator, LysR family	-0.1	0.900	-0.1	1.000
Mchl_1260	METDI1667	-	ABC transporter, periplasmic protein; putative taurine transporter	-0.6	0.280	-0.2	0.980
Mchl_1261	METDI1668	-	ABC transporter, permease component; putative taurine transporter	-0.5	0.520	-0.1	1.000
Mchl_1262	METDI1669	-	ABC transporter, ATPase; putative taurine transporter	-0.6	0.420	0.2	0.960
Mchl_1264	METDI1671	-	glutamyl-tRNA(Gln) amidotransferase subunit A	-0.0	1.000	-0.1	1.000
Mchl_1480	METDI1923	-	metal-dependent phosphohydrolase, HD subdomain	1.7	<1.000	0.1	1.000
Mchl_1481	METDI1924	-	transcriptional regulator, AraC family	-0.2	0.890	0.4	0.880
Mchl_1482	METDI1925	-	protein of unknown function	0.3	0.770	-0.9	0.630
Mchl_1484	METDI1927	-	protein of unknown function	0.8	0.048	0.3	0.9940
Mchl_1493	METDI1935	-	conserved protein of unknown function	0.1	0.830	0.3	0.930

Mchl_2249	METDI2736	-	protein of unknown function	-0.3	0.730	0.2	0.970
Mchl_2370	METDI2862	-	protein of unknown function	0.5	0.760	-0.9	0.670
Mchl_2863	METDI3394	-	protein of unknown function	0.5	0.280	-0.1	1.000
Mchl_3241	METDI3788	-	conserved protein of unknown function	0.4	0.390	0.1	0.990
Mchl_3333	METDI0326	-	putative sensory box-containing diguanylate cyclase/phosphodiesterase	0.4	0.390	0.7	0.870
Mchl_3438	METDI3901	-	conserved protein of unknown function	0.1	0.950	-0.3	0.960
Mchl_3466	METDI3928	-	putative addiction module antidote protein	-0.6	0.300	0.1	0.990
Mchl_3840	METDI4356	-	phage-related integrase (fragment)	0.9	0.930	0.3	0.930
Mchl_3913	METDI4612	-	conserved protein of unknown function	0.5	0.510	-0.4	0.900
Mchl_4141	METDI4814	-	protein of unknown function	0.9	0.096	-0.4	0.0001
Mchl_4205	METDI0774	-	protein of unknown function	-0.1	0.920	0.5	0.900
Mchl_4206	METDI0775	*sucD*	succinyl-CoA synthetase, alpha subunit	0.9	0.130	1.7	0.022
Mchl_4207	METDI0776	*sucC*	succinyl-CoA synthetase, beta subunit	0.6	0.320	1.5	0.056
Mchl_4208	METDI0777	-	putative succinate-semialdehyde dehydrogenase	0.6	0.290	1.4	0.120
Mchl_4209	METDI0778	-	major facilitator superfamily transporter	0.5	0380	1.9	0.008
Mchl_4211	METDI0781	-	glucose-methanol-choline oxidoreductase	-0.1	0.910	0.2	0.970
Mchl_4212	METDI0782	-	MFS transporter membrane protein	0.1	0.980	0.1	0.840
Mchl_4214	METDI0788	*xsc*	sulfoacetaldehyde acetyltransferase	-0.2	0.780	0.1	1.000
Mchl_4215	METDI0789	-	DNA-binding transcriptional repressor, IclR family	0.1	0.780	0.4	0.980
Mchl_4216	METDI0790	-	putative acetate kinase (AckA)	0.2	0.960	0.1	0.890
Mchl_4341	METDI4965	-	conserved protein of unknown function	0.7	0.092	0.1	1.000
Mchl_4342	METDI4966	-	conserved protein of unknown function	0.5	0.230	0.4	0.890
Mchl_4350	METDI4974	-	conserved protein of unknown function	0.6	0.200	0.7	0.670
Mchl_4444	METDI5071	-	protein of unknown function	-0.2	0.810	-0.5	0.840
Mchl_4550	METDI5180	-	conserved protein of unknown function	0.2	0.830	0.4	0.900
Mchl_4551	METDI5181	-	conserved protein of unknown function	0.1	0.930	0.5	0.900
Mchl_4659	METDI5293	-	conserved protein of unknown function	0.8	0.100	0.1	1.000
Mchl_4718	METDI5356	-	putative phage integrase	-0.1	0.840	0.4	0.880
Mchl_4719	METDI5357	-	putative transcriptional regulator	0.4	0.690	-0.4	0.960
Mchl_4720	METDI5358	-	putative metallo-beta-lactamase family protein	0.1	0.990	-0.5	0.910
Mchl_4721	METDI5359	-	conserved protein of unknown function	0.4	0.560	-0.4	0.890
Mchl_4722	METDI5360	-	protein of unknown function	0.6	0.260	0.6	0.800
Mchl_4723	METDI5361	-	protein of unknown function	0.4	0.670	0.6	0.780
Mchl_4729	METDI1222	-	protein of unknown function	0.5	0.270	1.4	0.120
Mchl_4742	METDI5367	-	protein of unknown function	1.4	<0.001	0.1	0.950
Mchl_4743	METDI5371	-	protein of unknown function	0.3	0.810	-0.4	0.930
Mchl_4744	METDI5372	-	putative DNA helicase related protein	0.1	0.860	1.4	0.590
Mchl_4749	METDI5375	-	putative ardC antirestriction protein	1.1	0.260	-1.3	0.690
Mchl_5330	METDI2680	-	putative SAM-dependent methyltransferase domain	0.5	0.290	0.1	1.000
Mchl_5331	METDI1912	-	protein of unknown function	0.6	0.280	1.1	0.370
Mchl_5447	METDI0315/ METDI0325	-	transposase	1.2	0.001	1.1/ 0.9	0.300/ 0.460
Mchl_5456	METDI4486	-	protein of unknown function	0,7	0.190	-0.3	0.960
Mchl_5457	METDI4485	*ispF*	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	-1.4	<0.001	-1.7	0.150
Mchl_5458	METDI4484	-	putative pseudouridine synthase, putative 1-deoxy-D-xylulose 5-phosphate reductoisomerase	-1.6	<0.001	-1.8	0.260
Mchl_5459	METDI4483	*gdhA*	glutamate dehydrogenase (NAD(P)+) oxidoreductase	-1.3	<0.001	-1.4	0.650
Mchl_5461	METDI4481	-	putative oxidoreductase	-1.8	<0.001	-1.1	0.780
Mchl_5546[c]	p1METDI0092[c]	-	conserved protein of unknown function	1.4/1 .4	0.056	0.2	0.920
Mchl_5549[c]	p1METDI0095[c]	-	protein of unknown function	1.4	0.085	0.2	1.000
Mchl_5552[c]	p1METDI0098[c]	-	conserved protein of unknown function	0.5	0.570	-0.1	1.000
Mchl_5555[c]	p1METDI0102[c]	-	conserved protein of unknown function	1.0	0.082	-0.3	0.940
Mchl_5558[c]	p1METDI0104[c]	*ardC*	antirestriction protein ArdC	0.7	0.240	0.5	0.750
Mchl_5559[c]	p1METDI0105[c]	-	conserved protein of unknown function	1.3	0.016	-0.6	0.091
Mchl_5564[c]	p1METDI0110[c]	-	conserved protein of unknown function	0.9	0.100	-0.4	0.970
Mchl_5567[c]	p1METDI0111[c]	-	putative methylase/helicase	0.9	0.041	-0.5	0.940
Mchl_5568[c]	p1METDI0112[c]	-	conserved protein of unknown function	1.0	0.045	1.1	0.220
Mchl_5569[c]	p1METDI0113[c]	-	conserved protein of unknown function	1.1	0.052	-0.3	0.960
Mchl_5571[c]	p1METDI0117[c]	-	protein of unknown function	1.8	0.450	-0.2	1.000
Mchl_5572[c]	p1METDI0118[c]	*traG*	conjugal transfer protein	-0.6	0.170	-0.1	1.000
Mchl_5573[c]	p1METDI0119[c]	*traD*	conjugal transfer protein	-0.8	0.048	-0.2	0.990

Mchl_5575[c]	p1METDI0121[c]	-	putative conjugal transfer protein (TraA)	0.5	0.370	-0.8	0.650
Mchl_5578[c]	p1METDI0122[c]	-	protein of unknown function	0.4	0.470	-0.1	1.000
Mchl_5579[c]	p1METDI0123[c]	-	protein of unknown function	0.6	0.170	-0.1	1.000
Mchl_5591[c]	p1METDI0125[c]	-	conserved membrane protein of unknown function	1.1	0.011	-0.4	1.000
Mchl_5592[c]	p1METDI0126[c]	*icmB*	putative IcmB/DotO-related protein	0.7	0.082	-0.8	0.094
Mchl_5595[c]	p1METDI0129[c]	*icmE*	putative IcmE-related protein	0.6	0.210	-0.4	0.980
Mchl_5596[c]	p1METDI0130[c]	*icmK*	putative IcmK-related protein	0.4	0.370	-1.1	0.620
Mchl_5597[c]	p1METDI0131[c]	*icmL*	conserved protein of unknown function; putative IcmL	0.0	0.950	-1.1	0.780
Mchl_5600[c]	p1METDI0134[c]	-	protein of unknown function	-0.0	1.000	-1.0	0.840
Mchl_5601[c]	p1METDI0135[c]	-	protein of unknown function	0.9	0.190	-0.3	0.990
Mchl_5602[c]	p1METDI0136[c]	-	putative DotC-related protein	0.4	0.490	-0.3	1.000
Mchl_5603[c]	p1METDI0137[c]	-	putative Dot/Icm secretion system ATPase DotB	0.3	0.640	-0.2	1.000
Mchl_5604[c]	p1METDI0138[c]	-	conserved exported protein of unknown function	0.8	0.087	-0.2	0.990
Mchl_5605[c]	p1METDI0139[c]	-	protein of unknown function	0.8	0.034	-0.9	0.610
Mchl_5606[c]	p1METDI0140[c]	-	conserved protein of unknown function	0.9	0.009	-0.6	0.930
Mchl_5607[c]	p1METDI0141[c]	-	protein of unknown function	1.1	0.021	-0.7	0.880
Mchl_5609[c]	p1METDI0002[c]	-	conserved protein of unknown function	1.0	0.004	-0.2	1.000
Mchl_5610[c]	p1METDI0003[c]	-	conserved exported protein of unknown function	1.2	<0.001	-0.3	0.960
Mchl_5611[c]	p1METDI0004[c]	-	protein of unknown function	1.2	<0.001	-0.1	1.000
Mchl_5612[c]	p1METDI0005[c]	-	protein of unknown function	1.3	<0.001	-0.4	0.960
Mchl_5613[c]	p1METDI0006[c]	-	protein of unknown function	1.6	<0.001	-0.2	1.000
Mchl_5614[c]	p1METDI0007[c]	-	fragment of transposase related to IS*481* family	1.6	<0.001	0.3	0.940
Mchl_5615[c]	p1METDI0008[c]	*repA*	plasmid partitioning protein RepA	0.4	0.380	0.4	0.120
Mchl_5616[c]	p1METDI0009[c]	*repB*	plasmid partitioning protein RepB	0.5	0.200	0.5	0.960
Mchl_5617[c]	p1METDI0010[c]	*repC*	replication protein C	0.4	0.340	0.4	1.000
Mchl_5620[c]	p1METDI0111[c]	-	conserved protein of unknown function (fragment)	0.3	0.750	0.3	0.960
Mchl_5622	p1METDI0112	-	conserved protein of unknown function	1.2/1.3	1.000	1.0	0.710
Mchl_5644	p1METDI0053	-	conserved protein of unknown function, PilT domain	-0.3	0.620	-0.3	0.770
Mchl_5646	p1METDI0052	-	putative phage integrase	0.9	0.020	0.4	0.870
MCv2_0417	METDI0198	-	protein of unknown function	0.0	0.990	0.3	0.930
MCv2_0842	METDI0940	-	protein of unknown function	-0.5	0.380	-0.1	1.000
MCv2_1071	METDI1130	-	protein of unknown function	0.5	0.710	0.7	0.750
MCv2_1101	METDI1161	-	protein of unknown function	-0.1	0.990	-0.1	0.980
MCv2_1114	METDI1171	-	conserved protein of unknown function	0.4	0.500	0.1	1.000
MCv2_1115	METDI1172	-	protein of unknown function	1.1	0.024	0.3	0.930
MCv2_1120	METDI1179	-	protein of unknown function	0.5	0.620	0.3	0.940
MCv2_1140	METDI1329	-	protein of unknown function	0.1	0.940	-0.6	0.750
MCv2_1149	METDI1339	-	protein of unknown function	-0.2	0.870	0.5	0.800
MCv2_1217	METDI1202	-	protein of unknown function	0.3	0.780	1.1	0.430
MCv2_1218	METDI1203	-	protein of unknown function	0.2	0.800	0.4	0.900
MCv2_1316	METDI1474	-	protein of unknown function	-0.3	0.790	-0.2	0.970
MCv2_1584	METDI1797	-	conserved protein of unknown function	1.5	0.003	-0.7	0.740
MCv2_1680	METDI1888	-	protein of unknown function	0.8	0.350	0.2	0.950
MCv2_1703	METDI1921	-	putative phosphoglycerate/bisphosphoglycerate mutase	0.2	0.750	-0.2	0.970
MCv2_1739	METDI1950	-	protein of unknown function	1.1	0.018	-0.3	0.960
MCv2_1740	METDI1951	-	protein of unknown function	0.8	0.490	-0.1	1.000
MCv2_1741	METDI1952	-	conserved protein of unknown function	-0.1	1.000	-0.5	0.900
MCv2_1751	METDI1961	-	conserved exported protein of unknown function	-0.3	0.740	-0.7	0.640
MCv2_2296	METDI2356	-	conserved protein of unknown function	0.7	0.320	0.5	0.840
MCv2_2310	METDI2369	-	protein of unknown function	0.3	0.760	-0.6	0.840
MCv2_2529	METDI2709	-	protein of unknown function	0.3	0.660	-0.5	0.810
MCv2_2712	METDI2882	-	protein of unknown function	1.3	0.060	0.8	0.580
MCv2_2972	METDI3136	-	protein of unknown function	0.5	0.460	0.5	0.820
MCv2_3197	METDI3358	-	protein of unknown function	0.3	0.720	0.2	0.950
MCv2_3231	METDI3407	-	conserved protein of unknown function	0.3	0.670	0.4	0.850
MCv2_3537	METDI3709	-	protein of unknown function	0.3	0.600	-0.8	0.580
MCv2_4420	METDI5044	-	protein of unknown function	0.1	0.980	0.8	0.590
MCv2_4813	METDI4912	-	protein of unknown function	-0.5	0.590	-0.3	0.940
MCv2_5307	METDI5397	-	protein of unknown function	0.2	0.900	-1.7	0.070
MCv2_5403	METDI4380	-	conserved protein of unknown function	1.0	0.046	0.1	1.000

[a] log2 fold-change (log2fc) values ≥ 2 or ≤ -2 mean that transcript abundance is higher in cultures grown with chlorinated methanes compared to methanol, or higher with methanol compared to chlorinated methanes, respectively.
[b] False discovery rate, only adjusted p-value under 0.1 were considered.
[c] In synteny region between plasmid pCMU01 and plasmid p1METDI.

3.3. RNAseq data of plasmid pCMU01-borne gene with chromosomal homologs

Among plasmid-borne homologs, 28 out of 98 had transcript found as more abundant in cultures grown with chloromethane compared to methanol (Table 3.3; no homologs found as differentially less abundant). Among those, 24 were involved in biosynthesis of the essential chloromethane dehalogenase corrinoid compound (Studer *et al*., 2001, 2002) (see section 3.2 article). Similarly, of the homologs encoding cobalt and cobalamin transporters, only copies located on plasmid pCMU01 has transcript with increased abundant in cultures grown with chloromethane compared to methanol, unlike core genome homologs (*czcA*, encoding a RND transporter or some transport proteins encoded by *tonB*, *exbB*/*exbD*, *tolQ*/*tolR* and *btuFCD*). This suggests that a yet uncharacterized chloromethane-dependent regulation mechanism favors the expression of genes of the variable genome over homologous genes of the core genome.

Table 3.2. RNAseq data for pCMU01-borne genes with chromosomal homologs

Homolog name & localization[a]		Gene	Product	aa Id (%)	Plasmid pCMU01		Chromosome	
pCMU01	Chromosome				log2fc[b]	Adjusted p-value[c]	log2fc	Adjusted p-value
Mchl_5400	Mchl_2934	-	putative nitrile hydratase activator protein	77.3	1.3	0.050	-0.6	0.700
Mchl_5405	Mchl_3481	-	putative H(+)/Cl(-) exchange transporter ClcA	68.0	5.5	<0.001	0.3	0.700
Mchl_5435	Mchl_1555	-	putative universal stress protein, UspA-like (fragment)	58.7	1.0	0.050	0.5	0.300
Mchl_5437	Mchl_4097	-	integrase family protein	58.4	-0.1	0.940	-0.1	0.960
Mchl_5444	Mchl_2145	*xoxF2*	putative PQQ-linked dehydrogenase of unknown specificity	88.2	0.9	0.019	0.5	0.240
Mchl_5445	Mchl_3593	-	sensor histidine kinase	40.6	0.9	0.047	-0.1	0.880
Mchl_5446	Mchl_3594	-	putative response regulator; CheY and HTH LuxR family	64.4	1.0	0.074	-0.2	0.890
Mchl_5455	Mchl_2361	-	putative cystathionine beta-synthase	82.1	1.0	0.006	-0.3	0.640
Mchl_5457	Mchl_3044	*ispF*	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	94.5	-1.4	<0.001	0.1	0.990
Mchl_5460	Mchl_0670	*serC2*	phosphoserine aminotransferase	95.2	-1.2	<0.001	-0.6	0.150
Mchl_5461	Mchl_2219	-	putative oxidoreductase	52.7	-1.8	0.001	-0.2	0.740
Mchl_5462	Mchl_2317	*hss2*	homospermidine synthase	96.4	-1.4	0.001	-0.6	0.210
Mchl_5475	Mchl_0044	-	putative sensor protein (fragment)	45.7	0.8	0.091	1.4	<0.001
Mchl_5476	Mchl_0272	-	putative sensor protein (fragment)	61.3	1.2	0.009	0.6	0.130
	Mchl_0517			46.2			1.0	0.034
	Mchl_1530			56.2			-0.1	1.000
Mchl_5478	Mchl_2189	-	conserved protein of unknown function	69.0	1.0	0.015	-0.7	0.110
Mchl_5491	Mchl_2220	*shc2*	squalene-hopene cyclase	56.7	-0.1	0.970	-0.5	0.430
Mchl_5495	Mchl_3651	-	RND efflux transporter, MFP subunit	47.8	0.9	0.270	-0.5	0.250
Mchl_5496	Mchl_1298	-	RND efflux transporter, HME family, translocase subunit	40.8	0.3	0.540	0.5	0.320
	Mchl_3650			60.2			-0.7	0.140

	Mchl_4139			40.3			0.4	0.630
	Mchl_3285			40.1			0.1	0.970
Mchl_5499	Mchl_1419	-	response regulator in two-component regulatory system	42.2	2.4	<0.001	0.3	0.700
	Mchl_2980			40.2			0.2	0.750
Mchl_5502	Mchl_1959	-	conserved protein of unknown function	72.2	1.3	0.120	-0.3	0.470
Mchl_5504	Mchl_0559	-	putative di-heme cytochrome c peroxidase (MauG-related)	41.3	1.0	0.021	0.3	0.650
	Mchl_2999			40.3			-0.1	1.000
Mchl_5509	Mchl_4119	-	conserved exported protein of unknown function	45.4	0.9	0.041	0.4	0.550
Mchl_5512	Mchl_5137	-	putative molybdopterin molybdenumtransferase (MoeA)	42.3	0.7	0.350	0.2	0.790
Mchl_5514	Mchl_4941	*modA2*	molybdate ABC transporter, periplasmic binding component	61.2	0.6	0.140	0.1	0.950
Mchl_5515	Mchl_3926	*modB2*	molybdate ABC transporter, membrane component	75.4	0.4	0.490	-0.1	0.720
Mchl_5516	Mchl_3928	*modC2*	molybdate ABC transporter, ATP-binding component	55.7	0.6	0.210	0.1	0.520
Mchl_5517	Mchl_2845	*mop2*	molybdenum-pterin-binding protein, putative molybdate homeostasis factor	69.6	1.0	0.003	0.1	0.890
Mchl_5520	Mchl_2371	-	conserved exported protein of unknown function	44.8	1.8	<0.001	0.1	1.000
	Mchl_2536			46.4			-1.9	<0.001
	Mchl_1417			40.7			-1.0	0.015
	Mchl_1055			43.5			0.4	0.740
Mchl_5526	Mchl_1112	-	transposase of IS*Mex11*	100	2.3	<0.001	0.7	0.410
	Mchl_1119			64.3			0.1	0.910
Mchl_5528	Mchl_2715	-	putative diguanylate cyclase (fragment)	42.1	1.0	0.240	-0.9	0.920
	Mchl_2736			45.5			0.1	0.920
	Mchl_2255			46.3			0.4	0.490
Mchl_5529	Mchl_0114	-	conserved protein of unknown function	58.5	1.4	0.003	0.1	0.980
	Mchl_1956			75.7			-0.2	0.850
Mchl_5531	Mchl_0082	-	putative transcriptional regulator, Ros/MucR family	46.4	1.0	0.052	0.1	0.900
	Mchl_4811			49.4			-0.2	0.730
	Mchl_4927			53.4			0.3	0.620
Mchl_5542	Mchl_2091	-	putative aminoglycoside phosphotransferase	43.4	0.6	<0.001	-0.1	<0.001
Mchl_5544	Mchl_2409	-	putative glycosyl transferase	61.6	1.4	0.038	0.3	0.730
Mchl_5546	Mchl_0114	-	conserved protein of unknown function	59.3	1.3	0.056	0.1	0.980
	Mchl_1956			75.7			-0.2	0.850
Mchl_5556	Mchl_5373	-	putative methyl-accepting chemotaxis protein	46.0	0.6	0.120	1.6	<0.001
	Mchl_0134			46.2			0.9	0.008
Mchl_5558	Mchl_4749	*ardC*	antirestriction protein ArdC	40.8	0.7	0.240	1.0	0.260
Mchl_5566	Mchl_0082	-	putative transcriptional regulator	40.8	1.0	0.001	0.1	0.900
	Mchl_4938			38.7			0.5	0.240
Mchl_5581	MCv2_4479	-	transposase (fragment)	90.9	2.3	0.011	-0.5	0.540
	Mchl_1117			77			1.6	0.038
Mchl_5582	Mchl_0531	-	class III alcohol dehydrogenase, S-(hydroxymethyl) glutathione dehydrogenase	58.4	1.8	<0.001	1.5	<0.001
	Mchl_2582			61.5			0.6	0.280
	Mchl_4105			59.2			0.5	0.380
Mchl_5618	Mchl_1956	-	protein of unknown function	69.4	2.2	<0.001	0.1	0.850
Mchl_5623	Mchl_1137	-	dehydrogenase/reductase (fragment)	41.6	0.6	0.590	0.1	1.000
	Mchl_5004			46.7			0.1	0.890
Mchl_5627	Mchl_4232	-	conserved protein of unknown function	67.3	0.9	0.190	0.7	0.400
Mchl_5639	Mchl_3405	-	conserved exported protein of unknown function	79.0	1.7	<0.001	1.7	<0.001
Mchl_5646	Mchl_4007	-	putative integrase	68.8	0.9	0.023	0.6	0.470
Mchl_5667	Mchl_3540	-	putative phosphohydrolase/phosphoesterase	52.9	1.3	<0.001	0.1	1.000
	Mchl_0982			42.0			0.5	0.470
	Mchl_1897			43.8			0.3	0.280
Mchl_5669	Mchl_0978	-	putative ATP-hydrolyzing enzyme	42.1	1.0	0.032	0.3	1.000
Mchl_5677	Mchl_3153	*btuF*	ABC transporter periplasmic binding component, putative vitamin B12 transporter (BtuF)	61.1	6.0	<0.001	0.1	0.990
Mchl_5678	Mchl_3152	*btuC*	ABC transporter, membrane component, putative vitamin B12 transporter subunit (BtuC)	73.5	5.3	<0.001	-0.5	0.590
Mchl_5679	Mchl_4153	*btuD*	ABC transporter ATP-binding component, putative vitamin B12 transport (BtuD)	40.0	5.0	<0.001	-0.7	0.870
	Mchl_3151			65.9			-0.1	0.970
Mchl_5680	Mchl_3150	-	protein of unknown function	66.3	4.8	<0.001	0.1	1.000
Mchl_5682	Mchl_1004	-	putative TonB-dependent receptor	78.3	4.4	<0.001	0.1	0.960
Mchl_5683	Mchl_2974	*gck2*	glycerate kinase	85.0	5.1	<0.001	-0.5	0.250

Mchl_5685	Mchl_1166	*cobM2*	precorrin-4 C(11)-methyltransferase	82.7	7.0	<0.001	-0.2	0.720
Mchl_5686	Mchl_1162	*cobE2*	putative cobalamin biosynthesis protein CobE	64.5	4.8	<0.001	0.9	0.490
Mchl_5687	Mchl_1161	*cobL2*	precorrin-6Y C(5,15)-methyltransferase (decarboxylating)	76.1	4.7	<0.001	-0.2	0.990
Mchl_5688	Mchl_5189	*cobK2*	precorrin-6A reductase	63.4	0.7	0.600	0.2	0.810
Mchl_5689	Mchl_1710	*cobJ2*	precorrin-3B C(17)-methyltransferase	79.1	5.5	<0.001	0.4	0.620
Mchl_5690	Mchl_1711	*cobI2*	precorrin-2 C(20)-methyltransferase	80.7	5.7	0.015	0.5	0.320
Mchl_5691	Mchl_1712	*cobH2*	precorrin-8X methylmutase	89.2	5.9	<0.001	0.5	0.370
Mchl_5693	Mchl_2536 Mchl_4258	-	conserved exported protein of unknown function, CoxB-related protein	40.5 49.4	8.9	<0.001	-1.9 -0.1	<0.001 0.960
Mchl_5701	Mchl_4908	*folC2*	putative folylpolyglutamate synthase and dihydrofolate synthase	46.3	7.8	<0.001	-0.2	0.720
Mchl_5702	Mchl_0872	*cobU2*	nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase	54.5	6.8	<0.001	-0.1	0.890
Mchl_5703	Mchl_3302	-	putative transposase (fragment)	85.7	2.4	<0.001	-0.1	0.990
Mchl_5704	Mchl_4231	-	putative transposon (fragment)	95.4	3.2	<0.001	-0.8	0.706
Mchl_5713	Mchl_0376	*czcR*	DNA-binding response regulator in two-component regulatory system, putative activator of the heavy-metal resistance system Czc	49.8	1.9	<0.001	-0.7	0.150
Mchl_5715	Mchl_1072	*czcA2*	RND divalent metal cation efflux transporter membrane component	43.4	2.3	<0.001	0.2	0.670
Mchl_5721	Mchl_1722	*cobP2*	bifunctional adenosylcobalamin biosynthesis enzyme (adenosylcobinamide kinase/adenosylcobinamide-phosphate guanylyltransferase)	73.6	6.7	<0.001	-0.3	0.630
Mchl_5722	Mchl_1723	*cobO2*	cob(I)yrinic acid a,c-diamide adenosyltransferase	83.5	6.5	<0.001	0.1	0.870
Mchl_5723	Mchl_5190	*cobQ2*	cobyric acid synthase	75.2	6.2	<0.001	-0.4	0.460
Mchl_5724	Mchl_5191	*cobD2*	cobalamin biosynthesis protein	76.4	2.3	<0.001	0.3	0.710
Mchl_5730	Mchl_5192	*cobC2*	L-threonine-O-3-phosphate decarboxylase domain	51.2	5.0	<0.001	-0.1	0.960
Mchl_5735	Mchl_0172 Mchl_4926 Mchl_1650	-	RNA polymerase ECF-type sigma factor	54.7 43.5 41.4	1.3	<0.001	0.9 -0.3 -1.3	0.074 0.580 <0.001
p1MCv2_0095	Mchl_0651	-	conserved protein of unknown function	44.3	2.3	<0.001	2.3	<0.001
p1MCv2_0107	Mchl_2145	-	XoxF, PQQ-linked dehydrogenase (fragment)	100	1.8	<0.001	0.5	0.240

[a] Genes encoding for transposon-related elements with no differential transcript abundance in the studied conditions are not listed. Genes are homologs when they have 80 % of identity on at least 80 % of the protein length.

[b] log2 fold-change (log2fc) values ≥ 2 or ≤ -2 mean that transcript abundance is higher in cultures grown with chlorinated methanes compared to methanol, or higher with methanol compared to chlorinated methanes, respectively.

[c] False discovery rate, only adjusted p-value under 0.1 were considered.

3.4. RNAseq data of genes with differential sense/antisense transcript abundance

Directional RNAseq was performed so that the information about the DNA strand from which the RNA was transcribed can be traced. Knowing the direction of the transcription is crucial for resolving overlapping genetic features, detecting antisense transcription and assigning the sense strand for non-coding RNA (ncRNA). Antisense RNAs referred to as *cis*-encoded RNAs are transcribed opposite annotated genes and thus share extensive complementarity with the corresponding transcripts although the antisense RNA may have additional targets in *trans* (Brantl, 2007). In the submitted paper, mRNA quantification was studied with the sum of sense and antisense reads in order to detect sets of genes with differential transcript abundance).

Among those, we searched for the occurrence of putative antisense transcription nested within predicted mRNA-encoding genes. We defined the "sense ratio" which was calculated by dividing the number of normalized sense reads by the total number of normalized total sense and antisense reads. A "sense ratio" < 0.5 may point out misannotated uncharacterized CDS nested within predicted CDS, or may result from transcription noise (Llorens-Rico *et al*., 2016), read-through of adjacent genes, or involve regulation mechanisms (Georg and Hess, 2011; Chang *et al*., 2015; Brophy and Voigt, 2016).

Three groups were defined on the basis of the "sense ratio" in cultures grown with chlorinated methanes compared to methanol (Table 3.4).

- Group A, the "sense ratio" is lower than 0.5 pointing out the prevalence of antisense transcription;
- Group B includes CDS for which the "sense ratio" is above 0.8 only in one growth condition, and at least half the value in the other tested growth condition. This suggests conditional antisense transcription in cultures grown with different one-carbon compounds;
- Group C includes CDS that do not fall in groups A and B. This may include genes with similar ratio of sense and antisense transcription.

Using these criteria, in *M. extorquens* CM4, 20 genes were detected of which 17 were plasmid pCMU01-borne including six genes of a uncharacterized highly transcribed gene cluster harbouring *cszA2* and *cobS*-like, seven mobile element-related pseudogenes, and gene *cobM2*. The CobM2 protein (69 % aa identity with the characterized *Pseudomonas denitrificans* precorrin-4 C(11)-methyltransferase (Debussche *et al*., 1993) was found directly upstream of a mobile-element gene (Mchl_5684), which may be responsible, at least in part, of *cobM2* antisense transcription. In *M. extorquens* DM4, of the 4 highly transcribed genes in methanol that had a majority of antisense reads, two were members of the gene family of uncharacterized ANAH-like domain containing proteins localized on GEI 197 (METDI1426; METDI4555).

Table 3.3. RNAseq data of genes with differential sense/antisense transcript abundance

Sense/ antisense group	Gene characteristics[a]			Directional RNAseq data[b]					
	Gene	Product	Genomic occurrence	log2fc	Adjusted p-value	Relative read count[c]		Sense ratio[d]	
						CM or DCM	CH_3OH	CM or DCM	CH_3OH
***M. extorquens* CM4**									
Group A: Antisense transcription									
Mchl_5465	-	transposase (fragment)	pCMU01	2.3	<0.001	1257	261	0.3	0.2
Mchl_5685	*cobM2*	precorrin-4 C(11)-methyltransferase	pCMU01	7.0	<0.001	432	3	0.4	0.1
Mchl_5618	-	conserved protein (fragment)	pCMU01	2.2	<0.001	293	67	0.3	0.1
Mchl_5684	-	putative transposase (fragment)	pCMU01	4.2	<0.001	2949	162	0.2	0.2
Mchl_5703	-	putative transposase (fragment)	pCMU01	2.4	0.015	208	43	0.0	0.1
Group B: Carbon source-dependent transcription orientation									
Mchl_1416	-	putative membrane protein	core	2.2	<0.001	226	50	0.9	0.5
Mchl_3283	*dctA*	C_4-dicarboxylate transport protein	core	-2.9	<0.001	149	1120	0.4	0.8
Mchl_5405	*clcA2*	putative $H(^+)/Cl(^-)$ exchange transporter	pCMU01	5.5	<0.001	3782	84	0.9	0.3
Mchl_5523	*acxA*	acetone carboxylase beta subunit	pCMU01	3.7	<0.001	1770	133	0.8	0.3
Mchl_5716[e]	-	conserved protein, CobS-like domain	pCMU01	7.1	<0.001	539	125	0.5	1.0
Mchl_5718[e]	-	putative membrane transport protein	pCMU01	7.1	<0.001	79188	578	1.0	0.5
Mchl_5719[e]	-	putative transport membrane protein	pCMU01	7.7	<0.001	81772	385	1.0	0.3
Group C: Similar ratio of sense and antisense transcription									
Mchl_5715[e]	*czcA2*	RND divalent metal cation efflux transporter SU	pCMU01	2.3	<0.001	204	42	0.5	0.3
Mchl_5663	-	conserved protein	pCMU01	2.1	<0.001	2072	484	0.5	0.3
p1MCv2_0379	-	protein of unknown function	pCMU01	2.4	<0.001	1514	296	0.5	0.7
p1MCv2_0380	-	protein of unknown function	pCMU01	2.1	<0.001	1233	295	0.2	0.6
***M. extorquens* DM4**									
Group A: Antisense transcription									
METDI0260	-	glucokinase (fragment)	GEI160	-4.3	<0.001	14	292	0.1	0.1
METDI0364	-	putative cytochrome c oxidase subunit (CbaB)	GEI 160	-3.4	<0.001	174	1791	0.4	0.4
METDI1414	*ardC*	antirestriction protein (ArdC)	GEI 270	-2.2	0.026	36	159	0.2	0.3
METDI1426	-	putative glycosyl transferase	GEI 270	-2.3	0.190	587	2867	0.2	0.2
METDI3864	-	protein of unknown function	specific	2.4	0.003	446	84	0.2	0.0
METDI4261	-	conserved protein of unknown function	core	-2.6	0.001	248	1525	0.0	0.0
METDI4557	-	protein of unknown function	GEI 197	-3.4	<0.001	1339	5840	0.4	0.4
Group B: Carbon source-dependent transcription orientation									
METDI0133	-	putative glycosyl transferase	variable	2.6	0.005	34	6	1.0	0.4
METDI0620	-	conserved protein of unknown function	core	3.8	<0.001	9568	681	1.0	0.4
METDI1755	-	conserved protein of unknown function	specific	2.1	0.004	446	104	0.9	0.2
METDI1762	-	conserved protein of unknown function	variable	-2.1	0.007	553	2370	0.5	0.9
METDI1842	-	conserved protein of unknown function	core	-2.1	0.005	82	343	0.2	0.7
METDI1978	-	conserved protein of unknown function	core	-2.1	0.036	31	130	0.4	0.8
METDI2432	-	putative phosphotransferase	core	2.7	<0.001	846	129	0.8	0.3
METDI2873	*fdh4B*	formate dehydrogenase subunit B	core	-3.2	<0.001	112	1020	0.5	1.0
METDI3359	-	protein of unknown function	core	2.1	0.002	610	143	0.8	0.2
METDI4507	-	putative cytochrome c	GEI 197	-3.3	<0.001	93	913	0.9	0.5
METDI5891	-	putative endonuclease	core	3.2	<0.001	3714	408	0.9	0.4
Group C: Similar ratio of sense and antisense transcription									
METDI0344	-	conserved protein of unknown function	GEI 160	2.2	0.013	139	31	0.5	0.2
METDI0365	-	conserved protein of unknown function	GEI 160	-3.3	<0.001	79	802	0.4	0.6
METDI0366	-	putative cytochrome c	GEI 160	-2.9	0.001	198	1438	0.6	0.4
METDI0367	-	putative cytochrome c	GEI 160	-2.7	<0.001	148	935	0.8	0.5
METDI0384	-	protein of unknown function	GEI 160	-2.5	0.004	174	977	0.7	0.5
METDI1598	-	protein of unknown function	variable	2.1	0.012	541	126	0.5	0.3
METDI2012	-	putative exported *cyd* operon protein	variable	-3.7	<0.001	424	5591	0.5	0.4
METDI4305	-	conserved protein of unknown function	variable	-2.4	0.004	1089	5589	0.6	0.4
METDI4504	-	putative cytochrome c oxidase SU	GEI 197	-3.0	<0.001	128	1055	0.5	0.4
METDI4506	-	putative cytochrome c	GEI 197	-2.5	0.006	189	1053	0.6	0.5
METDI4528	-	putative ion transport domain	GEI 197	-3.1	<0.001	308	2549	0.5	0.7
METDI4555	-	conserved protein with ANAH-like domain	GEI 197	-3.3	0.001	413	4165	0.5	0.4
METDI4564	-	putative cytochrome c	GEI 197	-2.6	0.003	661	3927	0.5	0.6
METDI4620	-	protein of unknown function	variable	2.2	0.002	476	106	0.7	0.4
METDI4695	*proV*	glycine/betaine/proline ABC transporter SU	variable	2.3	0.001	269	55	0.7	0.4

METDI4696	*opuAB*	glycine/betaine/proline ABC transporter SU	variable	2.0	1.000	15578	17755	0.5	0.3
METDI5532	*nrtB*	nitrate transport permease protein	core	-2.4	0.001	73	386	0.2	0.7

[a] MaGe annotation (https://www.genoscope.cns.fr/agc/microscope). Genes labelled as "core" encode for proteins with 80 % aa identity on at least 80 % of the protein length that are shared between *M. extorquens* genomes. Genes labelled as "variable" encode for homologs shared at least between 2 genomes of strains AM1, BJ001, CM4, DM4 and PA1. GEI stands for genomic islands as detailed in Table 4 of the submitted article.

[b] log2 fold-change (log2fc) values ≥ 2 or ≤ -2 mean that transcript abundance is higher in cultures grown with chlorinated methanes compared to methanol, or higher with methanol compared to chlorinated methanes, respectively.

[c] Averaged normalized read numbers divided by gene length in kb in different methylotrophic growth conditions.

[d] Number of normalized sense reads divided by total number of normalized sense and antisense reads.

[e] Adjacent to the Mchl_5714 to Mchl_5716 gene cluster, gene encoding Mchl_5717 (log2fc of 8.1) displayed a majority of sense transcript reads.

3.5. RNAseq data of formate dehydrogenase subunits encoding-genes

Formate is central to methylotrophic growth (see Chapter 1, Figure 1.10): For carbon assimilation, *M. extorquens* reduced formate into methylene-H_4F (CH_2=H_4F) (Marx *et al*., 2005) while for energy production, formate is oxidized to CO_2 by formate dehydrogenase (FDH). *M. extorquens* core genome harbour four characterized FDH (Chistoserdova *et al*., 2004, 2007). Some FDH-encoding SU genes were detected to have one-carbon compound utilization-dependent transcript abundance variations (summarized in Table 3.4).

Table 3.4. RNAseq data of formate dehydrogenase subunits encoding-genes

Formate dehydrogenase (cellular localization) [a]	Gene, encoded SU	*M. extorquens* CM4				*M. extorquens* DM4				Strain AM1[e]
		log2fc[b]	Adjusted p-value[c]	Relative read counts[d] CM	Relative read counts[d] CH_3OH	log2fc	Adjusted p-value	Relative read counts DCM	Relative read counts CH_3OH	
Fdh1	*fdh*1A, alpha	1.6	<0.001	9327	3164	2.3	0.150	17071	3401	methanol
tungsten-containing (cytoplasm)	*fdh*1B, beta	1.5	<0.001	7155	2480	1.8	0.350	10588	2981	methanol
Fdh2	*fdh*2A, alpha[f]	-0.7	0.085	2391	3891	0.9	0.800	8798	4767	methanol
NAD-dependent molybdenum containing (cytoplasm)	*fdh*2B, beta	-0.7	0.059	1571	2260	0.5	0.890	4856	3349	methanol
	*fdh*2C, gamma	-0.5	0.290	662	958	-0.4	0.960	970	1257	methanol
	*fdh*2D, delta	-0.9	0.029	1505	2821	0.9	0.450	3409	1801	methanol
Fdh3 (periplasm)	*fdh*3A, alpha	0.5	0.300	15087	10499	0.9	0.810	14985	8365	succinate
	*fdh*3B, beta	0.7	0.069	15221	9360	0.8	0.830	13849	7864	succinate
	*fdh*3C, gamma	0.5	0.270	10983	7894	0.8	0.840	1127	6360	succinate
Fdh4	*fdh*4A alpha	-0.7	0.048	1819	3041	-4.0	<0.001	140	2280	nd
(cytoplasm)	*fdh*4B, beta	-1.2	0.001	762	1779	-3.2	<0.001	112	1020	nd

[a] Data from Chistoserdova *et al*., 2004, 2007; Peyraud *et al*., 2011.
[b] log2 fold-change (log2fc) values ≥ 2 or ≤ -2 mean that transcript abundance is higher in cultures grown with chlorinated methanes compared to methanol, or higher with methanol compared to chlorinated methanes, respectively.
[c] False discovery rate, only an adjusted p-value under 0.1 was considered as significant.
[d] Averaged normalized read numbers divided by gene length in kb in different methylotrophic growth conditions.
[e] Microarray data from Okubo *et al*., 2007. Transcript abundance higher in cultures grown with methanol, succinate or unchanged (nd).
[f] Proteomic data from Muller *et al*., 2011. Protein more abundant in cultures grown with dichloromethane compared to methanol.

4. Conclusion

The transcriptomes for the utilization of methanol, chloromethane or dichloromethane as sole carbon and energy source were compared in two *M. extorquens* strains. The core genome represents 75 % (4,620 CDS) of their shared genes. The variable genome includes strain-specific pathways of chloromethane and dichloromethane that are initiated by the expression of dehalogenation modules that respectively are encoded by the *cmu* genes located on the 380 kb pCMU01 plasmid in strain CM4 and the *dcm* genes nested within a genomic island (GEI) in strain DM4. The core genome had a lower proportion of differentially expressed genes than the variable genome. A few common genes had higher transcript abundance under the following growth conditions:

(i) chlorinated methane, including stress response and related bioenergetics genes (*depP*, *hppA* encoding a membrane-bound proton translocating pyrophosphatase);

(ii) chloromethane, including the membrane-bound transhydrogenase *pnt* gene cluster for $NADP^+$ regeneration and the glycine cleavage complex *gcvPHT* gene cluster involved in C_1 carrier tetrahydrofolate (H_4F) metabolism;

(iii) dichloromethane, including a putative glutathione peroxidase;

(iv) methanol, including 8 genes of the PQQ-dependent C_1/C_2 dehydrogenase *mdh2*-associated gene module.

The variable genome of strain CM4 harbored the majority of the genes whose transcription was enhanced by chloromethane: carbon assimilation genes (*cmu, acxABC*, *gck2*), coenzyme-associated *cmu*-pathway genes for vitamin B_{12} (14 *cob* genes, *btuBCDF*, *bluB2*) and for H_4F (*folC2*, *folD*, *metF2*, *purU*). In the variable genome of strain DM4, the *dcm* genes were more expressed genes in cultures grown with dichloromethane compared to methanol, whereas central metabolism genes (*ackA*, *adh*-like, *xfp*, cytochrome-associated genes) encoded by redundant GEIs had increased transcript abundance in cultures grown with methanol. The common transcriptional response for *M. extorquens* grown with chloromethane or dichloromethane was limited to a few core genome genes. This suggests that transcription responses to growth with chloromethane and dichloromethane are predominantly uncorrelated. Redox status adjustments are probably necessary for growth with different C_1 compounds as deduced from the transcription profiles of genes of the core and the variable genome.

Chapter 4

Ecotypes of microbial chloromethane utilizers in a forest soil

Chapter 4. Ecotypes of microbial chloromethane utilizers in a forest soil

1. Introduction

Chloromethane (CH_3Cl) is an ozone-destroying volatile organic compound (VOC), which is mainly produced by natural processes (Galbally and Kirstine 2002; Denman *et al*., 2007; see Chapter 1). Its global emissions to the atmosphere are about 10 Gmol $year^{-1}$. Chloromethane is the most abundant halogenated carbon compound in the atmosphere (500-600 ppt) (Montzka *et al*., 2003). Chloromethane production is mainly associated with biological activities of plants and white rot fungi, and with the degradation of plant organic matter in soils (Harper 2000; Keppler *et al*., 2005; Schäfer *et al*., 2007). Chloromethane affects ozone formation, and causes 16 % of chlorine-catalyzed ozone destruction (Harper *et al*., 2003; Hines *et al*., 1998; Redeker *et al*., 2000; Rhew *et al*., 2007).

Another biogenic, plant-derived VOC is methanol. Global methanol (CH_3OH) emissions are much larger than those of CH_3Cl. Methanol represents the second most abundant organic compound in the atmosphere (0.1–10 ppb) after methane (1800 ppb). Global emissions of CH_3OH are estimated at 5 Tg y^{-1}. The main source of atmospheric CH_3OH is also associated with plants (Galbally and Kirstine 2002; Jacob 2005; Wohlfahrt *et al*., 2015), and specifically originates from methoxy groups of pectin and lignin upon production of plant biomass (Donnelly and Dagley, 1980; Schink and Zeikus, 1980; Fall and Benson, 1996; Warneke *et al*., 1999).

A significant part of naturally produced CH_3OH does not reach the atmosphere but is dissolved in water and thus readily utilized as a growth substrate by many methylotrophic bacteria, in particular in terrestrial soil environments (Kolb and Stacheter, 2012). In the case of the more volatile CH_3Cl, a filtering effect mediated by CH_3Cl degrading methylotrophs is also likely, in particular in the aerial parts of plants (Nadalig *et al.*, 2011; Farhan Ul Haque *et al.*, 2013; Bringel and Couée, 2015). CH_3Cl-utilizing bacteria have been isolated from soil of industrial sites (McDonald *et al*., 2001) and from woodland and forest soils (Borodina *et al*., 2005). Thus, microbial-driven reduction of CH_3Cl emissions from soils is in principle possible. However, this process remains to be characterized in detail, and its significance in soils has not yet been investigated.

Soil is a physicochemically heterogeneous environment, and top soil is highly biologically active. The number of bacterial cells per gram of top soil (Ah horizon) likely exceeds 10^{11}, and soil microbial diversity has been estimated at between 10^5 and 10^6 species of gram of soil bacteria and fungi, respectively (Peñuelas *et al.*, 2014). Forest soil is a typical habitat of methylotrophic microorganisms capable of using C_1 compounds for growth. Most of them are facultative methylotrophic, and strictly aerobic (Kolb *et al*., 2009). Their abundance was estimated in the range between 10^6 cells and $3 \cdot 10^8$ cells $g.soil^{-1}$ (Stacheter *et al*., 2013). There is evidence that methylotrophic bacteria are highly active in soil, and that they are able to use a large set of substrates, including CH_3OH and also some halogenated compounds including CH_3Cl (Kolb, 2009; Stacheter *et al*., 2013).

To date, the magnitude and kinetics of transformations involving CH_3Cl in forest soils remains a black box. CH_3Cl degradation in forest soil was estimated at around 137 $Gg.y^{-1}$ (Yoshida *et al*., 2004). Biotic and abiotic processes occurring in forest soil are complex and subject to high temporal and spatial variability, and contribute to the difficulty in estimating the global budget of CH_3Cl (Redeker *et al*., 2012). It has already been shown that exchange of CH_3Cl between forest ecosystems and the atmosphere are modulated by bacteria in forest soil (Schäfer *et al*., 2007). A biochemical pathway for CH_3Cl utilization (the *cmu* pathway) has been characterized in detail in *M. extorquens* CM4 (Vannelli *et al.*, 1999; Roselli *et al.*, 2013; see Chapter 1 and Chapter 3). This pathway was subsequently reported for several CH_3Cl-degrading isolates (reviewed in Schäfer, 2007) including isolates from forest soil (Doronina *et al*., 1996; Coulter *et al*., 1999, McAnulla *et al*., 2001; Miller *et al*., 2004). There is increasing evidence that the *cmu* pathway is not the sole pathway for CH_3Cl utilization (Chapter 1, Section 1.3.4., and Nadalig *et al.*, 2014).

Different approaches including degradation kinetics, use of pure bacterial cultures and stable isotope tracer experiments with aerated soil have shown that aerobic methylotrophic microbes in soil are able to use CH_3Cl at low concentrations such as those encountered in the environment (Shorter *et al*., 1995; Hines *et al*., 1998; Yoshida *et al*., 2006; Miller *et al*., 2004). Environmental CH_3Cl concentrations are so low that it is likely that CH_3Cl-degrading methylotrophs also rely on some of the very many other carbon compounds present in forest soil as growth substrates. For example, *Methylomicrobium album* BG8, for which CH_3Cl cannot

serve as the sole growth substrate, displays enhanced growth with CH_3Cl in the presence of CH_3OH (Han and Semrau, 2003).

In this study, bacterial CH_3Cl utilization in Steigerwald forest soil (Germany) and the effect of CH_3OH as a methylotrophic co-substrate in CH_3Cl utilization were investigated. To identify metabolically active key players in this process, taxonomic and functional gene biomarkers were employed in combination with a stable isotope labeling experiment. Genes encoding subunits of key enzymes of C_1 metabolism such as *mxaF* have been repeatedly used as biomarkers to detect methylotrophs in the environment (McDonald *et al*., 2008). Here, 4 key functional gene markers of methylotrophs (*cmuA*, *mxaF*, *xoxF*, and *mch*, Figure 4.1) were targeted by PCR in addition to the 16S rRNA gene. The gene *cmuA* encodes one of the two methyltransferases of the first reaction required for growth with CH_3Cl by the *cmu* pathway, and has already intensively been used as a biomarker for CH_3Cl-utilization (McAnulla *et al*., 2001a; Miller *et al*., 2004; Borodina *et al*., 2005; Schäfer *et al*., 2005; Nadalig *et al*., 2011). Genes *mxaF* and *xoxF* are both associated with the oxidation of CH_3OH for growth for a wide diversity of microorganisms. The gene *mxaF* encodes the catalytic subunit of pyrroloquinoline quinone (PQQ)-dependent CH_3OH dehydrogenase, and was proposed early on as a functional gene probe for methylotrophs (McDonald and Murrell, 1997). The *xoxF* gene, homologous to *mxaF*, occurs in the overwhelming majority of methylotroph genomes (Chistoserdova, 2011). Homologs of *xoxF* encode functional MDHs (Fitriyanto *et al*., 2011; Chu and Lidstrom 2016; Vu *et al*., 2016). The *xoxF* gene is present in a wide range of both methylotrophs and non-methylotrophs from *Alphaproteobacteria*, *Betaproteobacteria* and *Gammaproteobacteria* (Taubert *et al*., 2015). The gene *mch* encodes a methenyl cyclohydrolase involved in the interconversion of key intermediates of methylotrophic metabolism, and has already been used to resolve methylotrophic community structure and diversity (Kalyuzhnaya *et al*., 2004, 2005; Stacheter *et al*., 2013).

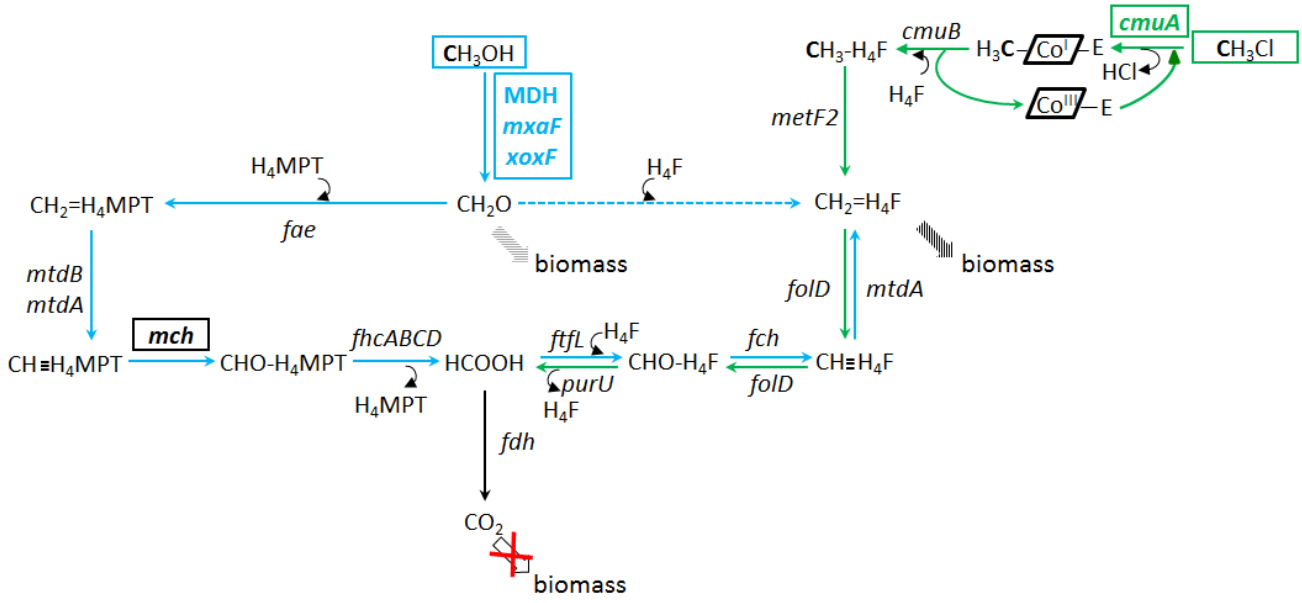


Figure 4.1. Targeted functional genes in methylotrophic utilization of chloromethane and methanol

The 3 genes *cmuA*, *mxaF*/*xoxF* and *mch* investigated by amplicon sequencing are boxed. Cofactors of central methylotrophic metabolism tetrahydrofolate (H_4F), tetrahydromethanopterin (H_4MPT) and metabolic intermediates including formaldehyde (CH_2O) and formate (HCOOH) are also indicated. The CH_3Cl utilization (*cmu*) pathway is shown in green, with the *cmuA* gene essential for growth with CH_3Cl as sole source of carbon and energy in *M. extorquens* CM4 (Vannelli *et al*., 1999). Reactions of the *cmu* pathway (green arrows) are shown on the right with associated genes encoding CH_3Cl dehalogenase subunits corrinoid-dependent methyltransferase (*cmuA*) and methyl-H_4F transferase (*cmuB*), methylene-H_4F reductase (*metF2*), bifunctional methylene-H_4F (CH_2=H_4F) dehydrogenase / methenyl-H_4F (CH ≡ H_4F) cyclohydrolase (*folD*) and 10-formyl-H_4F hydrolase (*purU*). Reactions and genes for transformation of methanol (CH_3OH) for energy production and incorporation into biomass are shown in blue, and include genes for oxidation reactions methanol dehydrogenase (*mxaF* and *xoxF* subunits) and methylene-tetrahydromethanopterin dehydrogenases (*mtdA* and *mtdB*). Other genes involved in oxidation of CH_3OH in *Methylobacterium* include formaldehyde activation enzyme (*fae*), methenyl cyclohydrolase (*mch*), and the formyltransferase-hydrolase complex (*fhcABCD)*. Formate dehydrogenase (*fdh*) is assumed to be used for the final oxidation step to CO_2 from formate originating from both CH_3OH and CH_3Cl. Genes involved in carbon assimilation from CH_3OH include formyltetrahydrofolate ligase (*ftfL)*, methylene-tetrahydrofolate cyclohydrolase (*fch*), and methylene-tetrahydromethanopterin dehydrogenases (*mtdA*). In *Methylobacterium*, carbon assimilation into biomass occurs by the serine cycle and the ethylmalonyl-CoA pathway via methylene-H_4F (striped thick arrow). In other organisms, carbon assimilation can occur by other pathways via formaldehyde or CO_2 (grey and white thick arrows). In the performed SIP labeling experiment, assimilation into biomass of CO_2 originating from mineralization of labeled carbon was avoided (red cross) as described in Material and Methods.

In previous studies, the *cmuA* gene probe and DNA stable isotope probing (DNA SIP) with fully [^{13}C]-CH_3Cl have been employed to target the diversity of CH_3Cl-utilizing bacteria in soil (Miller *et al*., 2004; Borodina *et al,* 2005). The SIP method makes use of the fact that the DNA of active bacteria that are able to assimilate the provided [^{13}C]-labeled carbon source will be enriched in [^{13}C]-carbon in their DNA. This "heavy" DNA can be separated by isopycnic ultracentrifugation from DNA of bacteria that did not assimilate the labeled compound. Such labeled DNA from active microbial populations can be analyzed, either by direct sequencing or by PCR amplification of key genes of interest (Coyotzi *et al.*, 2016; see Chapter 2). Here, we explored the structural and functional diversity of CH_3Cl-utilizers in top soil samples of forest soil amended with CH_3Cl and CH_3OH, added alone or together as the [^{13}C]-labeled form, or as the regular, [^{12}C]-unlabeled form of the compound.

2. Materials and methods

2.1. Study site and sampling

Soil was sampled from the Steigerwald forest area, located between Würzburg and Nürnberg (49° 37' N, 10° 17' O). Despite extensive development in this area, only few anthropogenic activities and no industrial activities exist in the forest, which covers a total area of about 1152 km^2. Vegetation mainly consists of beech (*Fagus sylvatica*), with minor stocks of oak (*Quercus* sp.). Five samples were taken from the upper soil layer within a 20 m circle, mixed, homogeneized and sieved through a 2 mm screen.

2.2. Soil activation and microcosm set-up

Physico-chemical characterization of soils was performed as described in Chapter 2. Sieved soil was aliquoted in batches of 500 g in sealed flasks, and 1 % CH_3Cl (Sigma Aldrich) was added to the gas phase to activate soils. Flasks (1 L) were kept at 20°C, and concentration of CH_3Cl was followed by gas chromatography until all CH_3Cl was consumed (within 2 weeks). For each microcosm, 70 g activated soil (approx. 50 mL) was transferred to a 500 mL flask sealed by a Viton stopper. In total, 8 different microcosms were prepared in duplicate in this study, and amended with unlabeled or [^{13}C]-CH_3Cl (Sigma Aldrich) and/or CH_3OH (Campro Scientific) (Table 4.1). Each microcosm received five carbon pulses (18 mM of carbon per pulse and

carbon source) over a period of 23 days (Figure 4.2). After each pulse, 5 aliquots of 1 g soil were sampled and immediately frozen at -80°C until further analysis.

Table 4.1. Microcosm setup and one-carbon supplementation

Carbon source added	Total carbon added (mM)[a]	Gas phase supplementation[b]	Liquid phase supplementation[c]
$[^{13}C]$-CH_3Cl	18	$[^{13}C]$-CH_3Cl	H_2O
$[^{13}C]$-CH_3Cl + CH_3OH	36	$[^{13}C]$-CH_3Cl	CH_3OH
CH_3Cl + $[^{13}C]$-CH_3OH	36	CH_3Cl	$[^{13}C]$-CH_3OH
$[^{13}C]$-CH_3OH	18	-	$[^{13}C]$-CH_3OH
CH_3Cl	18	CH_3Cl	H_2O
CH_3Cl + CH_3OH	36	CH_3Cl	CH_3OH
CH_3OH	18	-	CH_3OH
Control	-	-	H_2O

[a] Carbon added per pulse (5 pulses were performed by microcosm; Figure 4.2).

[b] A volume of 6 mL CH_3Cl was added for each CH_3Cl pulse. In total, a gas phase volume of 56 mL (air with or without CH_3Cl) was added to 500 mL flasks at each pulse in order to maintain slight overpressure in the flasks.

[c] Either 1 mL of milliQ water or of a 216 mM CH_3OH stock solution was added to soil of each microcosm per pulse.

2.3. Chemical analyses

Microcosms were monitored by measuring CH_3Cl and CO_2 concentrations by gas chromatography (HP 5890 Series II, Agilent) using a Porapak Q 80/100 (Supelco, Bellefonte, Pe) column. Injection of a new carbon pulse was performed when remaining CH_3Cl was low in microcosms amended with a CH_3Cl pulse. The same time-point was chosen for microcosms amended with a CH_3OH pulse. For CH_3Cl analysis, 1 mL of the gas phase was analyzed with the gas chromatograph fitted with an Electron Capture Detector (ECD). The injector temperature was fixed at 250°C, and ECD temperature was fixed at 300°C with a flow rate of 20 $mL.min^{-1}$, and a helium - methane (95:5) mixture as the carrier gas. A calibration curve with CH_3Cl standards allowed to correlate peak areas and CH_3Cl concentration. For CO_2 analyses, a volume of 100 µl of the gas phase was injected at 150°C into the gas chromatograph fitted with a thermal conductivity detector (TCD), and eluted at 250°C with a flow at 15 $mL.min^{-1}$, using an argon - methane mixture (95:5) as the carrier gas.

Values of pH were routinely measured on sieved soil, before pre-incubation, and after every carbon pulse was consumed, using an InLab R422 pH electrode (InLab Semi-Micro; Mettler-Toledo, Gießen, Germany). Soil moisture content was determined by weighing the soil before and after drying (at 60°C for 10 days).

2.4. [^{13}C]-CO_2 analysis

Analyses were performed at the UFZ - Helmholtz Centre for Environmental Research, Department of Molecular Systems Biology, Leipzig (Germany). GC-MS analysis was carried out using a Perkin–Elmer GC Clarus 600 system with an Rtx®-1 capillary column (60 m x 320 µm), operated with an ionization energy of 70 eV. Mass spectra were taken from 14 to 70 Da. Helium was used as the carrier gas with a constant flow at 300 kPa, and 10 µl gas samples (split ratio 10:1) were manually injected using gas-tight syringes. Each sample was measured five times. Total [^{12}C]-CO_2 and [^{13}C]-CO_2 were analyzed by extraction of masses 44 and 45 followed by peak integration. Peak areas were corrected for [^{12}C]-CO_2 and [^{13}C]-CO_2 indoor air values to calculate the 45/44 mass ratio of CO_2.

2.5. Nucleic acid extraction and RNA removal

Nucleic acids were extracted from soil after the third carbon pulse (chloromethane dissipation differed between each duplicate after the third pulse) was consumed in microcosms amended with CH_3Cl (Figure 4.2). DNA from both duplicates of each of the 8 microcosms was extracted according to the method described by Griffith *et al*. (2000), except that centrifugation was performed at 13'000 *g* and that NaCl was not added in the final precipitation step. Nucleic acids were resuspended in 30 µL DEPC-treated water and heat-treated for 3 min at 65°C. RNAs were removed by RNaseA treatment at 37°C for an hour, and DNA was precipitated with 21 µl isopropanol and 6 µl 5 M NaCl overnight at -20°C. After centrifugation at 14'000 *g* for 1 h at 4°C, the DNA pellet was washed with 70 % ethanol, air-dried at room temperature and resuspended in 30 µL DEPC-treated water. DNA was quantified by fluorescence using the Quant-iT™ PicoGreen® kit (Invitrogen).

2.6. DNA fractionation

A cesium chloride gradient solution was prepared as described by Neufeld *et al*., 2007). For each microcosm, a volume of 5 mL of this solution, containing 2 µg of RNA-free microcosm DNA in 20 µL of DEPC-treated water, was loaded in an ultracentrifugation tube (Quick-Seal, Beckman Coulter). Tubes were placed in a Vti 65.2 vertical rotor (Beckman Coulter) and centrifuged for 40 h at 177'000 *g* (LE-70 ultracentrifuge, Beckman Coulter) followed by 2 h of deceleration so as not to disturb the cesium chloride density gradient. Tubes were pierced at the bottom using a sterile 0.63 mm diameter needle, and a dye-containing water solution was injected from the top of the tube using a low-flow pump (BioRad Econo Pump), with a flow fixed at 450 µL.min^{-1}. Ten 450 µL fractions were collected for each tube for densities previously defined to be associated with heavy, middle and light DNA fractions (Neufeld *et al*., 2007). DNA from each fraction was precipitated as described (Neufeld *et al*., 2007), resuspended in 20 µL, and the defined 4 heavy, 3 middle and 3 light fractions were pooled together. Pooled heavy, middle and light DNA fractions of each microcosm were quantified with the Quant-iT™ PicoGreen® kit (Invitrogen). DNA concentrations ranged between 0.3 - 3.0 ng/µl for heavy fractions, 14.5 - 25.8 ng/µl for middle fractions and 18.5 -50.3 ng/µl for light fractions, respectively.

2.7. PCR amplification

PCR reactions of 16S rRNA gene, *mxaF*/*xoxF* and *mch* genes (Table 4.2) were carried out in 0.2 mL microcentrifuge tubes on a LabCycler thermal cycler (SensoQuest, Göttingen, Germany). The obtained PCR products were checked by agarose gel electrophoresis. The

50 µL reaction mix for amplification of the 16S rRNA gene consisted of 23.75 µL distilled water, 5 µL DNA template solution (between 1.5 - 15 ng), 2.5 µL of 10 µM degenerate primer stocks 341for and 785/805rev (Table 4.2) using Crystal Taq DNA polymerase (Biolab products GmbH, Lüneburg, Germany) as recommended by the supplier. After initial denaturation (95°C, 5 min), DNA amplification was achieved by 40 cycles of 1 min steps of denaturation at 95°C, annealing at 55°C, and extension at 72°C, followed by a final extension step of 5 min at 72°C. Amplification of the *cmuA* gene was performed as follows. A reaction mix of 5 µL of 10x buffer provided by the supplier (1x: Tris- HCl 10 mM, KCl 50 mM, $MgCl_2$ 1.5 mM, Triton 0.1 %, BSA 0.2 mg· mL^{-1}, pH 9.0, MP Biomedicals) 0.5 µL dNTPs (20 mM), 3.25 µL forward primer cmuAF422 and 0.75 µL of the reverse primer cmuAR422 (Table 4.2) as 20 µM stocks, 5 µL DNA

template solution (between 1.5 - 15 ng), and 33.5 µL of distilled water was prepared. After initial denaturation (95°C, 10 min), 2 µL TAQ polymerase mix (prepared as 2 µL of TAQ polymerase (5U/µL, MP biomedicals) added to 1 µL of 10x buffer and 7 µL H_2O) was added to the reaction mix. DNA amplification was then performed for 30 cycles consisting of 15 s denaturation at 95°C, annealing/extension for 30 s at 56°C, followed by a final extension step of 5 min at 72°C.

Table 4.2. Biomarker gene primer sequence, PCR amplicon size and OTU occurrence

Gene	Function	PCR Primer	Sequence (5'-3')[a]	Amplicon size (bp)	SNP/ OTU[b]	Total OTUs[c]	Labeled OTUs[d]	Reference
16S rRNA	ribosomal small subunit	341for	CCTACGGGNGGCWGCAG	464	9	117	5	Muyzer *et al.*, 1988
		785/805rev	GACTACHVGGGTATCTAATCC					Herlemann *et al.*, 2011
cmuA	chloromethane methyltransferase	cmuAF422	GARGTBGGITAYAAYGGHGG	422	38	8	5	This study[f]
		cmuAR422	TCRTTGCGCTCRTACATGTCICC					
mch	methylotrophy pathways	mch-2a	TGCCTCGGCTCKCAATATGCYGGBTGG	549	100	8	4	Vorholt *et al.*, 1999
		mch-3	GCGTCGTTKGTKCKBCCCAT					
mxaF_ xoxF[e]	methanol dehydrogenase	mdh1	GCGGIWSCAICTGGGGYT	430	39	6	6	This study[f]
		mdh2	GCGGIWSGAICTGGGGYT					
		mdhR	GAASGGYTCSYARTCCATGCA					

[a] Degenerate base mixtures: B (C,G,T), H (A,C,T), K (G,T), N (A,C,G,T), R (A,G), S (G,C), Y (C,T), V (A, C, T), W (A,T). Inosine (I) was used instead of the N mixture (Zheng *et al.*, 2008) in some cases.

[b] Maximal Single Nucleotide Polymorphism (SNP) positions possible within an OTU.

[c] Corresponding to the sum of OTUs detected in the 8 microcosms of the SIP experiment. Sequences were affiliated to the same OTU at cutoff values of 98 %, 90 %, and 80 % sequence identity at the nucleotide level for 16S rRNA gene, *cmuA* and *mxaF_xoxF*, and *mch* amplicons, respectively.

[d] See Materials and Methods for the criteria applied to define OTUs as "labeled".

[e] Primer pairs allow to amplify both *mxaF* and *xoxF* types of methanol dehydrogenase (MDH) subunits. Amplifications were performed with two different forward primers mdh1 and mdh2 in order to reduce primer degeneracy and improve PCR efficiency. Amplicons obtained with primer pairs mdh1/mdhR and mdh2/mdhR were pooled before sequencing.

[f] See Chapter 2 for details.

Amplifications of *mxaF*/*xoxF* and *mch* genes were performed in 20 µL reactions using 8 µL of 2.5x 5 Prime Mastermix ($MgCl_2$ final concentration 1.5 mM), 1.5 µL 3 % BSA (Biolab products GmbH, Lüneburg, Germany), 1 µL each of forward and reverse primers (10 µM; Table 4.2 for primer sequences), 5 µL DNA template solution (between 1.5 - 15 ng), 0.5 µL distilled water, and 3 µL 5x TaqMaster PCR Enhancer (5 Prime GmbH, Hamburg, Germany). In the case of *mxaF*/*xoxF* genes, two different PCR reactions with two different forward primers mdh1 or mdh2 (Table 4.2) and the same reverse primer mdhR (Table 4.2 for primer sequences) were performed, yielding amplicons for both *mxaF* and *xoxF* types of methanol dehydrogenase genes. After initial denaturation (95°C, 5 min), DNA amplification was performed by 40 cycles of 1 min denaturation at 95°C, annealing for 1 min at 55°C, extension for 1 min at 72°C and a final extension step of 5 min at 72°C.

2.8. DNA sequencing

PCR products of mdh1/mdhR and mdh2/mdhR amplicons were mixed prior to sequencing. In total, 24 samples each of 16S rRNA gene, *cmuA*, *mch*, and *mxaF_xoxF* amplicons were sequenced using Illumina MiSEQ technology (see Chapter 2) by a commercial supplier (LGC Genomics GmbH). Briefly, a barcode oligonucleotide identifying sequence was ligated to each PCR product. Equimolar pools of all libraries were assembled, and the resulting combined library was analyzed using a standard MiSeq paired end (2×300 bp) flow cell and reagent cartridge.

2.9. Analysis of the 16S rRNA gene

Illumina reads were analyzed using mothur software package v.1.33.2 (Kozich *et al*., 2013) with the default parameters of the MiSeq standard operating protocol (http://www.mothur.org/wiki/MiSeq_SOP). Briefly, read pairs were assembled into contigs. Contigs shorter than 420 bp or longer than 460 bp were discarded. Sequences were pre-clustered in groups of sequences with up to 2 nucleotide differences. Chimera sequences predicted by UCHIME (Edgar *et al*., 2011) were removed. Remaining sequences were assigned using naïve Bayesian taxonomic classification on the bacterial reference database SILVA (SSU_Ref database v.119), at a bootstrap cutoff set at 80 %. Remaining sequences were classified using the reference database SILVA (SSU_Ref database v.119), at a bootstrap cutoff set at 80 %. Sequences affiliated to Bacteria and Archaea domains were selected and other

sequences discarded. OTUs were defined at 98 *%* sequence identity using the automated protocol within Mothur. This also yielded a representative consensus sequence, for each OTU, chosen as the most abundant sequence in a given OTU, and used for subsequent sequence alignments.

2.10. Analysis of functional genes

Raw reads were processed using mothur software package v.1.33.2 (Kozich *et al*., 2013), including length filtering and quality trimming, and allowing sequence lengths within 20 nucleotides of the expected length of the amplicon. USEARCH software (Edgar, 2010) was then used for clustering the obtained filtered reads. Sequences occurring only once within the complete dataset of gene amplicon libraries were considered artefactual and removed, whereas singletons in individual amplicon libraries were retained. Reads were clustered iteratively at progressively lower cut-off values, and the maximum cut-off value at which the number of retrieved OTUs stabilized (90 % for *cmuA* and *mxaF*/*xoxF*, 80 % for *mch*) was selected (Stacheter *et al*., 2013). Consensus sequences for each OTU were those provided by mothur. These sequences were compared against a gene-specific database generated from GenBank using BLAST (http://blast.ncbi.nlm.nih.gov) for taxonomic identification. A cluster table for each gene and each sample was obtained and analyzed as described below to identify labeled OTUs.

2.11. Identification of labeled OTUs

Labeled OTUs (OTU detected specifically and dependently to the microcosm tested) were defined by following a previously reported protocol developed to minimize false positives in 'heavy' DNA fractions (Lueders *et al*., 2004; Liu 2007; Dallinger and Horn 2014). Briefly, the relative abundances in the 'heavy' DNA fraction of a given [^{13}C]-labeled microcosm and in its corresponding unlabeled microcosm (amended with the corresponding unlabeled ([^{12}C]-compound), as well as the relative abundances in 'heavy' and 'light' DNA fractions of a given [^{13}C]-labeled microcosm, were compared. For this study, the following criteria had to be met in order to define OTUs in microcosms amended with a [^{13}C]-labeled compound as "labeled": (1) the relative abundance of the OTU in the 'heavy' DNA fraction of the [^{13}C]-labeled microcosm was higher than in the 'heavy' DNA fraction of the corresponding unlabeled microcosm; (2) the relative abundance of the OTU was higher in the 'heavy' DNA fraction than

in the 'light' fraction of the [^{13}C]-labeled microcosm; (3) the relative abundance of the OTU in the 'heavy' DNA fraction of the [^{13}C]-labeled microcosm was ≥ 0.5 %; (4) the relative abundance difference of the OTU between 'heavy' and 'light' DNA fractions was ≥ 0.1 % higher in the [^{13}C]-labeled microcosm than in the corresponding unlabeled microcosm.

2.12. Statistical and phylogenetic analyses

Richness (as Sobs), Shannon index, and Simpson diversity indices were determined using the mothur routine as described (http://www.mothur.org/wiki/MiSeq_SOP), and with PAST software (http://folk.uio.no/ohammer/past). Shannon *H* index was $H = -\sum p_j \ln p_j$ and Simpson diversity $D_2 = 1/\sum p_j^2$, where p_j is the OTU relative abundance. Relationships between sequence datasets in different DNA fractions and microcosms were investigated by two-dimensional NMDS (non-metric multidimensional scaling) within mothur and visualized with Kaleidagraph (Synergy Software, Reading, PA, USA). Heatmaps were produced with R software (version 3.2.1, 2015-06-18) using packages ggplot2, RODBC, reshape, plyr and RColorBrewer. Consensus sequences of labeled OTUs were aligned with known reference sequences, and the alignment was used for Maximum likelihood analysis based on the Tamura-Nei model using MEGA 7 software (Tamura et Nei 1993; Kumar, Stecher, et Tamura 2016). Tree representations were generated and plotted within MEGA7.

3. Results

3.1. Microcosm setup for detection of CH_3Cl-utilizing methylotrophs in forest soil

CH_3Cl, a growth substrate for a minority of specialized methylotrophic microorganisms, is known to be naturally produced in top soils and especially in tree vegetated ecosystems such as forests. On the other hand, CH_3OH, another C_1 growth substrate of methylotrophs, is produced in much larger amounts in the same environments. A stable isotope labeling approach using the [^{13}C] isotopologue of CH_3Cl was chosen to identify microbial populations that assimilate carbon of CH_3Cl in temperate forest soils in the presence and absence of CH_3OH. Forest soil microcosms amended with [^{13}C]-CH_3Cl or [^{13}C]-CH_3OH were set up in order to (i) identify CH_3Cl-utilizers in forest soil, and ii) assess the impact of CH_3OH on CH_3Cl utilization and the corresponding soil microbial methylotroph community. Starting from sieved and homogeneized forest soil pre-activated for about 2 weeks by addition of 18 mM

CH_3Cl, a total of four [^{13}C]-labeled amendment conditions was tested on the same sieved soil: [^{13}C]-CH_3Cl, [^{13}C]-CH_3Cl with unlabeled CH_3OH, [^{13}C]-CH_3OH with unlabeled CH_3Cl, and [^{13}C]-CH_3OH (Table 4.1). Three control microcosms were treated with equivalent unlabeled carbon amendments, and another control microcosm was incubated without carbon amendment (Teble 4.1). Comparison with unlabeled microcosms was required for definition of labeled OTUs of used gene markers (see below, Section 4.3.3). CH_3Cl and/or CH_3OH were supplemented with 5 pulses of 18 mM unlabeled or [^{13}C]-labeled C_1 compound over a period of 23 days (Figure 4.2).

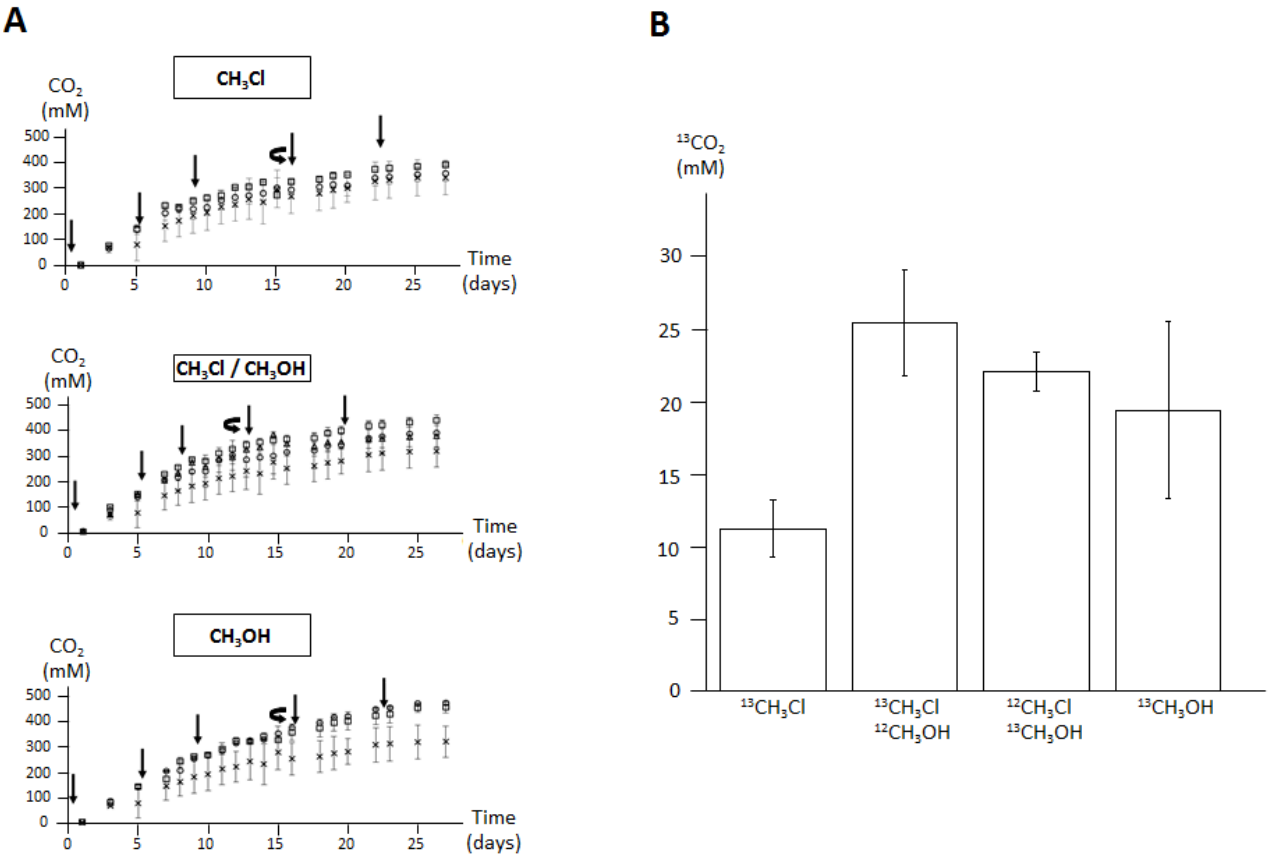


Figure 4.2. Carbon mineralization in the SIP experiment
(A) Time-course of total ([^{12}C]-CO_2 and [^{13}C]-CO_2) production of CO_2 in the 8 microcosms. When only CH_3Cl or CH_3OH was added (top and bottom panels), squares indicate CO_2 production from [^{13}C]-labeled C_1 compounds, and circles production from unlabeled compounds, respectively. When both CH_3OH and CH_3Cl were added (middle panel), squares represent the production from [^{13}C]-CH_3Cl and unlabeled CH_3OH, triangles production from unlabeled CH_3Cl and [^{13}C]-CH_3OH, and circles production from unlabeled compounds, respectively. Data for the control microcosm (no carbon addition, crosses) are given in all three graphs for comparison. The 5 carbon pulse time points are indicated (vertical arrow). Significant differences in microcosm duplicates were observed after the fourth pulse (data not shown). Thus, DNA extractions and PCR amplifications were performed after the third carbon pulse was consumed (curved arrow). (B) Carbon mineralization as [^{13}C]-labeled CO_2 is shown

only for the 4 microcosms to which [^{13}C]-labeled carbon was added. Data are shown at the time-point of DNA extraction, as the total [^{13}C]-labeled CO_2 after 3 carbon pulses (54 mM [^{13}C]-labeled carbon supplemented in total). Vertical bars show standard deviation of 2 biological replicates.

3.2. Mineralization and assimilation of chloromethane in forest soil microcosms

When methylotrophic microorganisms present in the microcosm actively utilize amended CH_3Cl and CH_3OH as energy source, this leads to CO_2 production (Figure 4.1; also detailed for strains of the genus *Methylobacterium* in Chapter 3). CO_2 production in microcosms was determined by gas chromatography (Figure 4.2A). As expected, a net increase in CO_2 production was observed in all microcosms in response to carbon supplementation (Figure 4.2A). Overall, a trend towards larger CO_2 release upon supplemental CH_3OH than upon supplemental CH_3Cl was observed. There was no significant difference in CO_2 production in microcosms to which labeled or unlabeled compounds were added, and thus the use of the [^{13}C] isotope did not affect carbon metabolism overall. Observed rates of CO_2 production in microcosms were in the range of 0.2-0.3 mmol g_{soil}^{-1} day^{-1} for microcosms to which CH_3Cl and CH_3OH were added.

Amendment with [^{13}C]-labeled compounds allowed to estimate the proportion of the carbon amendment that was used for energy production, based on produced [^{13}C]-CO_2 (Figure 4.2). The errors in the determinations are large, but the calculations suggest that about 10 mM of the total added 54 mM [^{13}C]-CH_3Cl were converted to [^{13}C]-CO_2 in the [^{13}C]-CH_3Cl amended microcosm. Thus, about 80 % of the supplemented carbon was assimilated into biomass. A significant increase in the extent of mineralization from labeled CH_3Cl (about 50 %) was observed when unlabeled CH_3OH was also provided. In the reverse case of microcosms amended with [^{13}C]-CH_3OH, however, the situation was different. Only a minor increase in mineralization was observed in the microcosm to which unlabeled CH_3Cl was provided together with [^{13}C]-CH_3OH. Moreover, methanol was mineralized to a significantly larger extent than CH_3Cl when provided alone.

3.3. Taxa associated with chloromethane utilization in forest soils

Taxonomical identification was based on sequence analysis of an approximately 460 nucleotide-long segment of the 16S rRNA gene hypervariable region that was PCR amplified.

The amplicons from heavy, middle and light gradient fractions retrieved of each of the 8 microcosms (Table 4.2) were sequenced by Illumina MiSeq, generating tens of thousands of sequences per amplicon (Table 4.3; Table S4.1).

Table 4.3. Number of filtered amplicon sequences obtained from heavy DNA fractions

Microcosm	*rrsA*	*cmuA*	*mxaF_xoxF*	*mch*
$[^{13}C]$-CH_3Cl	46110	4282	786	55892
$[^{13}C]$-CH_3Cl + CH_3OH	178042	3778	1534	38644
CH_3Cl + $[^{13}C]$-CH_3OH	76690	614	918	52586
$[^{13}C]$-CH_3OH	98332	1886	588	43088
CH_3Cl	72522	370	2616	27732
CH_3Cl + CH_3OH	47236	922	862	44356
CH_3OH	90822	246	2946	26518
Control	100032	184	2710	26804

See Table S4.1 for details of numbers of raw and filtered reads obtained in all DNA fractions.

The analysis of the recovered diversity in all DNA fractions at the 98 % sequence identity level suggested that, with one possible exception (i.e. the heavy fraction of the microcosm amended with $[^{13}C]$-CH_3OH and CH_3Cl, see below, section 4.3.4), datasets clustered according in a self-consistent pattern (Figure 4.3).

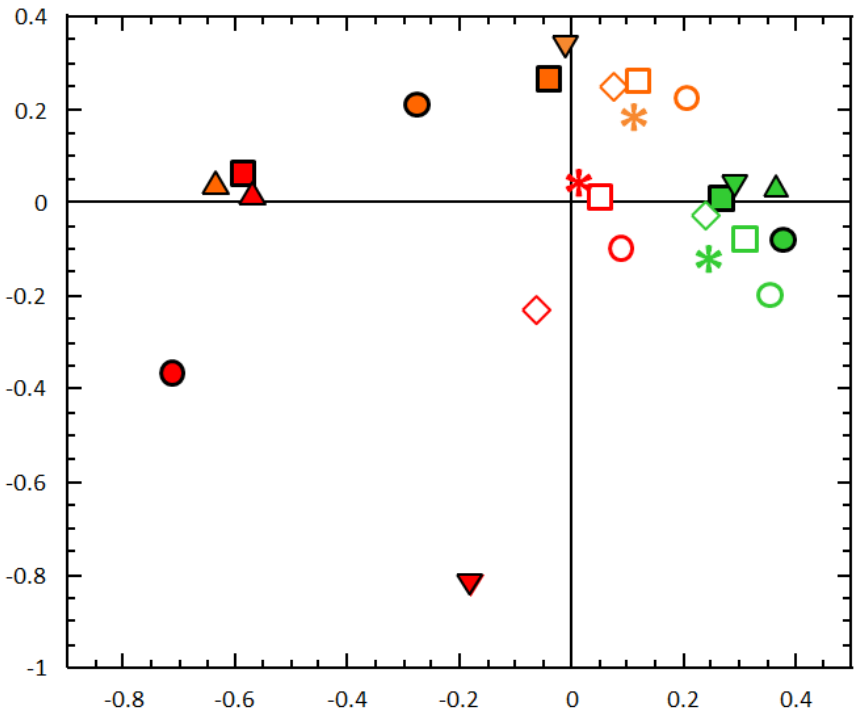


Figure 4.3. Clustering analysis of 16S rRNA gene amplicon sequence datasets from all microcosms
Relationships between sequence datasets were investigated by two-dimensional NMDS (non-metric multidimensional scaling) analysis within mothur. [^{13}C]-labeled microcosms and unlabeled microcosms are shown with full and empty symbols, respectively: microcosms supplemented with CH_3Cl only (squares); with [^{13}C]-labeled CH_3Cl and CH_3OH (upward pointing triangles); CH_3Cl with [^{13}C]-labeled CH_3OH (downward pointing triangles); with unlabeled CH_3Cl and CH_3OH (diamonds); with CH_3OH only (circles); and the unamended control (asterisks). Heavy, middle, and light fractions are shown in red, orange and green, respectively. Stress: 13.6 %.

A large and similar effect on community structure (Figure 4.3) was noticed on heavy fractions of microcosms labeled with [^{13}C]-CH_3Cl alone, or together with unlabeled CH_3OH. The effect of addition of [^{13}C]-CH_3OH alone was also prominent. Computed diversity indices of all DNA fractions then confirmed the observation from cluster analysis (Figure 4.3) that there was no significant difference in the recovered diversity of light fractions of all microcosms, as well as the unamended control (Table 4.4). The light fractions of [^{13}C]-labeled microcosms, as well as the middle fractions that were not investigated further, were also little affected by the amendments (Figure 4.3). In contrast, a significant decrease in diversity was observed in the heavy fractions of microcosms labeled with [^{13}C]-CH_3Cl alone or together with unlabeled CH_3OH, and also in the microcosm to which CH_3OH alone was added (Table 4.4).

Table 4.4. Diversity indices for 16S *rrnA* amplicon sequences obtained from heavy and light DNA fractions

Microcosm	SIP fraction	Sobs[a]	Shannon index[a]	Simpson diversity[a]
[^{13}C]-CH_3Cl	H	1762	5.95	44
	L	2870	7.82	223
[^{13}C]-CH_3Cl + CH_3OH	H	1629	6.11	36
	L	3150	7.55	354
CH_3Cl + [^{13}C]-CH_3OH	H	1567	5.13	12
	H[b]	1566	6.52	120
	L	3062	7.86	283
[^{13}C]-CH_3OH	H	1215	4.58	7
	L	3148	7.40	322
Unamended control	H	1804	6.45	102
	L	2904	7.39	217

[a] Calculated at the 98 % sequence identity level. See Materials and Methods for definitions. Simpson diversity is considered a conservative measure of the effective number of phylotypes (Haegeman *et al.*, 2014).

[b] Computed after removal of the *Curtobacterium* genus-level OTU from Microbacteriaceae, detected at 29.8 % relative abundance in the heavy DNA fraction of the [^{13}C]-CH_3OH labeled microcosm, but at 0.6 % relative abundance in all other 23 DNA fractions (see text).

After clustering at 98 %, a total of 117 family-level 16S RNA OTUs were detected in the eight microcosms (Table 4.2, Figure S4.1). However, the large set of sequence data obtained by Illumina MiSeq of PCR amplicons (Table S4.1) makes it difficult to assess directly which taxa are associated with CH_3Cl and/or CH_3OH utilization (Figures S4.1 to S4.4). The specific procedure for data analysis described in Material and Methods, basing on the approach described by Liu *et al.,* (2011), allowed to define and identify labeled taxa (OTUs) in a conservative manner. In total and across all [^{13}C]-labeled microcosms, only 5 families were defined as labeled (Figure 4.4).

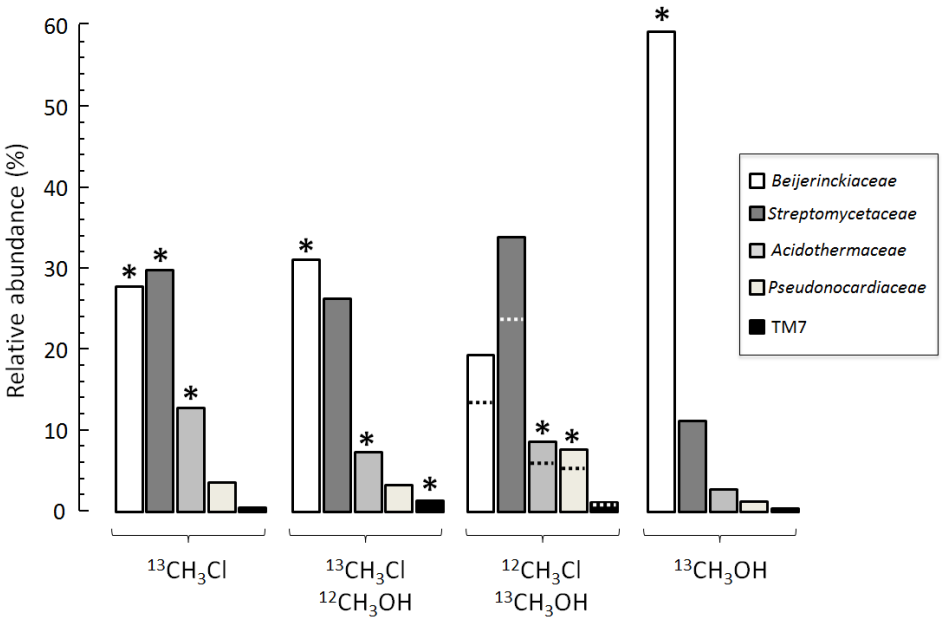


Figure 4.4. Taxonomic affiliation of labeled family-level 16S rRNA OTUs
Only microcosms exposed to [^{13}C]-labeling are shown. In total, 5 of the recovered 117 family-level OTUs satisfied the criteria defined for a [^{13}C]-labeled OTU (see Materials and Methods). Labeled OTUs in each microcosm are indicated by a star (see also Table 4.4). A 6th OTU (*Microbacteriaceae*), detected as highly abundant in the heavy fraction of the [^{12}C]-CH_3Cl/[^{13}C]-CH_3OH microcosm only, and thereby predicted to be labeled, is considered potentially doubtful (see text). Dotted lines in this microcosm indicate relative abundances of OTUs if this *Microbacteriaceae* OTU is included.

These were *Beijerinckiaceae* (*Alphaproteobacteria),* three families of *Actinobacteria* (i.e. *Acidothermaceae*, *Pseudonocardiaceae* and *Streptomycetaceae*) and a minor, uncharacterized family-level OTU belonging to the phylum TM7 (Table 4.5). A 6th OTU (*Microbacteriaceae*) was detected at lower than 0.6 % relative abundance in 22 gradient fractions, including the heavy fraction of the equivalent unlabeled microcosm, as well all other fractions of unlabeled and labeled microcosms amended with both CH_3Cl and CH_3OH. However, the *Microbacteriaceae* OTU was detected at very high (29.8 %) relative abundance in one heavy fraction of the [^{12}C]-CH_3Cl/[^{13}C]-CH_3OH microcosm, and thereby predicted to be labeled. Although this label may be considered as doubtful, it might indicate that these organisms are chloromethane degraders that dissimilate but not assimilate CH_3Cl (see Discussion below).

Four out of five family-level labeled taxa correspond to frequently labelled genera. *Methylovirgula*, *Acidothermus*, and *Streptomyces* represent over 95 % of the highly dominant

labeled family-level OTUs *Beijerinckiaceae*, *Actinomycetales* and *Streptomycetaceae*, respectively (Table 4.5). In the microcosm supplemented only with [^{13}C]-CH_3Cl, 3 of the 5 family-level OTUs (*Acidothermaceae*, *Beijerinckiaceae* and *Streptomycetaceae*) made for about 80 % of the total relative OTU abundance (Table 4.5).

Table 4.5. Dominant genera within labeled family-level [^{13}C] OTUs

Genus (*family*)	**Relative abundance (%) in heavy fractions of [^{13}C]-labeled microcosms[a]**			
	[^{13}C]-CH_3Cl	**[^{13}C]-CH_3Cl + CH_3OH**	**CH_3Cl + [^{13}C]-CH_3OH**	**[^{13}C]-CH_3OH**
Methylovirgula (*Beijerinckiaceae*)	27.0 (**27.9**)	29.9 (**31.1**)	(13.5)	58.5 (**59.4**)
Streptomyces (*Streptomycetaceae*)	29.9 (**29.9**)	(26.1)	(23.8)	(11.4)
Acidothermus (*Acidothermaceae*)	12.3 (**12.9**)	7.2 (**7.3**)	5.9 (**6.0**)	(2.8)
Pseudonocardia (*Pseudonocardiaceae*)	(3.5)	(3.1)	4.8 (**5.3**)	(1.4)
Curtobacterium (*Microbacteriaceae*)[b]	(0.3)	(0.6)	29.8 (**30.0**)	(0.2)
TM7[c]	(0.3)	**(1.3)**	(0.8)	(0.3)
(Total of family-level taxa)	(74.5)	(68.9)	(79.4)	(75.3)

[a] Percentage of total filtered reads affiliated by GreenGenes to the same dominant genus (featuring several OTUs at >98 % sequence identity) in the 5 family-level taxa defined as labeled in all microcosms (in brackets: % of total reads; labeled family-level taxa in each microcosm are shown in bold).

[b] The family-level OTU *Microbacteriaceae* was detected at 30 % relative abundance only in the heavy DNA fraction of this microcosm labeled with [^{13}C]-CH_3OH, and at ≤0.6 % relative abundance in the heavy fraction of the corresponding unlabeled microcosm as well as in all other 22 DNA fractions that were sequenced (see text and Figure 4.2).

[c] Uncharacterized family-level taxon (>80 % sequence identity) within the TM7 phylum.

In the microcosm supplemented only with [^{13}C]-CH_3OH, in contrast, only *Beijerinckiaceae* was labeled and amounted to an even larger percentage than in the other labeled microcosms. The minor abundant family-level labeled OTU of phylum TM7 was only labeled in the [^{13}C]-CH_3Cl treated microcosm that was additionally supplemented with unlabeled CH_3OH. The labeled OTU affiliated with *Pseudonocardiacae* was a dominant taxon in all CH_3Cl-amended

microcosms, but was only detected as labeled in the microcosm supplemented with [^{13}C]-CH_3OH together with unlabeled CH_3Cl.

Labeled OTUs differed significantly from the closest type strains as well as from previously described CH_3Cl-degrading strains (Figure 4.5).

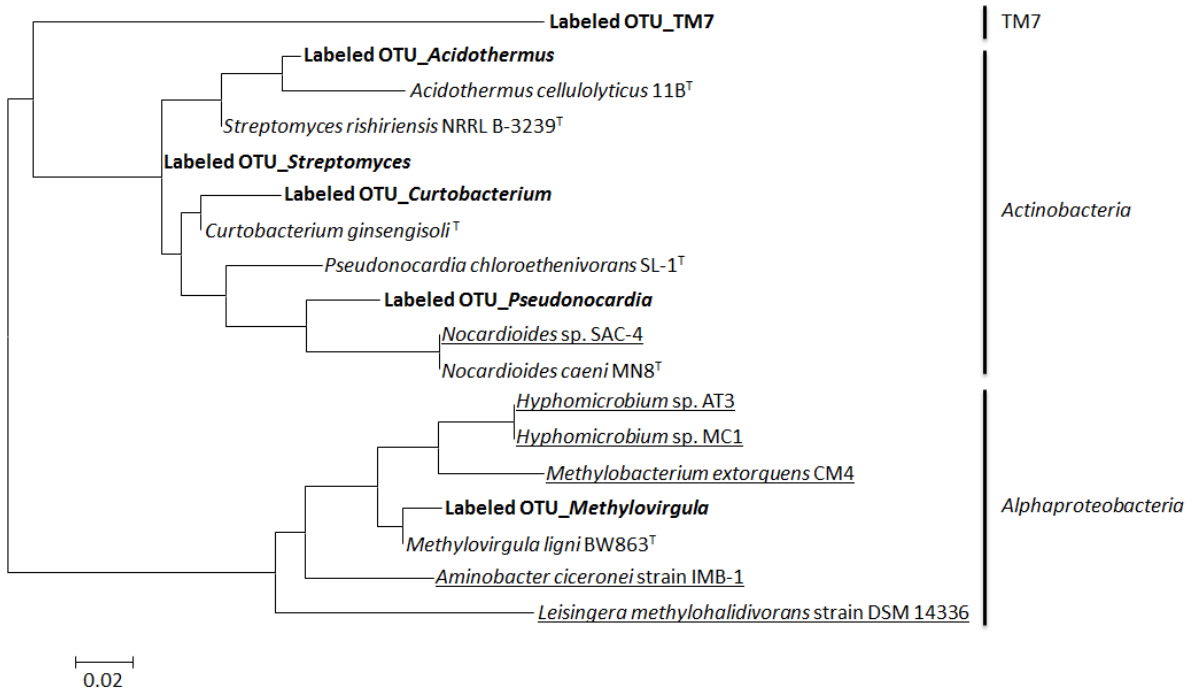


Figure 4.5. Phylogenetic relationships between partial sequences of the 16S RNA gene of labeled OTUs and of reference chloromethane-utilizing strains

Genus-level labeled OTUs are in bold. Their selected consensus 16S rRNA gene sequences (464 nucleotides) were aligned with those of corresponding type strains including *Acidothermus cellulolyticus* 11B^T (Acc. No. JQ899226), *Curtobacterium ginsengisoli*^{T} (EF587758), *Nocardiodes caeni* MN8^T (FJ423551), *Methylovirgula ligni* BSW863^T (FM252034), *Pseudonocardia chloroethenivorans* SL-1^T (AF454510), *Streptomyces rishiriensis* NRRL B-3239^T (EF178682), and those of characterized CH_3Cl-degrading strains (underlined) such as *M. extorquens* CM4 (AF198624, Doronina *et al*., 1996), *Hyphomicrobium* sp. MC1 (NR_074190, Hartmans *et al.,* 1986, Vuilleumier *et al.*, 2011), *Hyphomicrobium* sp. AT3 (FN667865, Nadalig *et al*., 2011), *Leisingera methylohalidovorans* strain DSM 14336 (DQ915607, Schaefer *et al.,* 2002), *Aminobacter ciceronei* strain IMB-1 (AF034798, Hancock *et al.,* 1998), and *Nocardioides* sp. SAC-4 (F279792.1, McAnulla *et al*., 2001). A Maximum Likelihood phylogenetic tree representation was constructed from the alignment basing on the Tamura-Nei model (Tamura *et al*., 1993).

3.4. Diversity of *cmu* pathway chloromethane utilizers in forest soil microcosms

Chloromethane: cobalamin methyltransferase *cmuA*, the key gene biomarker for CH_3Cl consumption, was amplified with new primers specifically designed for this study (see Chapter 2). Based on the amplicon sequence data in the heavy fractions (Table 4.3), OTUs in all fractions were defined and analyzed as for the 16S rRNA gene, but with a cut-off of 90 % sequence identity. In total, 8 OTUs were recovered, among which 5 satisfied the criteria defined in this work for labeled OTUs (see Materials and Methods, Table 4.2, Figure 4.6 and S4.2).

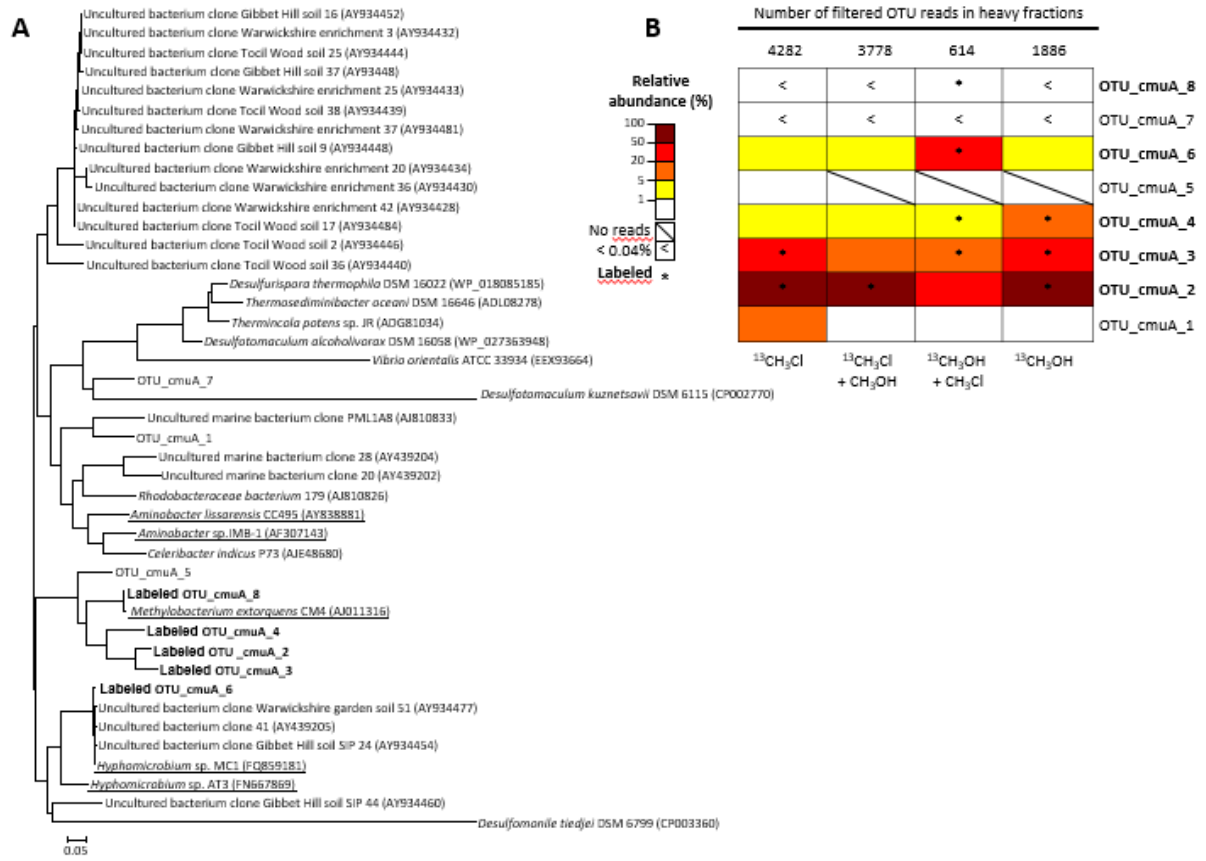


Figure 4.6. Affiliation and relative abundance of *cmuA* OTUs

OTUs shown in bold were defined as labeled according to the criteria defined in Materials and Methods. **(A)** Phylogenetic analysis of *cmuA* amplicon sequences. The tree was constructed from a 422 nucleotide sequence alignment of consensus sequences of the 8 labeled OTUs defined for the PCR amplicons obtained from microcosm heavy DNA fractions (OTU_cmuA_X; X ranging from 1 to 8); sequences of uncultured organisms detected in previous studies on CH_3Cl-utilization in three different soils (Gibbet Hill, Tocil Wood and Warwickshire from Miller *et al*., 2004) and Plymouth coastal water enrichments (Schäfer *et al*., 2005); and sequences from reference CH_3Cl degrading strains (*Aminobacter lissarensis*, *Aminobacter* sp. IMB-1, *Hyphomicrobium* sp. MC1, *Hyphomicrobium* sp. AT3 and *M. extorquens* CM4; underlined), as well as other strains of known genome sequence containing the *cmuA* gene. Branch lengths

are proportional to percentage sequence differences. Bootstrap was performed with 1,000 replicates. **(B)** Heatmap of the relative abundance of the 8 *cmuA* OTUs detected in the heavy DNA fractions in [^{13}C]-labeled microcosms. The total number of reads after PCR amplification of the *cmuA* gene (Table 4.2) is also shown.

Consensus sequences of these 8 OTUs were compared to *cmuA* sequences obtained in a previous SIP study of soil with [^{13}C]-CH_3Cl (Borodina *et al.*, 2005), and to the sequences of isolated strains of known genome sequence harboring a *cmuA* gene, several of which were shown to be CH_3Cl-degraders (Figure 4.6A). The 8 OTUs belong to 3 distinct groups. All labeled OTUs are closely related to the sequences of known CH_3Cl-degraders including *M. extorquens* CM4 (>99 % identity) or *Hyphomicrobium* sp. MC1, itself closely related to sequences and CH_3Cl-degrading strains previously obtained in the SIP CH_3Cl study of soil (Borodina *et al*., 2005). In the CH_3Cl condition, OTU_*cmuA*_2 and OTU_*cmuA*_3 were labeled and among the most dominant (relative abundance of 63.7 and 24 %, respectively). When CH_3OH was added to [^{13}C]-CH_3Cl, only OTU_*cmuA*_2 was still defined as labeled, and was even more abundant than in the microcosm amended only with CH_3Cl (over 83 % of all sequences, Figure 4.6B). In the microcosm supplemented with [^{13}C]-CH_3OH, in contrast, OTU_*cmuA*_3, OTU_*cmuA*_4, OTU_*cmuA*_6 were detected as labeled, whereas OTU_cmuA_2 was not, and OTU_cmuA_6 was found as abundant (35 % of reads). In the [^{13}C]-CH_3OH, OTU_cmuA_2, 3, 4 and 6 were detected, with a large abundance of OTU_*cmuA*_2 (64 %), but only OTU_cmuA _3 was labeled.

3.5. Diversity of methanol utilizers in forest soil microcosms

Sequences of the MDH amplicon in the heavy DNA fraction of [^{13}C]-labeled microcosms (Table 4.3) were grouped in 6 OTUs at the chosen cut-off of 90 % sequence identity (Table 4.2, Figure 4.7 and Figure S 4.3). Closest sequences to MDH_OTU_6 are from *xoxF* genes from *Hyphomicrobium* strains, whereas the 5 other OTUs cluster closely together, and are more closely related to *xoxF* sequences of *Methylobacterium* strains (Figure 4.7A). Thus, only *xoxF*-like sequences were recovered (Figure 4.7), despite the fact that primer pairs used to generate amplicons of genes encoding CH_3OH dehydrogenase (MDH) had the potential to amplify both *mxaF* and *xoxF* types of the target gene for a catalytic subunit of the enzyme (Table 4.2).

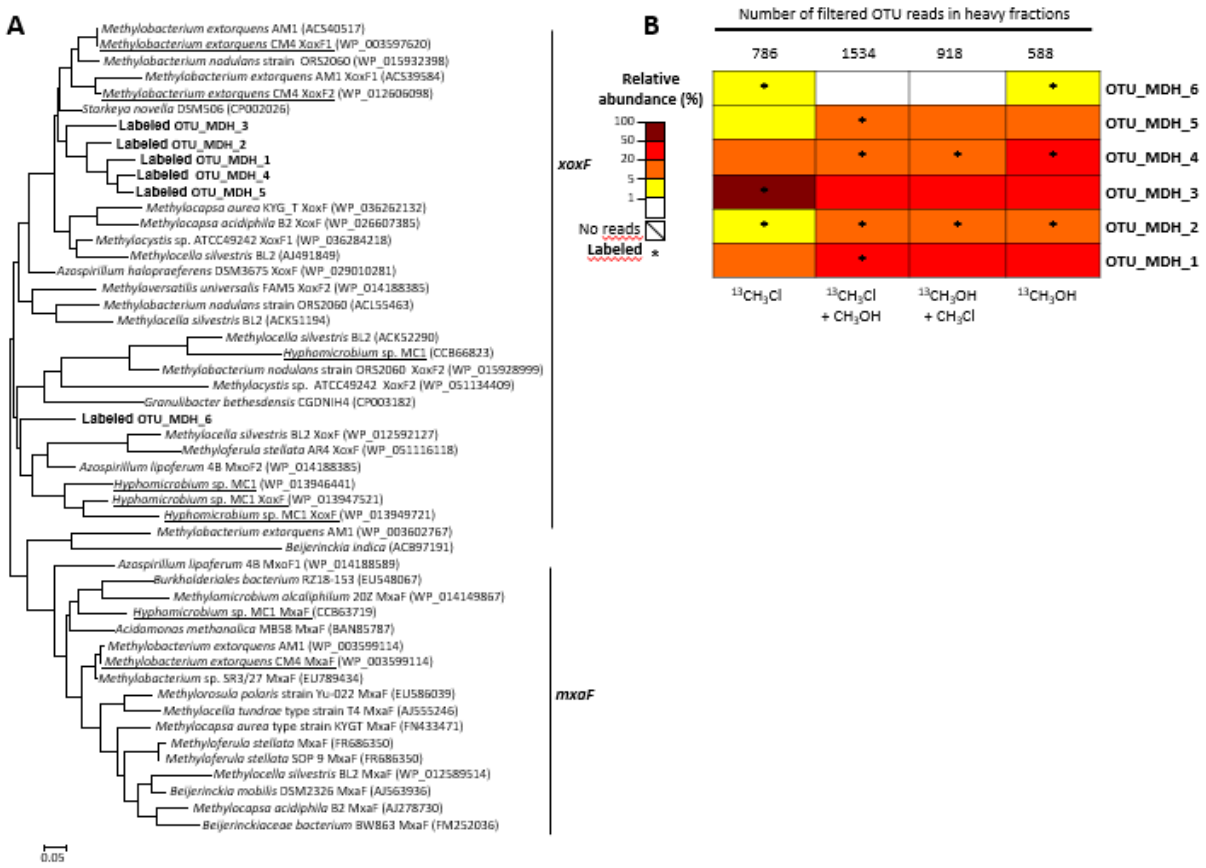


Figure 4.7. Affiliation and relative abundance of MDH OTUs
(A) Phylogenetic analysis of methanol dehydrogenase (MDH) amplicon sequences was performed with 46 unique 549 nucleotide long sequences including *xoxF* and *mxaF* sequences from reference strains (known CH_3Cl-degraders are underlined) and consensus sequences for OTUs (OTU_MDH) defined in heavy DNA fractions at the >90 % sequence identity level. All recovered MDH OTUs were defined as labeled according to the criteria defined in Materials and Methods. No *mxaF* or *xoxF* sequences are detected in CH_3Cl-degraders *Aminobacter lissarensis*, *Aminobacter* sp. IMB-1 and *Hyphomicrobium* sp. AT3. The tree was obtained from the alignment by the Maximum Likelihood method (Tamura Nei model, Tamura *et al*., 1993). Branch lengths are proportional to sequence differences. Bootstrap was performed with 1,000 replicates. **(B)** Heatmap of the relative abundance of OTUs in heavy DNA fractions of [^{13}C]-labeled microcosms.

The 6 MDH OTUs were defined as labeled in one or several of the [^{13}C]-labeled microcosms. OTU_MDH_2, OTU_MDH_3 and OTU_MDH_6 were labeled in the [^{13}C]-CH_3Cl treatment, with OTU_MDH_3 the most abundant OTU detected (69.7 %, see Table S4.2). In the condition in which unlabeled CH_3OH was added on top of [^{13}C]-CH_3Cl to forest soil, in contrast, OTU_MDH_3 and OTU_MDH_6 were not labeled, but the other OTUs were (Figure 4.7B). Of these, OTU_MDH_2 and OTU_MDH_4 were also labeled in the reciprocal condition (i.e. addition of [^{13}C]-labeled CH_3OH and unlabeled CH_3Cl). OTU_MDH_2 and OTU_MDH_6, were

detected at low abundance (3 %), while OTU_MDH_4 had higher abundance in the microcosm to which only [^{13}C]-CH_3OH was added.

Analysis of *mch* homologs showed an overall higher diversity compared to other studied functional genes. At the cut-off value of 80 % (Table 4.2, Stacheter *et al.*, 2013), 16 OTUs were defined, more than for the other two functional genes (Tables 4.2 and S4.2). Only 4 of the 16 OTUs were defined as labeled (OTU_mch_3, OTU_mch_9, OTU_mch_10 and OTU_mch_11; Figure 4.8 and Figure S 4.4).

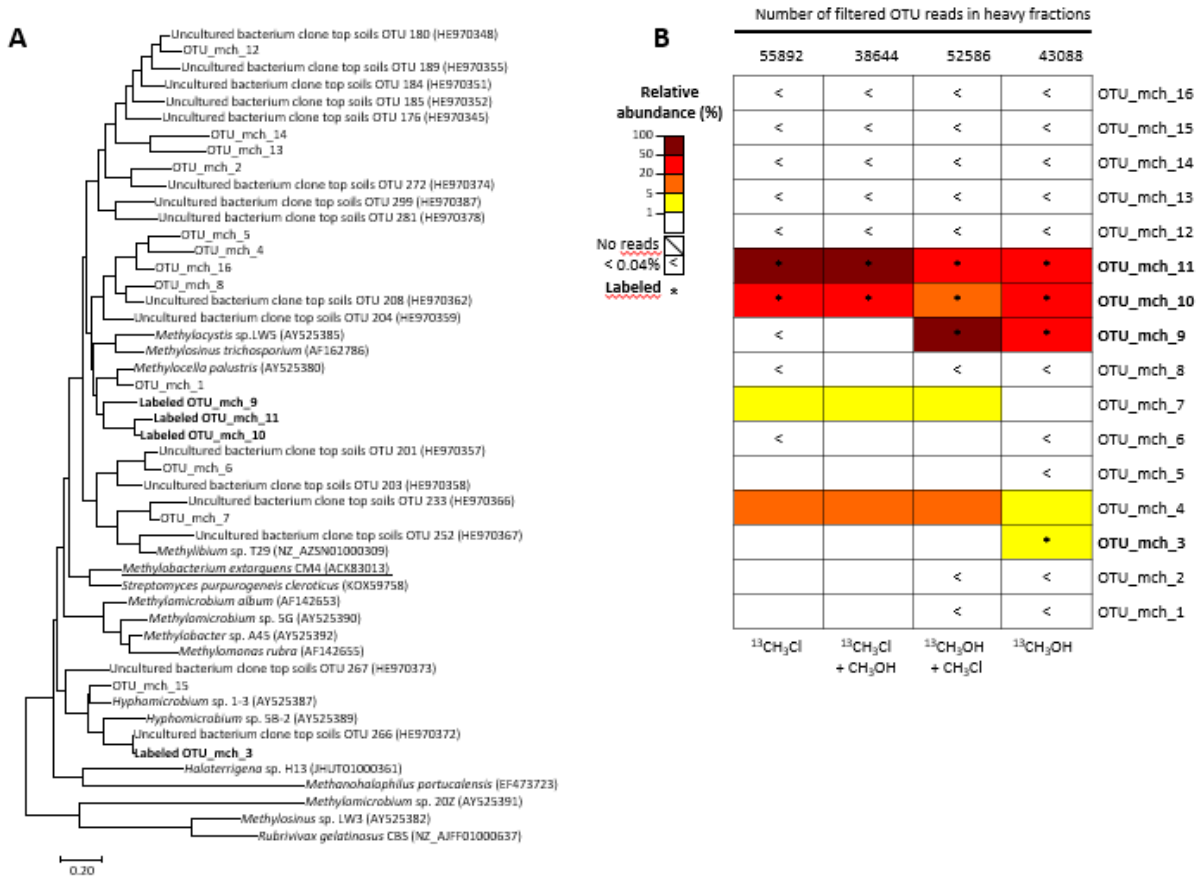


Figure 4.8. Affiliation and relative abundance of *mch* OTUs

(A) Phylogenetic analysis of methenyl cyclohydrolase (*mch*) amplicon sequences was performed with a set of 47 unique 430 nucleotide long sequences including sequences from reference strains (known CH_3Cl-degraders are underlined), sequences of isolates from a previous SIP study on CH_3OH utilization in top soils (Stacheter *et al.*, 2013), and consensus sequences for *mch* OTUs defined in heavy DNA fractions at the >80 % sequence identity level. OTUs defined as labeled according to the criteria defined in Materials and Methods are shown in bold. The tree was obtained from the alignment by the Maximum Likelihood method (Tamura Nei model, Tamura *et al.*, 1993). Branch lengths are proportional to sequence differences. Bootstrap was performed with 1,000 replicates. (B) Heatmap of the relative

abundance of recovered OTUs in heavy DNA fractions of [^{13}C]-labeled microcosms. Labeled OTUs are shown in bold.

In the [^{13}C]-CH_3Cl treatment, labeled OTU_mch_10 and OTU_mch_11 represent more than 90 % of *mch* sequences. These two OTUs had similar abundances (Table S4.2) and were also labeled in the corresponding treatment to which unlabeled CH_3OH was added together with [^{13}C]-CH_3Cl. Conversely, when CH_3OH was labeled but not CH_3Cl, genotypes OTUs_mch_10 and OTUs_mch_11 remained labeled, but their relative abundance decreased compared to that of OTU_mch_9 (Figure 4.8B). Finally, in the microcosm to which only [^{13}C]-labeled CH_3OH was added, the same 3 OTUs were labeled and also represented together over 80 % of the total *mch* sequences, although their relative abundances changed in the [^{13}C]-CH_3OH treatment. The minor frequent genotype OTU_mch_3 was only labeled in this treatment.

4. Discussion

In this study, forest soil samples were pre-incubated with unlabeled CH_3Cl to allow subsequent efficient CH_3Cl assimilation, and identify potential chloromethane degraders. Notably, the chosen experimental strategy also prevents oxidation products (labeled CO_2) to be subsequently fixed by taxa unable to directly assimilate CH_3Cl. Indeed, out of the five labeled OTUs in [^{13}C]-CH_3Cl and [^{13}C]-CH_3OH amended microcosms, four are clearly not autotrophs, although some strains of *Pseudonocardiaceae* capable of utilizing CH_3OH have been described as facultative autotrophs (Grostern and Alvarez-Cohen, 2013). Our data suggest that the experimental setup with alternating cycles of labeling and aeration prevented detectable cross feeding via [^{13}C]-CO_2 between taxa. During CH_3Cl labeling, a dilution of the labeling with unlabeled CH_3OH present in the forest soil might have occurred, although the high concentration of the C_1 compound used for amendment would in theory help minimize this process. Moreover, CO_2 production was very similar in all microcosms, as well as [^{13}C]-CO_2 production from CH_3Cl and CH_3OH (Figure 4.2). Incidentally, this clearly indicates that soil microbiomes have much higher capacity for oxidation of carbon nutrients than that required for prevailing ambient concentrations of both methanol and chloromethane. In other words, obtaining evidence for differential production of CO_2 from CH_3Cl or CH_3OH would require even higher concentrations than the concentrations used in this study. Conversely, the used

experimental setup, involving preincubation with high concentrations of chloromethane and relatively high substrate concentrations of methylotrophic substrates, may have led to overseeing active microorganisms growing with CH_3Cl and/or CH_3OH at ambient concentrations – a situation currently typical of the limitations of the SIP approach (Coyotzi *et al*., 2016).

4.1. New insights into CH_3Cl utilizers in forest soil

One might have speculated that pre-incubation with CH_3Cl might have narrowed down the diversity compared to the enrichment and selection isolation protocols used previously (Miller *et al*., 2004; Borodina *et al*., 2005; Nadalig *et al*., 2011). However, the detected taxonomic diversity in soil microcosms was quite high and generally unaffected by further pulses of CH_3OH and CH_3Cl (Table 4.4, Figure 4.2). Moreover, among the 117 different taxonomic OTUs described at the family-level based on the 16S rRNA gene, only 5 could be conservatively defined as labeled (Table 4.5). They belong to the three phyla *Proteobacteria, Actinobacteria,* and TM7. The latter is a major bacterial lineage which has been recalcitrant to cultivation and isolation until recently (He *et al*., 2015). All five differentially detected OTUs have been previously isolated or detected in soils, but until now, were not associated with CH_3Cl degradation. It was also striking to find that these family-level OTUs, with the possible exception of the poorly defined TM7 phylum OTU, essentially consisted of sequences that could be ascribed to a highly dominant genus (Table 4.5). Hence, the main detected genus within *Alphaproteobacteria* could be tentatively affiliated to the genus *Methylovirgula* largely found in acidic forest soil and decaying deadwood (Hoppe *et al*., 2015). *Streptomyces* is found predominantly in soils and decaying plant material and taxonomically diverse in forest soils (Adams *et al*., 2011; Amore *et al*., 2012). *Pseudonocardiaceae* representatives have also been isolated from soils (Warwick *et al*., 1994). *Acidothermus* comprises obligate aerobes, some with cellulolytic activity, and found in soils (Talia *et al*., 2012). Strikingly, the genus *Hyphomicrobium*, previously shown to dominate CH_3Cl-utilizing communities (Borodina *et al*., 2005) and for which strains were also readily isolated from soils (Borodina *et al.*, 2005) and the phyllosphere (Nadalig *et al.*, 2011), was not prominent and not detected as labeled, although present in the present study (Figure S4.1). *Pseudonocardiaceae* feature the closest known cultured CH_3Cl-degrader, affiliated with the genus *Norcardiodes*: strain SAC-4, a forest

top soil isolate unable to utilize CH_3OH for growth (McAnulla *et al*., 2001). Thus, the diversity of currently known cultivated CH_3Cl-degraders is likely not representative of microbial CH_3Cl utilization in soils (Figure 4.5), as suggested previously (Schäfer *et al*., 2007).

Of the five genus-level labeled phylotypes that were identified, only *Methylovirgula*, *Streptomyces*, and *Acidothermus* were labeled in microcosms that were solely amended with [^{13}C]-CH_3Cl. This clearly demonstrates that strains from these taxa have the ability to assimilate CH_3Cl as a carbon source. Moreover, representatives of the TM7 phylum were found to be labeled in microcosms supplemented with [^{13}C]-CH_3Cl in the presence of unlabeled CH_3OH, although they were detected at low relative abundance (0.3 to 1.3 %). This suggests that strains belonging to the TM7 phylum are also able to utilize CH_3Cl as a carbon source. Conversely, strains associated with the labeled family-level taxon *Pseudonocardiaceae* may rather be associated with strains incapable of carbon assimilation from CH_3Cl, since this phylotype was only labeled in the microcosm amended with [^{13}C]-CH_3OH in presence of [^{12}C]-CH_3Cl.

Hence, the relative abundance of the five-labeled phylotypes in the four microcosms amended with [^{13}C]-CH_3Cl and/or CH_3OH (Figure 4.4; Table 4.5), revealed three major metabolic profiles associated with dominant phylotypes with regard to utilization of CH_3Cl in the investigated forest soil:

- *Streptomyces* and *Acidothermus*: CH_3Cl amendments alone or combined with CH_3OH always confer high relative abundance. Organisms that use only CH_3Cl and not CH_3OH for growth, as found for the cultivable *Aminobacter* species (Coulter *et al*., 1999), label with [^{13}C]-CH_3Cl. A particularity was observed in the case of *Acidothermus.* This OTU was labeled with [^{13}C]-CH_3OH when unlabeled CH_3Cl was amended, which suggests that CH_3Cl may rather represent a source of energy for such organisms, and that CH_3OH is preferentially used as the carbon source for growth.

- *Methylovirgula*: CH_3Cl amendment lowers relative abundance. *Methylovirgula* was described to use CH_3OH produced during wood decomposition (Hope *et al*., 2014, Kielak *et al*., 2016). To date, CH_3Cl utilization in a *Methylovirgula* strain has not been described and, surprisingly, no genomes from this genus are yet publically available. Otherwise, *Methylovirgula* was defined as labeled in [^{13}C]-CH_3Cl treatments

supplemented or not with unlabeled CH_3OH, as in the [^{13}C]-CH_3OH condition. Its absence as a labeled taxon in the treatment with [^{13}C]-CH_3Cl and [^{12}C]-CH_3OH suggest that *Methylovirgula* representatives prefer CH_3OH over CH_3Cl as a carbon source. Cultivated *Alphaproteobacteria* which can utilize both CH_3OH and CH_3Cl simultaneously as carbon and energy source for growth have been described, and include well-characterized *Hyphomicrobium* and *Methylobacterium* strains (McDonald *et al*., 2001). It can be speculated that the combination of CH_3Cl and CH_3OH could decrease growth rate of representatives of *Methylovirgula* strains, as indeed was observed *in vitro* with *Methylobacterium extorquens* CM4 (Sandro Roselli PhD thesis, 2009), and explain why *Methylovirgula* was not detected as a labeled OTU in the condition in which unlabeled CH_3Cl was added together with [^{13}C]-CH_3OH.

- *Pseudonocardiaceae* and TM7 exhibit an intermediate metabolic profile compared to the other two profiles described above. No labeling was detected in treatments that were solely supplemented with one of the two C_1 compounds. Utilization of CH_3Cl as electron donor only but not as carbon source could potentially increase labeling with CH_3OH as found for *Pseudonocardia*. An extreme case of this scenario may be associated with the very high relative abundance of the *Microbacteriaceae* OTU, essentially consisting of the *Curtobacterium* genus, and only detected in amended microcosms with both [^{13}C]-CH_3OH and unlabeled CH_3Cl (Figure 4.4, Table 4.5). Considering the very high (30 %) abundance this OTU reached under this condition, however, it is surprising that no increase whatsoever was observed for this OTU in any fraction of the corresponding unlabeled microcosm (<0.6 %, data not shown). The intriguing abundance profile of this OTU in microcosms of Steigerwald forest soil under different conditions will now need to be ascertained by independent approaches such as qPCR.

In the reverse scenario, use of CH_3Cl as carbon source, and CH_3OH for energy could increase labeling with CH_3Cl, as found for the TM7 phylum. The TM7 phylum was only found labeled in treatments in which [^{13}C]-CH_3Cl was added together with unlabeled CH_3OH. Thus, methylotrophic organisms affiliated to the TM7 phylum might utilize CH_3Cl more for carbon assimilation, and CH_3OH more for energy production. The low abundance of the TM7 phylum family-level OTU also suggests that corresponding strains assimilate carbon from CH_3Cl

inefficiently or at a slow rate. A synergistic metabolism of CH_3OH and CH_3Cl utilization may also be envisaged in this context. This was reported for the marine methylotrophic bacterium HTCC2181 that used CH_3Cl as carbon source only in the presence of CH_3OH (Halsey *et al*., 2012).

As a final note, metabolic inhibition by chloromethane cannot be ruled out. Use of unlabeled CH_3Cl could have a toxic effect, potentially resulting in reduction of labeling. For instance, chloromethane was demonstrated to inhibit ammonia monooxygenase, an essential enzyme of the obligate chemolithoautotrophic aerobe, *Nitrosomonas europaea*, which uses ammonia as its sole natural energy source (Rasche *et al*., 1991; Keener and Arp, 1993). Conversely, as a fortuitous substrate of particular methane monooxygenase, chloromethane affects methylotrophic growth of *Methylomicrobium album* GB8 on CH_3OH (Han et Semrau, 2000).

4.2. *cmuA* OTUs and metabolic diversity in the environment

Eight *cmuA* OTUs were detected in this study, and five were labeled under different conditions (Figure 4.6B). The labeled *cmuA* OTUs were closely related to each other and clustered with cultured *Alphaproteobacteria* sequences of *Methylobacterium* and *Hyphomicrobium* as well as a few *cmuA* genotypes of uncultured bacteria that were previously detected in soil samples (Figure 4.6A; Miller *et al*., 2004: Borodina *et al*., 2005). OTU_cmuA_2 was detected as the most abundant OTU in all treatments. Taken together, in the studied forest soil, these data may be explained by the fact that few microorganism harbor *cmuA*-related sequences or alternatively, that the *cmuA* primers are inefficient to detect the *cmuA* diversity present in soils. Another explanation would be that *cmuA*-independent CH_3Cl utilization pathway predominate within CH_3Cl-degraders in soil (Nadalig *et al*., 2014). While OTU_cmuA_3 was labeled in the microcosm amended with [^{13}C]-CH_3Cl alone, only OTU_cmuA_2 was labeled when CH_3OH was also added. Conversely, in the microcosm where [^{13}C]-CH_3OH and unlabeled CH_3Cl were added, the diversity of *cmuA* OTUs increased. Interestingly, OTUs_cmuA_2, 3 and 4 were labeled in the microcosm treated with [^{13}C]-CH_3OH alone. Taken together, these results tentatively suggest that *cmuA* OTUs may cover the four phylotypes of CH_3Cl-degraders: (i) strains with OTU_cmuA_2 or OTU_cmuA_3 may be able to use both CH_3Cl and CH_3OH as carbon and energy source; (ii) strains with OTU_cmuA_4 may be able to use CH_3OH as carbon and energy source, and CH_3Cl as energy source; (iii) strains with OTU_cmuA_6 may assimilate carbon from CH_3OH and use CH_3Cl as energy source; (iv) low abundance strains with

OTU_cmuA_8 only found labeled in the microcosm labeled with [^{13}C]-CH_3OH and to which CH_3Cl was also added may use CH_3OH as carbon source and CH_3Cl only for energy.

4.3. Chloromethane-associated methylotrophy

Soil methylotrophs were assessed using two types of biomarker genes, *mxaF*/*xoxF* for methanol oxidation, and *mch* for tetrahydrofolate-dependent C_1-interconversion reactions. In heavy fractions of the [^{13}C]-CH_3Cl microcosms, OTUs_MDH_2, 3 and 6, with a large dominance of the OTU_MDH_3 (relative abundance of 69.7 %) were labeled (Figure 4.7). MDH diversity shifted in response to CH_3OH supplementation. Thus, CH_3Cl utilization impacts the diversity of detected methanol dehydrogenase phylotypes, with labeled MDH OTUs all affiliated to *xoxF*-sequences in this study (Figure 4.8). Similar effects of CH_3Cl were also observed in the analysis of the 4 labeled *mch* OTUs. OTU_mch_3 is most similar to the *mch* sequence of an uncultured organism from forest soil (Stacheter *et al*., 2013), whereas the other 3 labeled OTUs are clustered together but not affiliated to previously reported *mch* sequences of isolates or metagenomic datasets (Figure 4.8A). Methylotrophs harboring OTU_mch_9 may use CH_3Cl only in presence of CH_3OH, whereas CH_3OH appears to affect the relative abundance of OTU_mch_10. On the contrary, the OTU_mch_3 was detected only in the [^{13}C]-CH_3OH microcosm, suggesting that strains of this phenotype cannot use CH_3Cl. The obtained results are reflected into the phylogenetic tree, where OTU_mch_9, 10 and 11 are closely related, whereas OTU_mch_3 is relatively divergent (Figure 4.8). Whatever the basis of these observations may be, the diversity of OTUs associated with genes involved in methylotrophy metabolism, as well as their dynamics as a function of amendment with unlabeled or labeled CH_3Cl, clearly suggest that the diversity of methylotrophic communities is affected by chloromethane.

5. Conclusion

This study shows for the first time that chloromethane consumption in forest soil may be affected by the key methylotrophic substrate methanol. This study also provides the first deep coverage exploration of the bacterial diversity associated with chloromethane metabolism in soil. A main outcome of this study is the uncovering of new taxa associated with CH_3Cl utilization, including specific genera from *Alphaproteobacteria* and *Actinomycetales* for which

no degrading strains have yet been isolated or associated with CH_3Cl consumption in previous studies. Conversely, taxa corresponding to previously isolated CH_3Cl-degraders, such as *Hyphomicrobium* and *Methylobacterium*, were not detected as effective CH_3Cl-utilizers in the studied forest soil, despite the fact that *cmuA* genes closely related to those of corresponding isolated were detected. This suggests the possibility of active organisms in soils with known chloromethane utilization genes acquired by horizontal transfer. Future efforts will require the use of complementary approaches to improve knowledge of CH_3Cl-utilizers in the environment, including experiments without CH_3Cl pre-incubation and the use of concentrations of CH_3Cl closer to those prevailing in the soil environment. Moreover, further work is required to address the relevance of the detected labeled OTUs in the environment and to characterize new isolates from corresponding taxa, which may also feature as yet unknown CH_3Cl degradation pathways other than the canonical *cmu* pathway.

6. Supplemental materials

Table S4.1. Summary of sequence data obtained in the SIP experiment

Microcosm	DNA fraction[a]	16S rRNA *rrsA* gene reads			Functional gene reads						
		Raw reads[b]	Total filtered reads[c]	Filtered reads	Raw reads[b]	Total filtered reads[c]			Filtered reads		
						cmuA	*mch*	*mxaF_xoxF*	*cmuA*	*mch*	*mxaF_xoxF*
$[^{13}C]$-CH_3Cl	H	47530	306404	46110	96190	7438	130172	3356	4282	55892	786
	M	111914		107946	58556				2056	37524	1118
	L	157600		152348	54408				1100	36756	1452
$[^{13}C]$-CH_3Cl + CH_3OH	H	183860	331804	178042	74270	10320	124876	4794	3778	38644	1534
	M	107282		103360	83720				4384	48252	1692
	L	52098		50402	57828				2158	37980	1568
CH_3Cl + $[^{13}C]$-CH_3OH	H	79510	349990	76690	76850	11448	143134	3800	614	52586	918
	M	173144		167078	74756				5584	44510	1394
	L	110046		106222	72968				5250	46038	1488
$[^{13}C]$-CH_3OH	H	100600	227354	98332	72356	6144	115594	2862	1886	43088	588
	M	94430		91136	66856				3384	41322	1086
	L	39066		37886	45940				874	31184	1188
CH_3Cl	H	74708	248254	72522	57694	5438	125532	6692	370	27732	2616
	M	107742		103850	77772				1988	51800	2216
	L	74522		71882	70190				3080	46000	1860
CH_3OH + CH_3Cl	H	49080	219992	47236	75074	6524	118428	3844	922	44356	862
	M	84564		81572	46122				2230	28196	1300
	L	93876		91184	70152				3372	45876	1682
CH_3OH	H	93620	244402	90822	50286	4578	98696	5434	246	26518	2946
	M	51602		49984	63712				1872	43490	1422
	L	106682		103596	43526				2460	28688	1066
Control	H	102994	240724	100032	50886	4690	113260	6154	184	26804	2710
	M	82628		79894	64516				876	42884	1772
	L	62760		60798	65778				3630	43572	1672

[a] H, M and L: heavy, middle and light DNA fractions respectively

[b] Sequences containing the intact barcode sequence and the forward primer sequence with less than 3 mismatches

[c] Sum of filtered reads in fractions H, M and L for each microcosm

[d] For 16S rRNA sequences, filtered reads were obtained from raw reads by following the SOP in mothur (www.mothur.org/wiki/MiSeq_SOP). Amplicons of functional genes were processed as described in Materials and Methods. Briefly, amplicons differing by more than 20 nucleotides from the expected size, sequences with a quality score lower than 20, and singletons were all removed. As a final step, OTUs for which the reference sequence did not match the target gene by BLAST were also discarded.

Table S4.2. Relative abundance (%) and labeled status of functional gene OTUs in heavy DNA fractions of [^{13}C]-labeled microcosms

OTU/microcosm[a]	[^{13}C]-CH_3Cl	[^{13}C]-CH_3Cl + CH_3OH	CH_3Cl + [^{13}C]-CH_3OH	[^{13}C]-CH_3OH
OTU_cmuA_1	6.1	0.5	0.0	0.0
OTU_cmuA_2	**63.8**	**83.8**	42.7	**64.1**
OTU_cmuA_3	**24.0**	12.9	**18.2**	**26.6**
OTU_cmuA_4	4.0	1.4	**3.3**	**7.4**
OTU_cmuA_5	0.1	NR[b]	NR	NR
OTU_cmuA_6	2.1	1.4	**35.2**	1.9
OTU_cmuA_7	LR	LR	LR	LR
OTU_cmuA_8	LR	LR	**0.7**	LR
OTU_MDH_1	13.2	**28.3**	26.4	25.5
OTU_MDH_2	**3.6**	**12.4**	**5.9**	**11.6**
OTU_MDH_3	**69.7**	29.3	46.0	23.1
OTU_MDH_4	7.1	**15.4**	**12.0**	**23.5**
OTU_MDH_5	3.8	**13.7**	9.2	13.3
OTU_MDH_6	**2.5**	0.9	0.7	**3.1**
OTU_mch_1	0.1	0.1	LR	LR
OTU_mch_2	0.1	0.2	LR	LR
OTU_mch_3	0.2	0.3	0.5	**3.7**
OTU_mch_4	7.2	9.5	8.2	1.0
OTU_mch_5	0.1	0.1	0.1	LR
OTU_mch_6	LR	0.1	0.1	LR
OTU_mch_7	1.0	1.7	1.1	0.2
OTU_mch_8	LR	0.1	LR	LR
OTU_mch_9	LR	0.2	**52.5**	**27.2**
OTU_mch_10	**32.3**	**31.9**	**15.4**	**23.5**
OTU_mch_11	**59.1**	**55.9**	**22.3**	**44.3**
OTU_mch_12	LR	LR	LR	LR
OTU_mch_13	LR	LR	LR	LR
OTU_mch_14	LR	LR	LR	LR
OTU_mch_15	LR	LR	LR	LR
OTU_mch_16	LR	LR	LR	LR

[a] OTUs defined as labeled in a given microcosm are shown in bold; [b] NR, no reads; [c] LR, low read number, relative abundance less than 0.04 %.

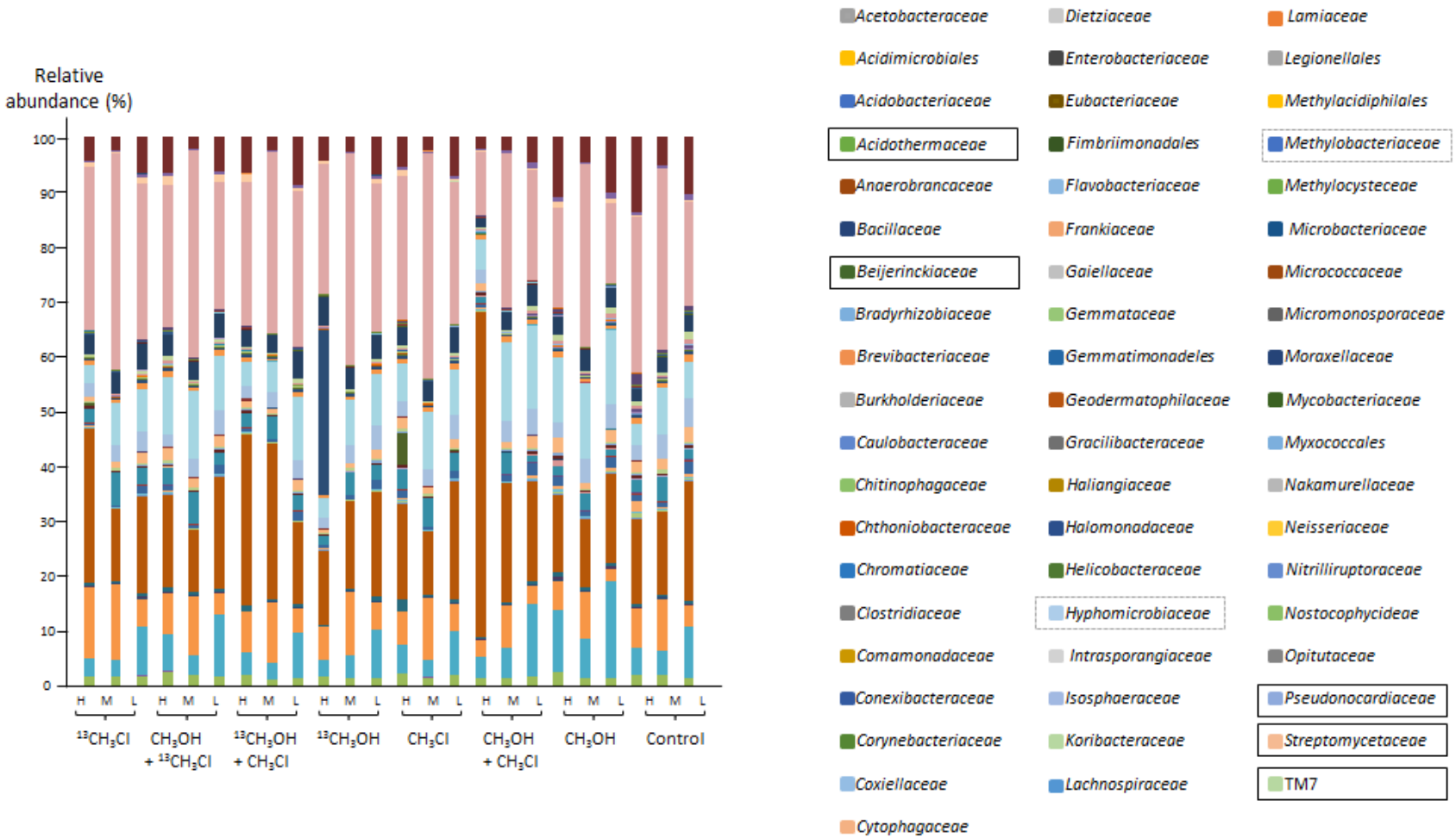


Figure S4.1. Relative abundance of family-level taxa in amplicon libraries of the 16S rRNA gene

Families (OTUs with >80 % sequence identity) which were defined as labeled by comparison of heavy DNA fractions of the labeled and corresponding unlabeled microcosms are framed.

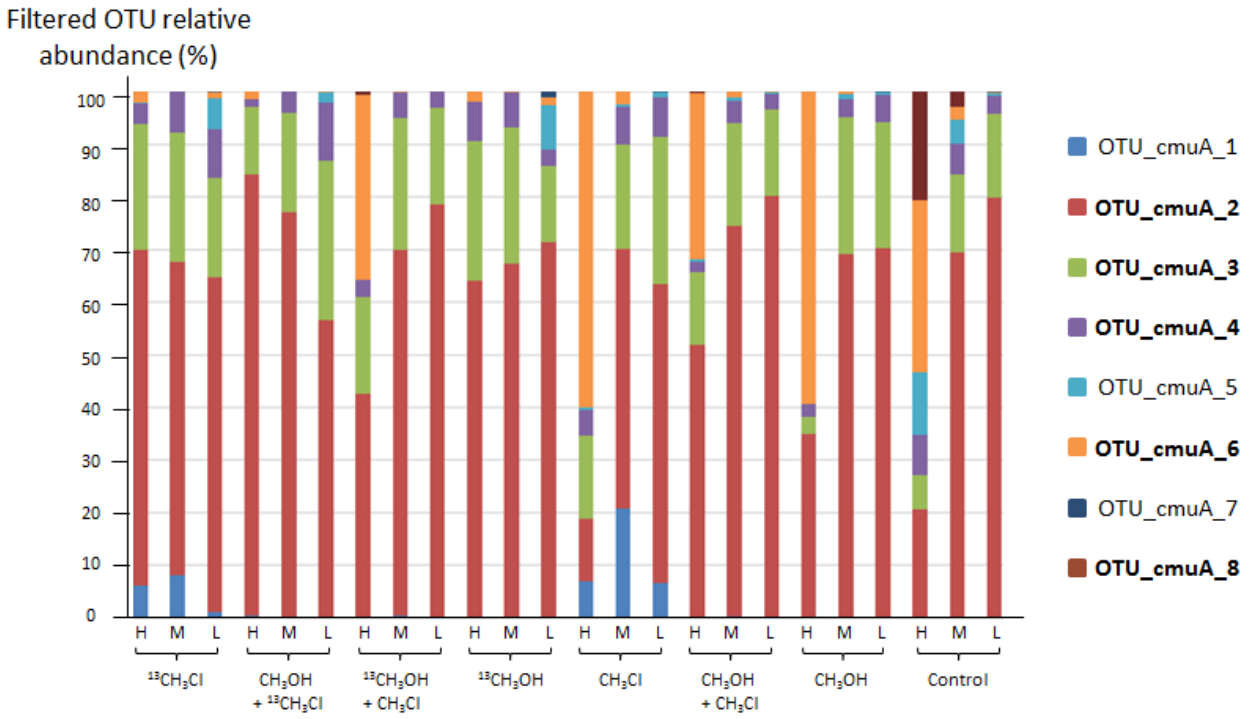


Figure S4.2. Relative abundance of OTUs in amplicon libraries of the *cmuA* gene

H, M and L represent heavy, middle and light DNA fractions in the 8 microcosms. OTUs were defined at the 90 % sequence identity level, and OTUs defined as labeled (see Materials and Methods) are shown in bold.

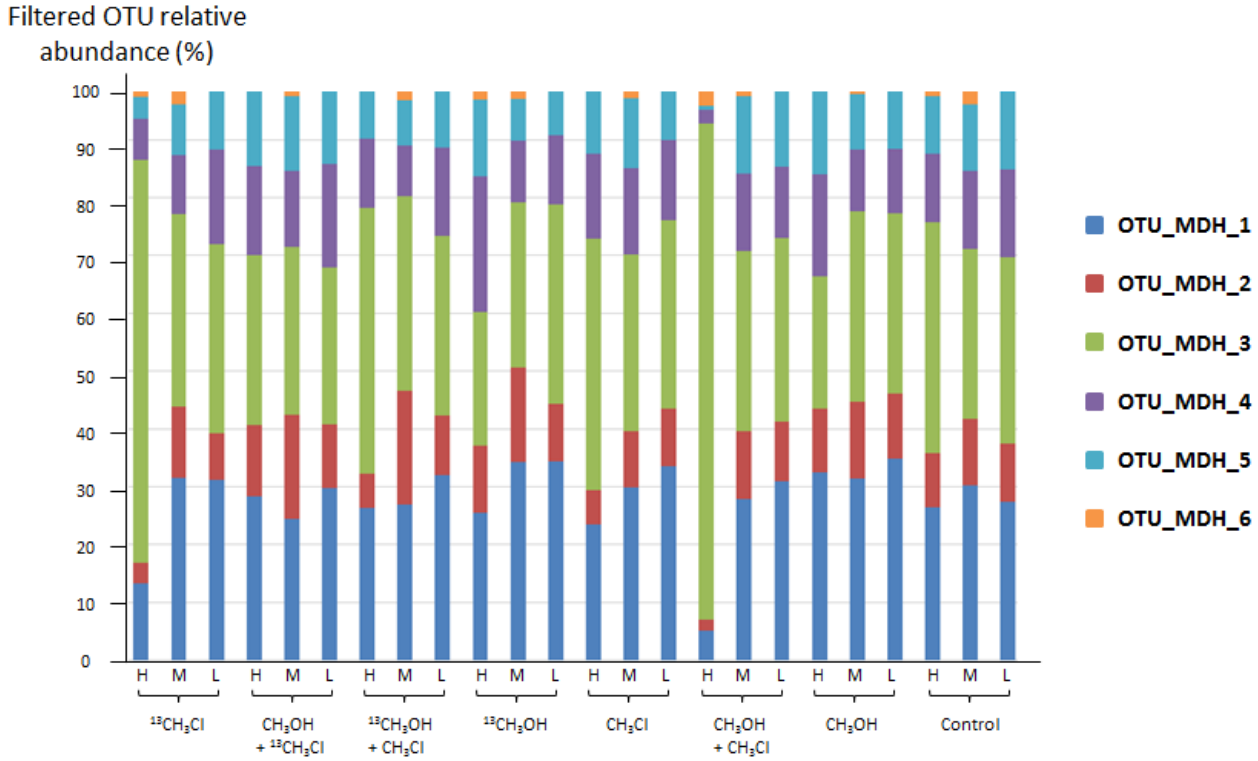


Figure S4.3. Relative abundance of OTUs in amplicon libraries of the *mxaF* and *xoxF* genes

H, M and L represent heavy, middle and light DNA fractions in the 8 microcosms. All OTUs at the 90 % sequence identity level were defined as labeled according to the defined criteria (see Materials and Methods) in microcosms incubated with [^{13}C]-labeled compounds.

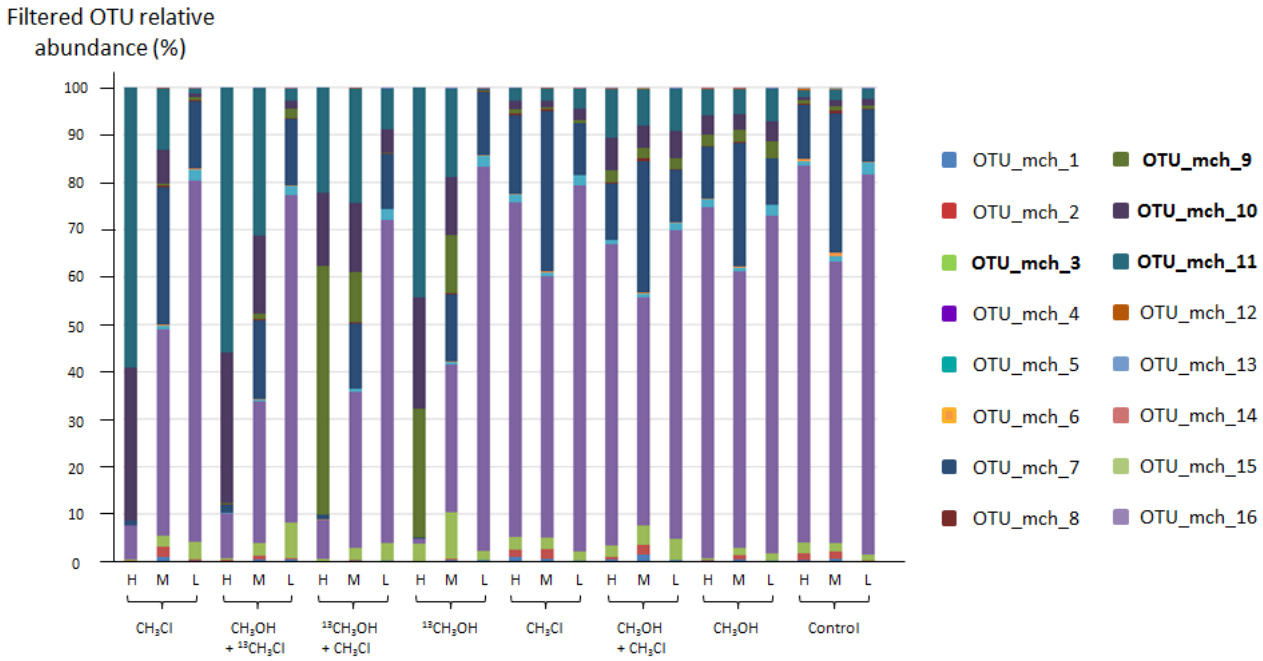


Figure S 4.4. Relative abundance of OTUs in amplicon libraries of the *mch* gene
H, M and L represent heavy, middle and light DNA fractions from the 8 microcosms. OTUs at the 80 % sequence identity level shown in bold were defined as labeled as described in Materials and Methods.

Chapitre 5

Conclusion and perspectives

Chapter 5. Conclusions and perspectives

In this final Chapter, the major conclusions of my PhD studies are presented, and discussed in the context of the future experiments that I propose (Figure 5.1).

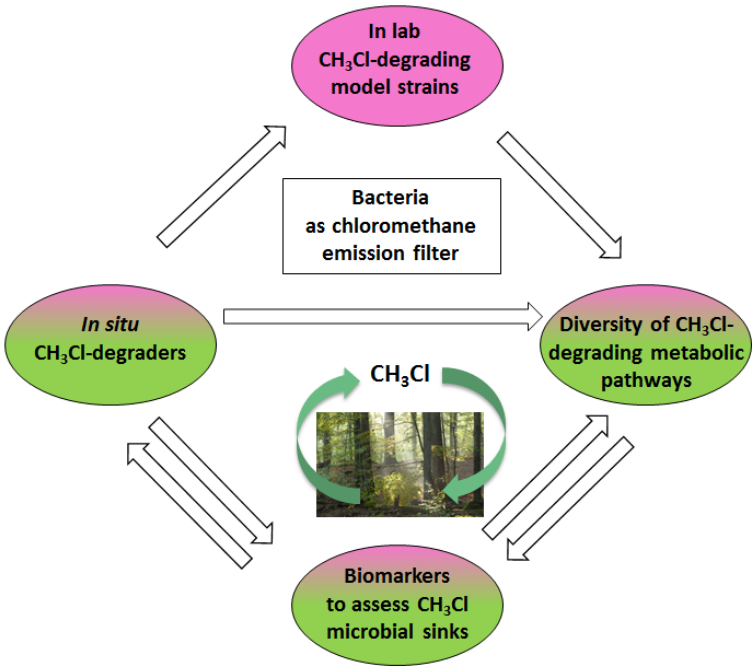


Figure 5.1. Overview of research avenues to assess the role of bacteria as chloromethane emission filters in forest soil

Four major tasks are proposed (ovals), and color-coded for experiments performed *in labo* (pink) or *in situ* (green). These tasks are based on culture-dependent and independent methods. As suggested by the arrows, these research axes are interconnected in an interactive observe/test/validate workflow.

In lab controlled experiments with cultivated model strains are needed for a better understanding of chloromethane metabolism by the different bacterial chloromethane-degrading isolates which are available. For instance, one can assess if these isolates are able to use chloromethane as the sole source of carbon and energy, as a source of carbon only, or as a source of energy only, and whether co-utilization of other carbon compounds found in forest soils occurs in the context of chloromethane degradation.

In-depth knowledge of catabolic pathways and adaptation to growth with chloromethane is needed to discover alternative biomarkers to obtain a full picture of the entire diversity of bacterial chloromethane degradation in forest soil.

Exploring without a *priori* the diversity of chloromethane-utilizers *in situ* using molecular-based techniques and discovering as yet unknown degradation pathways is also needed to uncover chloromethane-degrading ecotypes active *in situ* in their entirety.

Biomarkers are biological measurable indicators that are representative of some biological state, function or condition. For the questions at hand, biomarkers that report on microbial filtering of chloromethane in soils are needed.

1. News insights in chloromethane utilization deduced from the study of the model strain *M. extorquens* CM4

1.1. In-depth study of the bacterial adaptive response to chloromethane utilization

Analysis of the genes that were found with higher transcript abundance in *M. extorquens* CM4 cultures grown with chloromethane compared to methanol allowed the identification of key genes in the adaptive response to growth with chloromethane. Hereafter, I will focus on three genes that were not correlated with growth with chloromethane previously.

- **The *hppA* gene encoding a membrane-bound proton translocating pyrophosphatase**. The energy of pyrophosphate hydrolysis may serve as the driving force for proton extrusion to build up a proton motive, force as demonstrated in other bacteria (Lopez-Marques *et al*., 2004). The proton produced upon chloromethane dechlorination in the cytoplasm will be excreted to help maintain intracellular pH build up a proton motive force. Cells may thereby also be able to coexcrete chloride ions, as suggested by the study of *hppA* homologs demonstrated to contribute to salt resistance that were retrieved from the metagenome of brines and moderate-salinity rhizosphere (Mirete *et al*., 2015). In the dichloromethane-degrading strain of the same species, *M. extorquens* DM4, *hppA* was also found with higher transcript abundance in cultures grown with dichloromethane. Thus, the *hppA* gene appears to be involved in adaptation to dechlorination reactions during growth with chlorinated methanes chloromethane and dichloromethane.

- **The *pnt* gene cluster encoding the transhydrogenase PntA**. The Pnt complex of PntAA, PntAB and PntB subunits allows the regeneration of $NADP^+$ by transfer of reducing equivalents between NAD(H) and NADP(H) for translocation of protons across the membrane, as demonstrated in *M. extorquens* AM1 (Chou *et al*., 2015). Basing on this study, it can be speculated that the $NADP^+$ pool is a limiting factor for growth with chloromethane, since two molecules of $NADP^+$ are needed in two successive reactions of the *cmu* pathway catalyzed by MetF2 and FolD (Fig. 3.1, Chapter 3). If this hypothesis is correct, $NADP^+$ pools are predicted to be low in chloromethane-utilizing methylotrophs grown with chloromethane, as compared to other carbon substrates.

- **The gene cluster encoding the acetone carboxylase subunits AcxA, AcxB, AcxC.** There is growing evidence for a possible link between the *acxABC* gene cluster and chloromethane utilization. We observed higher expression of all three *acxABC* genes

in the transcriptome of the chloromethane-grown strain. Previous comparative genomics revealed that the *acx* cluster is part of the shared variable genome of chloromethane-degrading strains of known genome sequence *M. extorquens* CM4, *Hyphomicrobium* sp. MC1, and in part for *Roseovarius* sp. 217 (Roselli *et al*., 2013; Vuilleumier *et al*., 2011). Strain CM4 was previously demonstrated to grow with acetone as the sole source of carbon and energy (Roselli *et al*., 2013), but this has not been tested in other chloromethane-utilizers. Whether acetone, in comparison with other C_2 substrates, has a specific positive growth effect on growth with chloromethane and not with other C_1 compounds may be assessed in *M. extorquens* CM4 and other chloromethane-utilizing isolates. [^{13}C]-methanol labeling and liquid chromatography mass spectrometry analyses have been used in the past to study the different fates of the carbon from two substrates that are co-utilized in *M. extorquens* (Peyraud *et al*., 2012). A metabolomic approach using labeled [^{13}C]-acetone could be envisaged to characterize metabolic intermediates for chloromethane utilization in presence of acetone (Tang *et al*., 2012).

In the investigated conditions, the transcriptional responses of *M. extorquens* strains to growth with chloromethane or dichloromethane were mostly unrelated. However, high-throughput transcriptomic analysis only captures an averaged global snapshot of the response of bacteria in terms of gene expression at a given point in time. In my PhD project, all RNAs were extracted from cultures in mid-log growth phase in the presence of a unique source of carbon and energy. Assessing transcriptional patterns under growth conditions of chloromethane and methanol co-utilization, including high and low concentrations of growth substrates, may uncover genes not previously associated with growth with chloromethane, but relevant under environmental conditions.

1.2. Search of chloromethane-utilization-dependent regulatory elements

A bacterial cell needs to adapt to varying environments and nutrient availability, by quick, sensitive and effective responses (Francez-Charlot *et al*., 2015; Balleza *et al*., 2009). This may involve regulation processes at the transcriptional and translational levels, but also modulation of the metabolite pool (Figure 5.2). Relationships between gene and transcript abundance and corresponding protein-encoded processes are commonly assumed, yet rarely

observed (Rocca *et al*., 2015). Basic research on regulatory elements and mechanisms that control chloromethane-utilization is still needed to establish predictive links between genomic, transcriptomic and proteomic data in ecosystem processes.

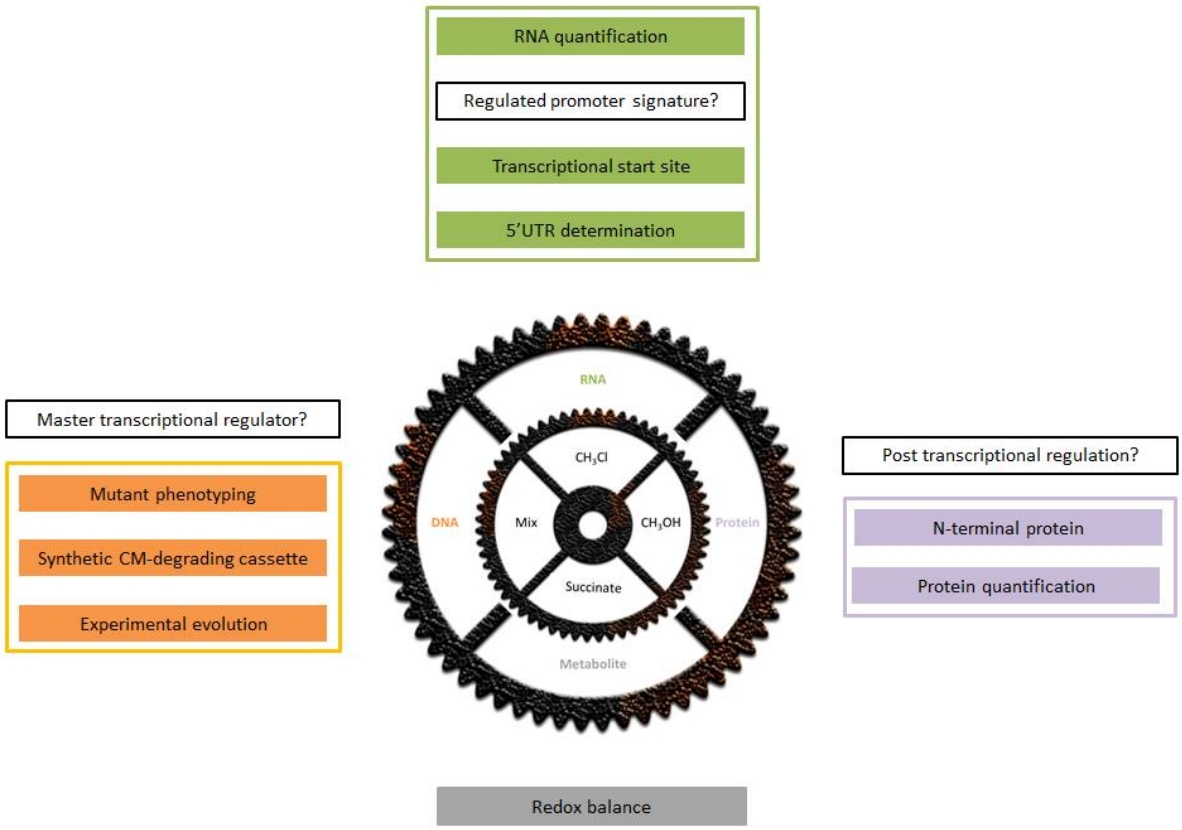


Figure 5.2. Central questions for future in lab experiments with model strains
Up to this day, the master regulator involved in transcriptional induction of the *cmu* gene remains unknown. Neither have the chloromethane-dependent genetic elements (promoter and associated regulating factors) been characterized. Different approaches are proposed to better understand the adaptive response to chloromethane utilization, at the DNA level (orange), transcriptional level (green), proteomic level (purple), and combined with a metabolite analysis (grey). For instance, dosage of $NADP^+$ pools may address the question whether $NADP^+$ pool is a limiting factor for growth with chloromethane. All the proposed approaches, including the comparison of data obtained in cultures grow with chloromethane alone, or provided with other C_1 or multi-carbon substrates, are needed to discover specific regulatory elements of chloromethane utilization.

The protein encoded by *fmdB* nearby *cmuA* was proposed as a putative regulator of chloromethane utilization in *Hyphomicrobium chloromethanicum* strain CM2, based on its very low amino acid identity (24 %) with the *Methylophilus methylotrophus* FmdB regulator protein (Wyborn *et al*., 1996; McAnulla *et al*., 2001). To our knowledge, no FmdB-like protein has been experimentally investigated. The corresponding gene in *M. extorquens* CM4 is also located adjacent to *cmuA*, and had increased transcript abundance in cultures grown with

chloromethane compared to methanol (Chapter 3). Among possible experiments that would help in finding the regulator of *cmu* genes, a targeted knockout of putative regulators would be a straightforward experiment to test their role in chloromethane-dependent gene regulation and growth. Unfortunately, as recently discovered, reverse genetics by allelic gene replacement does not work with *M. extorquens* CM4, contrary to other tested *Methylobacterium* strains (Marx *et al*., 2008), and unpublished data from both Strasbourg and Idaho laboratories. Also, knockout mutants have not been obtained so far in strain MC1, which belongs to the *Hyphomicrobium* genus, known to be recalcitrant to genetic analysis. Thus, another strategy to characterize the master regulator of the *cmu* pathway is needed. Developing a CRISPR/Cas9 system (Jakočiūnas *et al*., 2016) for *Methylobacterium* would be of great interest, but this still remains to be achieved. Alternatively, an experimental evolution approach of chloromethane utilization could also be envisaged. A previous experimental evolution was successfully developed for dichloromethane utilization. After transfer of a synthetic *dcmA*-based gene cassette for dichloromethane utilization, *M. extorquens* strains initially unable to degrade dichloromethane acquired the ability to grow on dichloromethane as the sole carbon and energy source (Michener *et al*., 2014). Accordingly, a synthetic gene cassette for chloromethane-degradation containing a minimal set of *cmu* pathway-encoding genes was constructed (Michener, Vuilleumier, Bringel, Marx, unpublished data). This *cmu* gene cassette was introduced by transformation into methylotrophic strains unable to grow with chloromethane, and the ability of the obtained transconjugants to grow with chloromethane was tested. This synthetic chloromethane-utilization gene cassette may be used to study the effect of genes of interest, including the *fmdB* gene, and by doing so, to assess their role in chloromethane utilization (Figure 5.2).

Post-transcriptional regulation has recently been increasingly demonstrated to be involved in modulation of gene expression in changing environments. It involves regulation elements which are for instance implied in translation of RNA into protein or RNA stability (Van Assche *et al*., 2015). The majority of known mechanisms of regulation of translation target the translation initiation process (Duval *et al*., 2015). Identification of 5' untranslated regions (UTRs) of mRNA would be a prerequisite to explore possible post-transcriptional regulation mechanisms. To find and annotate 5'UTRs, transcriptional start sites (TSS) and N-termini of proteins (to determine the translational start) need to be experimentally determined (as outlined in Figure 5.3).

The gene promoter region is composed of the TSS numbered +1 by convention, with two upstream sets of hexanucleotides-long consensus sequences elements approximately located at -10 and -35 nucleotides with an optimal spacing of 17 bp. The identification of TSS is now possible using high throughput sequencing in an approach called differential RNAseq (dRNA-seq), with selective sequencing of primary transcripts (Sharma *et al*., 2010) (Figure 5.3A). In this PhD work, a dRNAseq approach has been initiated for *M. extorquens* CM4 cultures grown with chloromethane compared to methanol, as well as for *M. extorquens* DM4 grown with DCM compared to methanol. RNAs from *M. extorquens* CM4 and DM4 cultures grown with methanol or chlorinated methanes were extracted (a total of eight biological independent samples corresponding to 4 different conditions). Then, the DNase-treated RNA extracts were sent to Vertis Biotechnologie AG (Freising, Germany) for cDNA library construction and sequencing. The dRNAseq data of eight TSS banks will be mapped soon to the corresponding genomes in collaboration with S. Cruveiller (Génoscope, Evry, France). Comparative TSS mapping in closely-related strains is expected to provide a large overview of what is conserved, what is strain specific (Dugar *et al*., 2013), and what is specific to growth with chloromethane.

The identification of the N-terminal position of proteins may be performed by the “dN-TOP” approach (Figure 5.3B). The N-terminome protein analysis by the “dN-TOP” approach has already been validated in *M. extorquens* DM4 grown with dichloromethane and with methanol in collaboration with C. Carapito (Laboratoire de Spectrométrie de Masse BioOrganique or LSMBO, Strasbourg, France).

The combined RNAseq and proteomic approaches will improve genome annotation and our understanding of genome expression of chloromethane-degrading reference strains grown in controlled lab conditions of C_1 and chloromethane utilization.

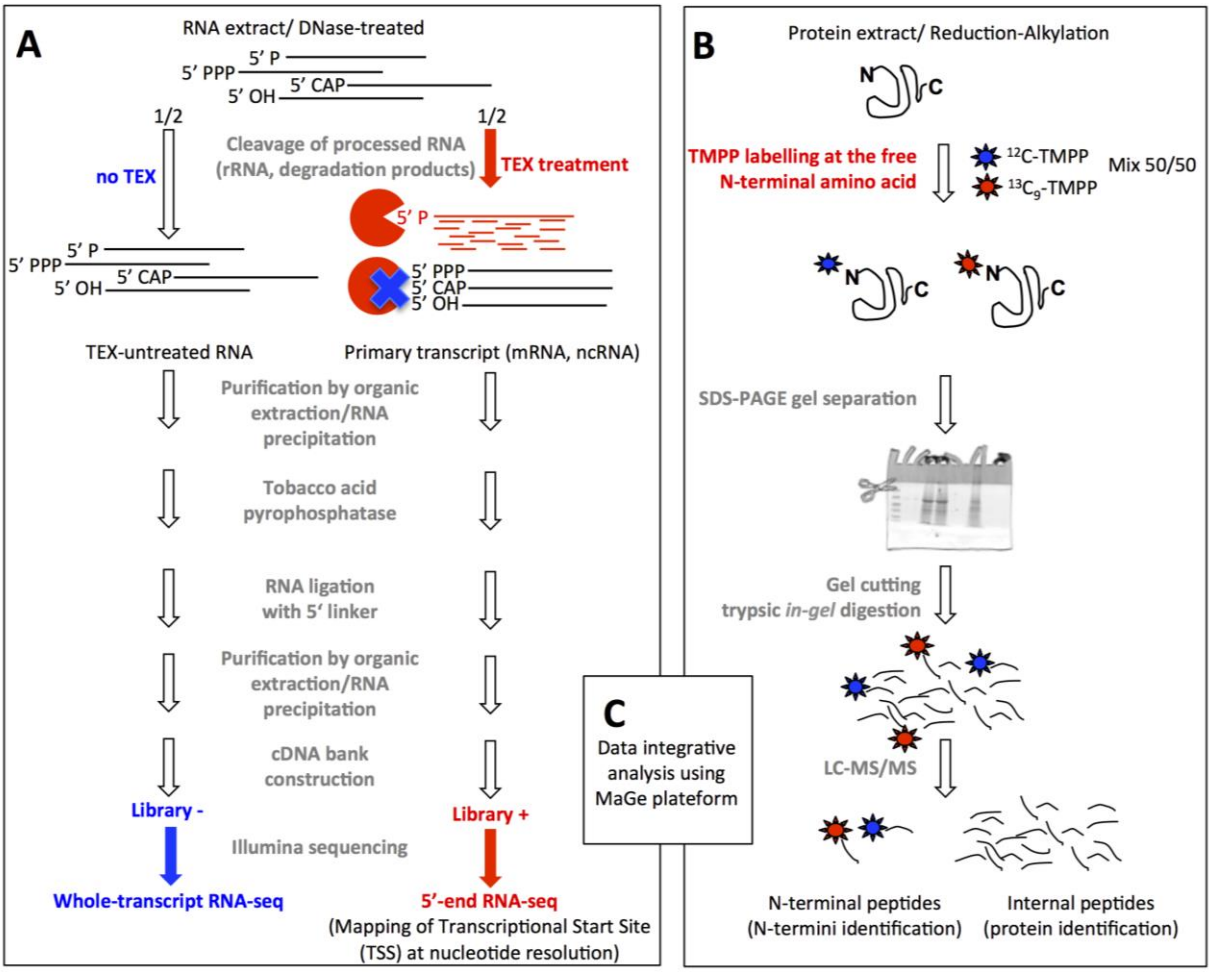


Figure 5.3. Exploration of chloromethane-dependent regulation mechanisms in model strains

A) Transcription Start Site (TSS) identification using dRNAseq (Sharma and Vogel 2014). Sequencing yields enrichment patterns of the cDNA coverage at the 5' end of primary transcripts which can be used to annotate TSS. The method is based on the fact that the bacterial cellular RNA pools consists of processed or cleaved transcripts characterized by a 5'monophosphate end (5'P), which is digested by the 5'P-dependent terminator exonuclease (TEX), while primary transcripts which have 5' triphosphate (5'PPP) or, less frequently, a 5' hydroxyl (5'OH) group are resistant to TEX (Evguenieva-Hackenberg and Klug 2011). Two libraries are prepared in parallel. Library – (in blue) captures all RNAs. Library + (in red) has been enriched for primary transcripts (Sharma and Vogel, 2014);

B) N-terminome protein analysis by the "dN-TOP" approach. The trimethoxyphenyl phosphonium (TMPP) labeling approach is an efficient N-terminomic approach that allows the characterization of both N-terminal and internal peptides in a single experiment (Bertaccini *et al*., 2013);

C) Integrative data analysis will, among other discoveries, facilitate in the annotation of 5'UTR. The difference of the +1 of the transcription compared to N-terminal position gives access to the 5' UTR of mRNAs, involved in post transcriptional regulation (Duval *et al*., 2015; Van Assche *et al*., 2015).

2. Identification of chloromethane-utilizing ecotypes in forest soil microcosms

The DNA-SIP approach gives access to active microorganisms independently of their abundance in soil and their cultivability (Coyotzi *et al*., 2016). During my PhD thesis, I used [^{13}C]-labeled chloromethane, alone or in combination with [^{13}C]-labeled methanol to assess the role of co-utilization of different key substrates of methylotrophic metabolism (Chapter 4), and thereby identified at least three new phylotypes associated with utilization of chloromethane as a carbon and/or energy source in forest soil. In addition to DNA-SIP, RNAs have been extracted from the labeled microcosms and reverse transcribed into cDNA. RNA-SIP is more sensitive than DNA-SIP, does not require cell division to labeled RNAs, and has already been used for microbial soil studies (Leng *et al*, 2016). Unfortunately, with the cDNA used as template, no PCR amplification was observed with primer sets targeting *cmuA*, *mch* and *mxaF*/*xoxF*, and I did not have the chance to attempt these experiments again.

For efficient [^{13}C]-labeling of chloromethane-utilizers in forest soil, a pre-incubation step prior to [^{13}C]-labeling was performed. In addition, high chloromethane concentrations were used. Within this experimental setup, underestimation of the diversity of chloromethane-utilizers is a possibility. A new SIP labeling strategy has been proposed recently in the literature using $H_2{}^{18}O$ (Rettedal *et al*., 2015). $H_2{}^{18}O$ is perhaps the only universal "substrate" targeting all active microorganisms in a given environment. A dual labeling with low [^{13}C]-CH_3Cl and $H_2{}^{18}O$ would in principle be more effective to assess among the active microorganisms the diversity of chloromethane utilizers at the prevailing low ambient forest soil chloromethane concentrations. Use of high levels of labeled [^{13}C]-CH_3Cl concentrations nevertheless allowed the identification of new chloromethane-associated phylotypes among families that have not been described so far to feature representatives capable of degrading chloromethane (Chapter 4). To validate and characterize if and how these phylotypes utilize chloromethane, additional experiments need to be performed. Cultural approaches may give access to these phylotypes for detailed characterization of their metabolism in the laboratory. For instance, physiological characteristics such as generation times upon growth with chloromethane and co-utilization abilities with other carbon sources could easily be assessed. Dehalogenation kinetics can be estimated by dosage of chloride production in resting-cells as previously performed in the laboratory for different chloromethane-degrading isolates of the phyllosphere (Nadalig *et al*., 2011; S. Roselli PhD thesis, 2009). Additional molecular tools such as PCR amplification or Southern hybridization with a *cmuA* probe may provide information

about *cmuA* occurrence and *cmu* gene organization. Evidence for new pathways of chloromethane utilization can be suggested by chloromethane isotope fractionation signatures (H and C), as demonstrated for *Leisingera methylohalidovorans* MB2 (Nadalig *et al*., 2014). Genome sequencing may also help predict methylotrophic central metabolism and inform on the presence of the *cmu* pathway. If genetic tools for the new isolates are available or can be developed (gene manipulation, cloning, site-directed mutagenesis, transformation, conjugation), they will greatly facilitate the identification and characterization of new genes involved in growth with chloromethane, as indeed was the case for characterization of the cmu pathway in strain CM4 (Vannelli *et al.*, 1998, 1999). However, many strains of interest may not be cultivable under laboratory conditions and are not amenable to genetic investigations. Thus, global culture-independent approaches based on direct content studies of proteins (metaproteomics), RNAs (metatranscriptomics), and DNA (metagenomics) are increasingly successful in many environmental studies (Blagodatskaya *et al*., 2013). In addition and in the case of a soil environment such as the forest soil investigated here, Fluorescence Hybridization In Situ (FISH) has also been used successfully to quantify taxa abundance (Dumont and Murrell 2005). To target active microorganisms in soil, a combination of FISH and microautoradiography (MAR) called FISH-MAR has been developed (Hirsch *et al*., 2010). This technique involves a short incubation time (to avoid cross feeding) of the environmental sample with radioactively labeled substrate. Then, if a bacterium is detected with both the fluorescent and radioactive probes under the microscope, this implies that it consumes the labeled substrate. Notably, the FISH-MAR technique is more sensitive than DNA-SIP, since the radiotracer is incorporated into all macromolecules (nucleic acids, proteins, lipids) and not only into nucleic acids (Wagner *et al*., 2006). However, soil is a complex and heterogeneous system, with autofluorescent properties and large amounts of particles of different sizes, so that its direct imaging is challenging (Vos *et al*., 2013). In addition, the use of radiotracers is expensive, requires specialized training, registered facilities, dedicated instruments, and generates potential hazardous wastes. An alternative way to avoid radioelement-based experiments is to combine FISH with flow cytometry for single-cell isolation and study of individual cells (Kalyuzhnaya *et al*., 2006). Nevertheless, for single-cell approaches to work with soil samples, soil particles need to be minimized to a minimal amount for efficient image analysis. Coupling phylogenetic identity (with FISH) and metabolic function (using stable isotope labeling incubation experiment) of single cells is possible for complex bacterial communities from the environment. In particular, those labeling technics coupled with a high-

resolution secondary ion mass spectrometry (NanoSims) has been recently developped to investigate soils. This allows to link a high resolution microscopy with isotopy, providing information on molecular and isotopic compositions of the investigated microorganisms in their native environment as already described in marine environments (Polerecky *et al*., 2012). However, NanoSIMS has only recently been applied to studies of bacterial activity in soil ecosystems, when it has been possible to efficiently dissociate bacterial cells from soil particles (Eichorst *et al*., 2015). Ideally, single-cell studies of incubations with [^{13}C]-CH_3Cl followed by NanoSIMS mass spectrometry analysis could be used to investigate the role of bacteria in chloromethane consumption in forest soil with unprecedented detail.

3. Assessing the diversity of chloromethane-utilizing pathway in soil

Besides the well-characterized oxic degradation of chloromethane mediated by the *cmu* pathway (Vannelli *et al*., 1999), at least one other pathway for chloromethane assimilation under oxic conditions exists. No *cmuA*-like homologs was detected in the genome of *Leisingera methylohalidovorans*, a strain affiliated to *Alphaproteobacteria* and isolated from a marine environment, and able to grow with chloromethane (Schaefer 2002; Nadalig *et al*., 2014). How *L. methylohalidovorans* and possibly related labeled chloromethane-utilizing ecotypes in forest soil grow with chloromethane, the genes involved and their connection to methylotrophic metabolism, remains to be assessed. More generally, a quantitative PCR (qPCR) analysis targeting 16S rRNA-encoding gene and *cmuA* gene to quantify in native soil samples and [^{13}C]-CH_3Cl-incubated soil, the ratio between *cmuA* copy number as a proxy for *cmuA*-dependent chloromethane utilization potential and 16S rRNA copy number as a proxy for bacterial abundance of labeled ecotypes, remains to be performed. The absence of correlation for a given ecotype active in chloromethane-degradation in forest soil could then be taken to suggest a *cmuA*-independent pathway for chloromethane utilization. Methods based on DNA amplification to assess the microbial diversity in forest soil are quite challenging, since humic acids bind DNA and interfere with Taq polymerase in PCR amplification reactions, and should be combined with “omics” approaches (Coyotzi *et al*., 2016). Thus, funds permitting, high throughout sequencing of metagenomes extracted from microbial communities of native forest soil or [^{13}C]-heavy fractions of SIP samples, should represent an approach of choice to allow the direct capture and analysis of soil microbial

communities (Lee *et al*., 2004). This global approach would enable the inventory of chloromethane utilization genes diversity in forest soil, and give access to *cmuA* gene in its entirety (including canonical and more divergent sequences), as well as the search of alternative chloromethane-degrading pathway-encoding signature genes.

However, the abundance of a gene within a metagenome does not suffice to assess its expression (RNA-level information) and the associated enzymatic activity (protein-level information). A direct quantification of CmuA protein may constitute more convincing evidence for occurrence of *in situ* chloromethane dehalogenation in forest soil. Mass spectrometry is indeed a very sensitive technique (Arsène-Ploetze *et al*., 2015), which has already been applied in investigations of forest soil (Mancuso *et al*., 2015). Molecular culture-independent methods such as this one may provide novel information about relevant gene products and proteins (Bastida *et al*., 2009). In addition, the bacterial fluorescent bioreporter developed by the team (Figure 1.7) (Farhan Ul Haque *et al*., 2013) may help to identify hotspots of chloromethane production in forest soil, and increase the chance to identify chloromethane degraders.

4. Need for new functional biomarkers for evaluation of the microbial chloromethane sinks

Several functional genes have been used as biomarkers of methylotrophic metabolism in environmental studies, including in forest soil. For example, genes *pmoA* and *mmoX*, encoding membrane-associated and soluble methane monooxygenases, respectively, were used as biomarkers to establish the diversity of methanotrophs involved in trichloroethylene degradation in soil (Shukla *et al*., 2009).

With regard to chloromethane utilization, the *cmuA* gene is still the best functional gene marker available. Nevertheless, there is growing evidence of *cmuA*-independent chloromethane utilization (Nadalig *et al*., 2014), confirming that the gene *cmuA* cannot adequately cover the entire diversity of chloromethane degraders. In some environments in which other methyl halides are present, *cmuA* may be inadequate as a specific biomarker for chloromethane. For instance, *cmuA* will also be associated to bacterial bromomethane utilization in marine environments (Cox *et al*., 2012). Thus, a larger toolbox of biomarkers of microbial chloromethane degradation indication is required to investigate the diversity and distribution of chloromethane-utilizing microorganisms in the environment. In Chapter 4,

potential new chloromethane-utilizing taxa were identified. Culture-based, or culture-independent characterization of genes involved in chloromethane utilization by such ecotypes may provide new biomarkers which might be more representative of the diversity of chloromethane utilizers in forest soil and complementary to the "historical" *cmuA* gene marker.

Defining novel biomarkers for the soil environment appears especially worthwhile considering the now strong evidence for bacteria-driven chloromethane degradation in soil (Borodina *et al*., 2005; Miller *et al*., 2004; and Chapter 4), and its implications for the role of bacteria in mitigating chloromethane emissions from soil to the atmosphere. The diversity and abundance of active chloromethane-utilizing ecotypes remains to be further investigated with in-depth genetic investigations. Finally, quantitative evaluation of the microbial contribution of terrestrial chloromethane sinks may be facilitated by chromatography/mass spectrometry/isotope ratio mass spectrometry (GC-MS-IRMS) and measurements of $\delta^{13}C$ and $\delta^{2}H$ values of chloromethane (Keppler *et al*., 2005; Greule *et al*., 2012, Nadalig *et al.*, 2014), as in the ongoing ANR-DFG Chlorofilter project associating the Keppler group in Heidelberg with both partner laboratories of my PhD project.

Bibliography

Adams, A.S., Jordan, M.S., Adams, S.M., Suen, G., Goodwin, L.A., Davenport, K.W., Currie, C.R., and Raffa, K.F. (2011). Cellulose-degrading bacteria associated with the invasive woodwasp Sirex noctilio. ISME J. *5*, 1323–1331.

Amore, A., Pepe, O., Ventorino, V., Birolo, L., Giangrande, C., and Faraco, V. (2012). Cloning and recombinant expression of a cellulase from the cellulolytic strain *Streptomyces* sp. G12 isolated from compost. Microb. Cell Factories *11*, 1.

Anke, H., and Weber, R.W.S. (2006). White-rots, chlorine and the environment – a tale of many twists. Mycologist *20*, 83–89.

Antony, C.P., Doronina, N.V., Boden, R., Trotsenko, Y.A., Shouche, Y.S., and Murrell, J.C. (2012). *Methylophaga lonarensis* sp. nov., a moderately haloalkaliphilic methylotroph isolated from the soda lake sediments of a meteorite impact crater. Int. J. Syst. Evol. Microbiol. *62*, 1613–1618.

Arsène-Ploetze, F., Bertin, P.N., and Carapito, C. (2015). Proteomic tools to decipher microbial community structure and functioning. Environ. Sci. Pollut. Res. *22*, 13599–13612.

Bai, Y., Müller, D.B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., Dombrowski, N., Münch, P.C., Spaepen, S., Remus-Emsermann, M., *et al.* (2015). Functional overlap of the *Arabidopsis* leaf and root microbiota. Nature *528*, 364–369.

Balleza, E., López-Bojorquez, L.N., Martínez-Antonio, A., Resendis-Antonio, O., Lozada-Chávez, I., Balderas-Martínez, Y.I., Encarnación, S., and Collado-Vides, J. (2009). Regulation by transcription factors in bacteria: beyond description. FEMS Microbiol. Rev. *33*, 133–151.

Bastida, F., Moreno, J.L., Nicolás, C., Hernández, T., and García, C. (2009). Soil metaproteomics: a review of an emerging environmental science. Significance, methodology and perspectives. Eur. J. Soil Sci. *60*, 845–859.

Bastviken, D., Thomsen, F., Svensson, T., Karlsson, S., Sandén, P., Shaw, G., Matucha, M., and Öberg, G. (2007). Chloride retention in forest soil by microbial uptake and by natural chlorination of organic matter. Geochim. Cosmochim. Acta *71*, 3182–3192.

Beck, D.A.C., McTaggart, T.L., Setboonsarng, U., Vorobev, A., Goodwin, L., Shapiro, N., Woyke, T., Kalyuzhnaya, M.G., Lidstrom, M.E., and Chistoserdova, L. (2015). Multiphyletic origins of methylotrophy in *Alphaproteobacteria*, exemplified by comparative genomics of Lake Washington isolates: Genomics of alphaproteobacterial methylotrophs. Environ. Microbiol. *17*, 547–554.

Bertaccini, D., Vaca, S., Carapito, C., Arsène-Ploetze, F., Van Dorsselaer, A., and Schaeffer-Reiss, C. (2013). An improved stable isotope N-terminal labeling approach with light/heavy TMPP to automate proteogenomics data validation: dN-TOP. J. Proteome Res. *12*, 3063–3070.

Blagodatskaya, E., and Kuzyakov, Y. (2013). Active microorganisms in soil: Critical review of estimation criteria and approaches. Soil Biol. Biochem. *67*, 192–211.

Blom, D., Fabbri, C., Connor, E.C., Schiestl, F.P., Klauser, D.R., Boller, T., Eberl, L., and Weisskopf, L. (2011). Production of plant growth modulating volatiles is widespread among rhizosphere bacteria and strongly depends on culture conditions: Volatile-mediated impact of bacteria on *Arabidopsis thaliana*. Environ. Microbiol. *13*, 3047–3058.

Boden, R., Thomas, E., Savani, P., Kelly, D.P., and Wood, A.P. (2008). Novel methylotrophic bacteria isolated from the River Thames (London, UK). Environ. Microbiol. *10*, 3225–3236.

Bonfante, P., and Anca, I.A. (2009). Plants, mycorrhizal fungi, and bacteria: A network of interactions. Annu. Rev. Microbiol. *63*, 363–383.

Borodina, E., Cox, M.J., McDonald, I.R., and Murrell, J.C. (2005). Use of DNA-stable isotope probing and functional gene probes to investigate the diversity of methyl chloride-utilizing bacteria in soil. Environ. Microbiol. *7*, 1318–1328.

Bosch, G., Skovran, E., Xia, Q., Wang, T., Taub, F., Miller, J.A., Lidstrom, M.E., and Hackett, M. (2008). Comprehensive proteomics of *Methylobacterium extorquens* AM1 metabolism under single carbon and nonmethylotrophic conditions. Proteomics *8*, 3494–3505.

Brantl, S. (2007). Regulatory mechanisms employed by cis-encoded antisense RNAs. Curr. Opin. Microbiol. *10*, 102–109.

Bringel, F., and Couée, I. (2015). Pivotal roles of phyllosphere microorganisms at the interface between plant functioning and atmospheric trace gas dynamics. Front. Microbiol. *6*.

Brophy, J.A., and Voigt, C.A. (2016). Antisense transcription as a tool to tune gene expression. Mol. Syst. Biol. *12*, 854–854.

Brunner, W., Staub, D., and Leisinger, T. (1980). Bacterial degradation of dichloromethane. Appl. Environ. Microbiol. *40*, 950–958.

Campbell, M.M., and Sederoff, R.R. (1996). Variation in lignin content and composition (mechanisms of control and implications for the genetic improvement of plants). Plant Physiol. *110*, 3.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., *et al*. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. *6*, 1621–1624.

Chang, H., Replogle, J.M., Vather, N., Tsao-Wu, M., Mistry, R., and Liu, J.M. (2015). A *cis* -regulatory antisense RNA represses translation in *Vibrio cholerae* through extensive complementarity and proximity to the target locus. RNA Biol. *12*, 136–148.

Chen, Y., and Murrell, J.C. (2010). When metagenomics meets stable-isotope probing: progress and perspectives. Trends Microbiol. *18*, 157–163.

Chistoserdova, L., Vorholt, J.A., Thauer, R.K., and Lidstrom, M.E. (1998). C_1 transfer enzymes and coenzymes linking methylotrophic bacteria and methanogenic *Archaea*. Science *281*, 99–102.

Chistoserdova, L., Chen, S.-W., Lapidus, A., and Lidstrom, M.E. (2003). Methylotrophy in *Methylobacterium extorquens* AM1 from a genomic point of view. J. Bacteriol. *185*, 2980–2987.

Chistoserdova, L., Laukel, M., Portais, J.-C., Vorholt, J.A., and Lidstrom, M.E. (2004). Multiple formate dehydrogenase enzymes in the facultative methylotroph *Methylobacterium extorquens* AM1 are dispensable for growth on methanol. J. Bacteriol. *186*, 22–28.

Chistoserdova, L., Crowther, G.J., Vorholt, J.A., Skovran, E., Portais, J.C., and Lidstrom, M.E. (2007). Identification of a fourth formate dehydrogenase in *Methylobacterium extorquens* AM1 and confirmation of the essential role of formate oxidation in methylotrophy. J. Bacteriol. *189*, 9076–9081.

Chistoserdova, L., Kalyuzhnaya, M.G., and Lidstrom, M.E. (2009). The expanding world of methylotrophic metabolism. Annu. Rev. Microbiol. *63*, 477–499.

Chistoserdova, L. (2011). Modularity of methylotrophy, revisited. Environ. Microbiol. *13*, 2603–2622.

Chistoserdova, L., and Lidstrom, M. (2013). Aerobic methylotrophic prokaryotes. In The Prokaryotes, pp. 267–285.

Chou, H.H., Marx, C.J., and Sauer, U. (2015). Transhydrogenase promotes the robustness and evolvability of *Escherichia coli* deficient in NADPH production. PLoS Genet. *11*, e1005007.

Chu, F., and Lidstrom, M.E. (2016). XoxF acts as the predominant methanol dehydrogenase in the Type I methanotroph *Methylomicrobium buryatense*. J. Bacteriol. *198*, 1317–1325.

Clerbaux, C., Cunnold, D.M., Anderson, J., Engel, A., Fraser, P.J., Mahieu, E., Manning, A., Miller, J., Montzka, S.A., Nassar, R., *et al*. (2007). Long-lived organic compounds (Chapter 1). Sci. Assess. Ozone Deplet. 2006 1–63.

Coulter, C., Hamilton, J.T.G., McRoberts, W.C., Kulakov, L., Larkin, M.J., and Harper, D.B. (1999). Halomethane: bisulfide/halide ion methyltransferase, an unusual corrinoid enzyme of environmental significance isolated from an aerobic methylotroph using chloromethane as the sole carbon source. Appl. Environ. Microbiol. *65*, 4301–4312.

Cox, M.L., Sturrock, G.A., Fraser, P.J., Siems, S.T., Krummel, P.B., and O'Doherty, S. (2003). Regional sources of methyl chloride, chloroform and dichloromethane identified from AGAGE observations at Cape Grim, Tasmania, 1998–2000. J. Atmospheric Chem. *45*, 79–99.

Cox, M.J., Schäfer, H., Nightingale, P.D., McDonald, I.R., and Murrell, J.C. (2012). Diversity of methyl halide-degrading microorganisms in oceanic and coastal waters. FEMS Microbiol. Lett. *334*, 111–118.

Coyotzi, S., Pratscher, J., Murrell, J.C., and Neufeld, J.D. (2016). Targeted metagenomics of active microbial populations with stable-isotope probing. Curr. Opin. Biotechnol. *41*, 1–8.

Crowther, G.J., Kosály, G., and Lidstrom, M.E. (2008). Formate as the main branch point for methylotrophic metabolism in *Methylobacterium extorquens* AM1. J. Bacteriol. *190*, 5057–5062.

Cui, J., Good, N.M., Hu, B., Yang, J., Wang, Q., Sadilek, M., and Yang, S. (2016). Metabolomics revealed an association of metabolite changes and defective growth in *Methylobacterium extorquens* AM1 overexpressing *ecm* during growth on methanol. PloS one *11*, e0154043.

Dallinger, A., and Horn, M.A. (2014). Agricultural soil and drilosphere as reservoirs of new and unusual assimilators of 2,4-dichlorophenol carbon: CAP-SIP of 2,4-DCP degraders in soil and drilosphere. Environ. Microbiol. *16*, 84–100.

David, M., Bringel, F., Pagni, M., Gilmartin, N., BOUBAKRI, H., Nadalig, T., Simonet, P., Vogel, T., Vuilleumier, S. (2008). Diversité et évolution des déshalogénases bactériennes : détection bioinformatique et perspectives de recherche. In: *7ème colloque national "Ressources génétiques"* (p. 83-94). Presented at 7ème colloque national "Ressources génétiques", Strasbourg, FRA (2008-10-13).

Debussche, L., Thibaut, D., Cameron, B., Crouzet, J., and Blanche, F. (1993). Biosynthesis of the corrin macrocycle of coenzyme B_{12} in *Pseudomonas denitrificans*. J. Bacteriol. *175*, 7430–7440.

Dekker, R.F., Vasconcelos, A.-F.D., Barbosa, A.M., Giese, E.C., and Paccola-Meirelles, L. (2001). A new role for veratryl alcohol: regulation of synthesis of lignocellulose-degrading enzymes in the ligninolytic ascomyceteous fungus, *Botryosphaeria* sp.; influence of carbon source. Biotechnol. Lett. *23*, 1987–1993.

DeLeon-Rodriguez, N., Lathem, T.L., Rodriguez-R, L.M., Barazesh, J.M., Anderson, B.E., Beyersdorf, A.J., Ziemba, L.D., Bergin, M., Nenes, A., and Konstantinidis, K.T. (2013). Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric implications. Proc. Natl. Acad. Sci. U. S. A. *110*, 2575–2580.

Delmotte, N., Knief, C., Chaffron, S., Innerebner, G., Roschitzki, B., Schlapbach, R., von Mering, C., and Vorholt, J.A. (2009). Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. Proc. Natl. Acad. Sci. *106*, 16428–16433.

Denman, S.E., Tomkins, N.W., and McSweeney, C.S. (2007). Quantitation and diversity analysis of ruminal methanogenic populations in response to the antimethanogenic compound bromochloromethane: Monitoring of rumen methanogenic *Archaea*. FEMS Microbiol. Ecol. 62, 313–322.

Donnelly, M.I., and Dagley, S. (1980). Production of methanol from aromatic acids by *Pseudomonas putida*. J. Bacteriol. 142, 916–924.

Doronina, N.V., Sokolov, A.P., and Trotsenko, Y.A. (1996). Isolation and initial characterization of aerobic chloromethane-utilizing bacteria. FEMS Microbiol. Lett. *142*, 179–183.

Dourado, M.N., Aparecida Camargo Neves, A., Santos, D.S., and Araújo, W.L. (2015). Biotechnological and agronomic potential of endophytic pink-pigmented methylotrophic *Methylobacterium* spp.. BioMed Res. Int. *2015*, 1–19.

Dugar, G., Herbig, A., Förstner, K.U., Heidrich, N., Reinhardt, R., Nieselt, K., and Sharma, C.M. (2013). High-resolution transcriptome maps reveal strain-specific regulatory features of multiple *Campylobacter jejuni* isolates. PLoS Genet. *9*, e1003495.

Dumont, M.G., and Murrell, J.C. (2005). Stable isotope probing—linking microbial identity to function. Nat. Rev. Microbiol. *3*, 499–504.

Dumont, M.G., Pommerenke, B., Casper, P., and Conrad, R. (2011). DNA-, rRNA- and mRNA-based stable isotope probing of aerobic methanotrophs in lake sediment: Stable isotope probing of methanotrophs. Environ. Microbiol. *13*, 1153–1167.

Duval, M., Simonetti, A., Caldelari, I., and Marzi, S. (2015). Multiple ways to regulate translation initiation in bacteria: Mechanisms, regulatory circuits, dynamics. Biochimie *114*, 18–29.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460–2461.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics *27*, 2194–2200.

Eichorst, S.A., Strasser, F., Woyke, T., Schintlmeister, A., Wagner, M., and Woebken, D. (2015). Advancements in the application of NanoSIMS and Raman microspectroscopy to investigate the activity of microbial cells in soils. FEMS Microbiol. Ecol. *91*, fiv106.

Erb, T.J., Rétey, J., Fuchs, G., and Alber, B.E. (2008). Ethylmalonyl-CoA mutase from *Rhodobactsr sphaeroides* defines a new subclade of conzyme B*12*-dependent Acyl-CoA mutases. J. Biol. Chem. *283*, 32283–32293.

Evershed, R.P., Crossman, Z.M., Bull, I.D., Mottram, H., Dungait, J.A., Maxfield, P.J., and Brennand, E.L. (2006). ^{13}C-Labelling of lipids to investigate microbial communities in the environment. Curr. Opin. Biotechnol. *17*, 72–82.

Evguenieva-Hackenberg, E. (2005). Bacterial ribosomal RNA in pieces: rRNA fragmentation in bacteria. Mol. Microbiol. *57*, 318–325.

Evguenieva-Hackenberg, E., and Klug, G. (2011). New aspects of RNA processing in prokaryotes. Curr. Opin. Microbiol. *14*, 587–592.

Fall, R., and Benson, A. (1996). Leaf methanol-the simplest natural product from plants. Trends in Plant Science, *1*(9), 296-301.

Farhan Ul Haque, M., Nadalig, T., Bringel, F., Schaller, H., and Vuilleumier, S. (2013). Fluorescence-based bacterial bioreporter for specific detection of methyl halide emissions in the environment. Appl. Environ. Microbiol. *79*, 6561–6567.

Fiddaman, P.J., and Rossall, S. Effect of substrate on the production of antifungal volatiles from *Bacillus subtilis*. J. Appli. Bacteriol. *76*, 395-405.

Fitriyanto, N. A., Mako F., Mika M., Ambar P., Tomonori I., et Keiichi Kawai. (2011). Molecular structure and gene analysis of Ce^{3+}-Induced methanol dehydrogenase of *Bradyrhizobium* Sp. MAFF211645. J. Biosco. Bioeng. *111*, 613-617.

Forczek, S.T., Laturnus, F., Doležalová, J., Holík, J., and Wimmer, Z. (2015). Emission of climate relevant volatile organochlorines by plants occurring in temperate forests. Plant Soil Environ. *61*, 103–108.

Francez-Charlot, A., Frunzke, J., Reichen, C., Ebneter, J.Z., Gourion, B., and Vorholt, J.A. (2009). Sigma factor mimicry involved in regulation of general stress response. Proc. Natl. Acad. Sci. *106*, 3467–3472.

Francez-Charlot, A., Kaczmarczyk, A., Fischer, H.-M., and Vorholt, J.A. (2015). The general stress response in *Alphaproteobacteria*. Trends Microbiol. *23*, 164–171.

Frascari, D., Zanaroli, G., and Danko, A.S. (2015). In situ aerobic cometabolism of chlorinated solvents: a review. J. Hazard. Mater. *283*, 382–399.

Freedman, D.L., Swamy, M., Bell, N.C., and Verce, M.F. (2004). Biodegradation of chloromethane by *Pseudomonas aeruginosa* strain NB1 under nitrate-reducing and aerobic conditions. Appl. Environ. Microbiol. *70*, 4629–4634.

Frey-Klett, P., Garbaye, J., and Tarkka, M. (2007). The mycorrhiza helper bacteria revisited. New Phytol. *176*, 22–36.

Galbally, I.E., and Kirstine, W. (2002). The production of methanol by flowering plants and the global cycle of methanol. J. Atmospheric Chem. *43*, 195–229.

Gälli, R., and Leisinger, T. (1985). Specialized bacterial strains for the removal of dichloromethane from industrial waste. Conserv. Recycl. *8*, 91–100.

Georg, J., and Hess, W.R. (2011). Cis-antisense RNA, another level of gene regulation in bacteria. Microbiol. Mol. Biol. Rev. *75*, 286–300.

Goodwin, K.D., Lidstrom, M.E., and Oremland, R.S. (1997). Marine bacterial degradation of brominated methanes. Environ. Sci. Technol. *31*, 3188–3192.

Gourion, B., Rossignol, M., and Vorholt, J.A. (2006). A proteomic study of *Methylobacterium extorquens* reveals a response regulator essential for epiphytic growth. Proc. Natl. Acad. Sci. *103*, 13186–13191.

Greule, M., Huber, S.G., and Keppler, F. (2012). Stable hydrogen-isotope analysis of methyl chloride emitted from heated halophytic plants. Atmos. Environ. *62*, 584–592.

Gribble, G.W. (2003). The diversity of naturally produced organohalogens. Chemosphere *52*, 289–297.

Grostern, A., and Alvarez-Cohen, L. (2013). RubisCO-based CO_2 fixation and C_1 metabolism in the *Actinobacterium Pseudonocardia dioxanivorans* CB1190: Autotrophic growth by *Pseudonocardia dioxanivorans*. Environ. Microbiol. *15*, 3040-3053.

Gruffaz, C., Muller, E.E.L., Louhichi-Jelail, Y., Nelli, Y.R., Guichard, G., and Bringel, F. (2014). Genes of the N-methylglutamate pathway are essential for growth of *Methylobacterium extorquens* DM4 with monomethylamine. Appl. Environ. Microbiol. *80*, 3541–3550.

Guenther, A.B., Jiang, X., Heald, C.L., Sakulyanontvittaya, T., Duhl, T., Emmons, L.K., and Wang, X. (2012). The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions. Geosci. Model Dev. *5*, 1471–1492.

Guo, X., and Lidstrom, M.E. (2008). Metabolite profiling analysis of *Methylobacterium extorquens* AM1 by comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry. Biotechnol. Bioeng. *99*, 929–940.

Halsey, K.H., Carter, A.E., and Giovannoni, S.J. (2012). Synergistic metabolism of a broad range of C_1 compounds in the marine methylotrophic bacterium HTCC2181: C_1 metabolism in the marine isolate HTCC2181. Environ. Microbiol. *14*, 630–640.

Hamilton, J.T.G. (2003). Chloride methylation by plant pectin: an efficient environmentally significant process. Science *301*, 206–209.

Han, J.-I., and Semrau, J.D. (2000). Chloromethane stimulates growth of *Methylomicrobium album* BG8 on methanol. FEMS Microbiol. Lett. *187*, 77–81.

Hancock, T.L.C., Costello, A.M., Lidstrom, M.E., and Oremland, R.S. (1998). Strain IMB-1, a novel bacterium for the removal of methyl bromide in fumigated agricultural soils. Appl. Environ. Microbiol. *64*, 2899–2905.

Harper, D., and Kennedy, J. (1986). Effect of growth conditions on halomethane production by *Phellinus* species: biological and environmental implications. Microbiology *132*, 1231–1246.

Harper, D., Kennedy, J., and Hamilton, J.T. (1988). Chloromethane biosynthesis in poroid fungi. Phytochemistry 27, 3147–3153.

Harper, D.B. (2000). The global chloromethane cycle: biosynthesis, biodegradation and metabolic role. Nat. Prod. Rep. *17*, 337–348.

Harper, D.B., Hamilton, J.T.G., Ducrocq, V., Kennedy, J.T., Downey, A., and Kalin, R.M. (2003). The distinctive isotopic signature of plant-derived chloromethane: possible application in constraining the atmospheric chloromethane budget. Chemosphere *52*, 433–436.

Hartmans, S., Schmuckle, A., Cook, A.M., and Leisinger, T. (1986). Methyl chloride: naturally occurring toxicant and C-1 growth substrate. Microbiology *132*, 1139–1142.

He, X., McLean, J.S., Edlund, A., Yooseph, S., Hall, A.P., Liu, S.Y., Dorrestein, P.C., Esquenazi, E., Hunter, R.C., Cheng, G., *et al.* (2015). Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. Proc. Natl. Acad. Sci. *112*, 244–249.

Herlemann, D.P., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J.J., and Andersson, A.F. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J. *5*, 1571–1579.

Herrmann, A.M., Ritz, K., Nunan, N., Clode, P.L., Pett-Ridge, J., Kilburn, M.R., Murphy, D.V., O'Donnell, A.G., and Stockdale, E.A. (2007). Nano-scale secondary ion mass spectrometry

— A new analytical tool in biogeochemistry and soil ecology: A review article. Soil Biol. Biochem. *39*, 1835–1850.

Hines, M.E., Crill, P.M., Varner, R.K., Talbot, R.W., Shorter, J.H., Kolb, C.E., and Harriss, R.C. (1998). Rapid consumption of low concentrations of methyl bromide by soil bacteria. Appl. Environ. Microbiol. *64*, 1864–1870.

Hirsch, P.R., Mauchline, T.H., and Clark, I.M. (2010). Culture-independent molecular techniques for soil microbial ecology. Soil Biol. Biochem. *42*, 878–887.

Holmes, A.J., Costello, A., Lidstrom, M.E., and Murrell, J.C. (1995). Evidence that participate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. FEMS Microbiol. Lett. *132*, 203–208.

Hope, P.R., Bohmann, K., Gilbert, M.T.P., Zepeda-Mendoza, M.L., Razgour, O., and Jones, G. (2014). Second generation sequencing and morphological faecal analysis reveal unexpected foraging behaviour by *Myotis nattereri* (*Chiroptera, Vespertilionidae*) in winter. Front. Zool. *11*, 1.

Hoppe, B., Krger, K., Kahl, T., Arnstadt, T., Buscot, F., Bauhus, J., and Wubet, T. (2015). A pyrosequencing insight into sprawling bacterial diversity and community dynamics in decaying deadwood logs of *Fagus sylvatica* and *Picea abies*. Sci. Rep. *5*, 9456.

Horiuchi, J., Badri, D.V., Kimball, B.A., Negre, F., Dudareva, N., Paschke, M.W., and Vivanco, J.M. (2007). The floral volatile, methyl benzoate, from snapdragon (*Antirrhinum majus*) triggers phytotoxic effects in *Arabidopsis thaliana*. Planta *226*, 1–10.

Hu, L., Yvon-Lewis, S.A., Liu, Y., Salisbury, J.E., and O'Hern, J.E. (2010). Coastal emissions of methyl bromide and methyl chloride along the eastern Gulf of Mexico and the east coast of the United States: Coastal emissions of CH_3Br and CH_3Cl. Glob. Biogeochem. Cycles *24*.

Huang, Y.Y., Tanaka, M., Vecchio, D., Garcia-Diaz, M., Chang, J., Morimoto, Y., and Hamblin, M.R. (2012). Photodynamic therapy induces an immune response against a bacterial pathogen. Expert Rev. Clin. Immunol. *8*, 479–494.

Huang, B., Lei, C., Wei, C., and Zeng, G. (2014). Chlorinated volatile organic compounds (Cl-VOCs) in environment — sources, potential human health impacts, and current remediation technologies. Environ. Int. *71*, 118–138.

Hung, W.L., Wade, W.G., Boden, R., Kelly, D.P., and Wood, A.P. (2011). Facultative methylotrophs from the human oral cavity and methylotrophy in strains of *Gordonia*, *Leifsonia*, and *Microbacterium*. Arch. Microbiol. *193*, 407–417.

Insam, H., and Seewald, M.S.A. (2010). Volatile organic compounds (VOCs) in soils. Biol. Fertil. Soils *46*, 199–213.

Jacob, D.J. (2005). Global budget of methanol: Constraints from atmospheric observations. J. Geophys. Res. *110*.

Jakočiūnas, T., Jensen, M.K., and Keasling, J.D. (2016). CRISPR/Cas9 advances engineering of microbial cell factories. Metab. Eng. *34*, 44–59.

Janssen, D.B., Dinkla, I.J.T., Poelarends, G.J., and Terpstra, P. (2005). Bacterial degradation of xenobiotic compounds: evolution and distribution of novel enzyme activities. Environ. Microbiol. *7*, 1868–1882.

Jehmlich, N., Schmidt, F., von Bergen, M., Richnow, H.-H., and Vogt, C. (2008). Protein-based stable isotope probing (Protein-SIP) reveals active species within anoxic mixed cultures. ISME J. *2*, 1122–1133.

Kai, M., Crespo, E., Cristescu, S.M., Harren, F.J.M., Francke, W., and Piechulla, B. (2010). *Serratia odorifera*: analysis of volatile emission and biological impact of volatile compounds on *Arabidopsis thaliana*. Appl. Microbiol. Biotechnol. *88*, 965–976.

Kalyuzhnaya, M.G., Lidstrom, M.E., and Chistoserdova, L. (2004). Utility of environmental primers targeting ancient enzymes: methylotroph detection in Lake Washington. Microb. Ecol. *48*, 463–472.

Kalyuzhnaya, M.G., Zabinsky, R., Bowerman, S., Baker, D.R., Lidstrom, M.E., and Chistoserdova, L. (2006). Fluorescence in situ hybridization-flow cytometry-cell sorting-based method for separation and enrichment of type I and type II methanotroph populations. Appl. Environ. Microbiol. *72*, 4293–4301.

Kalyuzhnaya, M.G., Hristova, K.R., Lidstrom, M.E., and Chistoserdova, L. (2008). Characterization of a novel methanol dehydrogenase in representatives of *Burkholderiales*: implications for environmental detection of methylotrophy and evidence for convergent Evolution. J. Bacteriol. *190*, 3817–3823.

Käserer, H. (1906). Über die oxidation des wasserstoves und des methans durch mikroorganismen. Cent. Bakteriol Parasitenkd Infekt. Hyg Abt II *15*, 573–576.

Kaviraj, A., Unlu, E., Gupta, A., and El Nemr, A. (2014). Biomarkers of environmental pollutants. BioMed Res. Int. *2014*, 1–2.

Kayser, M.F., and Vuilleumier, S. (2001). Dehalogenation of dichloromethane by dichloromethane dehalogenase/glutathione S-transferase leads to formation of DNA adducts. J. Bacteriol. *183*, 5209–5212.

Kayser, M.F., Ucurum, Z., and Vuilleumier, S. (2002). Dichloromethane metabolism and C_1 utilization genes in *Methylobacterium* strains. Microbiology *148*, 1915–1922.

Keener, W., and Arp, D. (1993). Kinetic studies of ammonia monooxygenase Inhibition in *Nitrosomonas europaea* by hydrocarbons and halogenated hydrocarbons in an optimized whole-cell assay. Appl Environ Microbiol *59*, 2501–2510.

Keiblinger, K.M., Wilhartitz, I.C., Schneider, T., Roschitzki, B., Schmid, E., Eberl, L., Riedel, K., and Zechmeister-Boltenstern, S. (2012). Soil metaproteomics – Comparative evaluation of protein extraction protocols. Soil Biol. Biochem. *54*, 14–24.

Keppler, F., Borchers, R., Elsner, P., Fahimi, I., Pracht, J., and Schöler, H.F. (2003). Formation of volatile iodinated alkanes in soil: results from laboratory studies. Chemosphere *52*, 477–483.

Keppler, F., Harper, D.B., Röckmann, T., Moore, R.M., and Hamilton, J.T.G. (2005). New insight into the atmospheric chloromethane budget gained using stable carbon isotope ratios. Atmospheric Chem. Phys. *5*, 2403–2411.

Kiefer, P., Portais, J.C., and Vorholt, J.A. (2008). Quantitative metabolome analysis using liquid chromatography–high-resolution mass spectrometry. Anal. Biochem. *382*, 94–100.

Kiefer, P., Delmotte, N., and Vorholt, J.A. (2011). Nanoscale ion-pair reversed-phase HPLC−MS for sensitive metabolome analysis. Anal. Chem. *83*, 850–855.

Kielak, A.M., Scheublin, T.R., Mendes, L.W., van Veen, J.A., and Kuramae, E.E. (2016). Bacterial community succession in pine-wood decomposition. Front. Microbiol. *7*.

Knief, C., Frances, L., and Vorholt, J.A. (2010). Competitiveness of diverse *Methylobacterium* strains in the phyllosphere of *Arabidopsis thaliana* and identification of representative models, including *M. extorquens* PA1. Microb. Ecol. *60*, 440–452.

Knief, C., Delmotte, N., Chaffron, S., Stark, M., Innerebner, G., Wassmann, R., von Mering, C., and Vorholt, J.A. (2012). Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. ISME J. *6*, 1378–1390.

Kögel-Knabner, I. (2002) The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter. Soil Biology and Biochemistry, *34*, 139-162.

Kohler-Staub, D., and Leisinger, T. (1985). Dichloromethane dehalogenase of *Hyphomicrobium* sp. strain DM2. J. Bacteriol. *162*, 676–681.

Kolb, S. (2009). Aerobic methanol-oxidizing Bacteria in soil. FEMS Microbiol. Lett. *300*, 1–10.

Kolb, S., and Stacheter, A. (2013). Prerequisites for amplicon pyrosequencing of microbial methanol utilizers in the environment. Front. Microbiol. *4*, 268.

Korotkova, N., and Lidstrom, M.E. (2001). Connection between poly-hydroxybutyrate biosynthesis and hrowth on C_1 and C_2 compounds in the methylotroph *Methylobacterium extorquens* AM1. J. Bacteriol. *183*, 1038–1046.

Koskimäki, J.J., Pirttilä, A.M., Ihantola, E.-L., Halonen, O., and Frank, A.C. (2015). The Intracellular scots pine shoot symbiont *Methylobacterium extorquens* DSM13060 aggregates around the host nucleus and encodes eukaryote-like proteins. mBio *6*, e00039-15.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and Schloss, P.D. (2013). Development of a dual-Index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. *79*, 5112–5120.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. msw054.

Kurylo, M., and Rodriguez, J. (1998). Short-lived ozone-related compounds. (Scientific assessment of ozone depletion). Chapter 2

La Roche, S.D., and Leisinger, T. (1990). Sequence analysis and expression of the bacterial dichloromethane dehalogenase structural gene, a member of the glutathione S-transferase supergene family. J. Bacteriol. *172*, 164–171.

Laukel, M., Rossignol, M., Borderies, G., Völker, U., and Vorholt, J.A. (2004). Comparison of the proteome of *Methylobacterium extorquens* AM1 grown under methylotrophic and nonmethylotrophic conditions. Proteomics *4*, 1247–1264.

Lee, S.W., Won, K., Lim, H.K., Kim, J.C., Choi, G.J., and Cho, K.Y. (2004). Screening for novel lipolytic enzymes from uncultured soil microorganisms. Appl. Microbiol. Biotechnol. *65*, 720–726.

Leng, L., Chang, J., Geng, K., Lu, Y., and Ma, K. (2015). Uncultivated Methylocystis Species in Paddy Soil Include Facultative Methanotrophs that Utilize Acetate. Microb. Ecol. 70, 88–96.

Lidstrom, M. (2006). Aerobic methylotrophic procaryotes. In The Prokaryotes, pp. 618–634.

Lin, C., Owen, S., and Penuelas, J. (2007). Volatile organic compounds in the roots and rhizosphere of *Pinus* spp. Soil Biol. Biochem. *39*, 951–960.

Liu, X.S. (2007). Getting started in tiling microarray analysis. PLoS Comput. Biol. 3.

Llorens-Rico, V., Cano, J., Kamminga, T., Gil, R., Latorre, A., Chen, W.H., Bork, P., Glass, J.I., Serrano, L., and Lluch-Senar, M. (2016). Bacterial antisense RNAs are mainly the product of transcriptional noise. Sci. Adv. *2*, e1501363–e1501363.

Llusia, J., Peñuelas, J., Guenther, A., and Rapparini, F. (2013). Seasonal variations in terpene emission factors of dominant species in four ecosystems in NE Spain. Atmos. Environ. *70*, 149–158.

Loew, O. (1892). Ueber einen *Bacillus*, welcher ameisensäure und formaldehyd assimiliren kann. Cent. Für Bakteriol. Parasitenkd. Infekt. Hyg. Abt. II *12*, 462–465.

Lopez-Marques, R.L., Perez-Castineira, J.R., Losada, M., and Serrano, A. (2004). Differential regulation of soluble and membrane-bound inorganic pyrophosphatases in the photosynthetic bacterium *Rhodospirillum rubrum* provides insights into pyrophosphate-based stress bioenergetics. J. Bacteriol. *186*, 5418–5426.

Lueders, T., Manefield, M., and Friedrich, M.W. (2003). Enhanced sensitivity of DNA- and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients: Quantitative analysis of SIP gradients. Environ. Microbiol. *6*, 73–78.

Mägli, A., Messmer, M., and Leisinger, T. (1998). Metabolism of dichloromethane by the strict anaerobe *Dehalobacterium formicoaceticum*. Appl. Environ. Microbiol. *64*, 646–650.

Mancuso, S., Taiti, C., Bazihizina, N., Costa, C., Menesatti, P., Giagnoni, L., Arenella, M., Nannipieri, P., and Renella, G. (2015). Soil volatile analysis by proton transfer reaction-time of flight mass spectrometry (PTR-TOF-MS). Appl. Soil Ecol. *86*, 182–191.

Marison, I.W., and Attwood, M.M. (1982). A possible alternative mechanism for the oxidation of formaldehyde to formate. Microbiology *128*, 1441–1446.

Marx, C.J., O'Brien, B.N., Breezee, J., and Lidstrom, M.E. (2003). Novel methylotrophy genes of *Methylobacterium extorquens* AM1 Identified by using transposon mutagenesis including a putative dihydromethanopterin reductase. J. Bacteriol. *185*, 669–673.

Marx, C.J., Dien, S.J.V., and Lidstrom, M.E. (2005). Flux analysis uncovers key role of functional redundancy in formaldehyde metabolism. PLoS Biol *3*, e16.

Marx, C.J. (2008). Development of a broad-host-range *sacB*-based vector for unmarked allelic exchange. BMC Res. Notes 1, 1.

Marx, C.J., Bringel, F., Chistoserdova, L., Moulin, L., Farhan Ul Haque, M., Fleischman, D.E., Gruffaz, C., Jourand, P., Knief, C., Lee, M.C., *et al.* (2012). Complete genome sequences of six strains of the genus *Methylobacterium*. J. Bacteriol. *194*, 4746–4748.

McAnulla, C., Woodall, C.A., McDonald, I.R., Studer, A., Vuilleumier, S., Leisinger, T., and Murrell, J.C. (2001). Chloromethane utilization gene cluster from *Hyphomicrobium chloromethanicum* strain CM2T and development of functional gene Probes to detect halomethane-degrading bacteria. Appl. Environ. Microbiol. *67*, 307–316.

McCulloch, A., and Aucott, M.L. (1999). Global emissions of hydrogen chloride and chloromethane from coal combustion, incineration and industrial activities: reactive chlorine emissions inventory. J. Geophys. Res-Atmos. *104*, 8391-8403.

McCulloch, A., Midgley, P.M., and Ashford, P. (2003). Releases of refrigerant gases (CFC-12, HCFC-22 and HFC-134a) to the atmosphere. Atmos. Environ. *37*, 889–902.

McDonald, I.R., and Murrell, J.C. (1997). The methanol dehydrogenase structural gene *mxaF* and its use as a functional gene probe for methanotrophs and methylotrophs. Appl. Environ. Microbiol. *63*, 3218–3224.

McDonald, I.R., Doronina, N.V., Trotsenko, Y.A., McAnulla, C., and Murrell, J.C. (2001). *Hyphomicrobium chloromethanicum* sp. nov. and *Methylobacterium chloromethanicum* sp. nov., chloromethane-utilizing bacteria isolated from a polluted environment. Int. J. Syst. Evol. Microbiol. *51*, 119–122.

McDonald, I.R., Radajewski, S., and Murrell, J.C. (2005). Stable isotope probing of nucleic acids in methanotrophs and methylotrophs: A review. Org. Geochem. *36*, 779–787.

McDonald, I.R., Bodrossy, L., Chen, Y., and Murrell, J.C. (2008). Molecular Ecology Techniques for the Study of Aerobic Methanotrophs. Appl. Environ. Microbiol. 74, 1305–1315.

Michener, J.K., Vuilleumier, S., Bringel, F., and Marx, C.J. (2014a). Phylogeny poorly predicts the utility of a challenging horizontally transferred gene in *Methylobacterium* strains. J. Bacteriol. *196*, 2101–2107.

Michener, J.K., Neves, A.C., Vuilleumier, S., Bringel, F., and Marx, C.J. (2014b). Effective use of a horizontally-transferred pathway for dichloromethane catabolism requires post–transfer refinement. Elife *3*, e04279.

Miller, L.G., Warner, K.L., Baesman, S.M., Oremland, R.S., McDonald, I.R., Radajewski, S., and Murrell, J.C. (2004). Degradation of methyl bromide and methyl chloride in soil microcosms: Use of stable C isotope fractionation and stable isotope probing to identify reactions and the responsible microorganisms. Geochim. Cosmochim. Acta *68*, 3271–3283.

Mirete, S., Mora-Ruiz, M.R., Lamprecht-Grandío, M., de Figueras, C.G., Rosselló-Móra, R., and González-Pastor, J.E. (2015). Salt resistance genes revealed by functional metagenomics from brines and moderate-salinity rhizosphere within a hypersaline environment. Front. Microbiol. *6*.

Montzka, S.A., and Fraser, P. (2003). Relative contributions of greenhouse gas emissions to global warming.

Montzka, S.A., Dlugokencky, E.J., and Butler, J.H. (2011). Non-CO_2 greenhouse gases and climate change. Nature *476*, 43–50.

Moore, R.M., Gut, A., and Andreae, M.O. (2005). A pilot study of methyl chloride emissions from tropical woodrot fungi. Chemosphere *58*, 221–225.

Muller, E.E.L., Hourcade, E., Louhichi-Jelail, Y., Hammann, P., Vuilleumier, S., and Bringel, F. (2011a). Functional genomics of dichloromethane utilization in *Methylobacterium extorquens* DM4. Environ. Microbiol. *13*, 2518–2535.

Muller, E.E.L., Bringel, F., and Vuilleumier, S. (2011b). Dichloromethane-degrading bacteria in the genomic age. Res. Microbiol. *162*, 869–876.

Nacke, H., Fischer, C., Thürmer, A., Meinicke, P., and Daniel, R. (2014). Land use type significantly affects microbial gene transcription in soil. Microb. Ecol. *67*, 919–930.

Nadalig, T., Farhan Ul Haque, M., Roselli, S., Schaller, H., Bringel, F., and Vuilleumier, S. (2011). Detection and isolation of chloromethane-degrading bacteria from the *Arabidopsis thaliana* phyllosphere, and characterization of chloromethane utilization genes: chloromethane-degrading bacteria from *Arabidopsis thaliana*. FEMS Microbiol. Ecol. *77*, 438–448.

Nadalig, T., Greule, M., Bringel, F., Vuilleumier, S., and Keppler, F. (2013). Hydrogen and carbon isotope fractionation during degradation of chloromethane by methylotrophic bacteria. MicrobiologyOpen *2*, 893–900.

Nadalig, T., Greule, M., Bringel, F., Keppler, F., and Vuilleumier, S. (2014). Probing the diversity of chloromethane-degrading bacteria by comparative genomics and isotopic fractionation. Front. Microbiol. *5*. 523

Nagatoshi, Y., and Nakamura, T. (2007). Characterization of three halide methyltransferases in *Arabidopsis thaliana*. Plant Biotechnol. *24*, 503–506.

Nayak, D.D., and Marx, C.J. (2014). Genetic and phenotypic comparison of facultative methylotrophy between *Methylobacterium extorquens* strains PA1 and AM1. PLoS ONE *9*, e107887.

Nercessian, O., Noyes, E., Kalyuzhnaya, M.G., Lidstrom, M.E., and Chistoserdova, L. (2005). Bacterial populations active in metabolism of C_1 compounds in the sediment of Lake Washington, a freshwater lake. Appl. Environ. Microbiol. *71*, 6885–6899.

Neufeld, J.D., Vohra, J., Dumont, M.G., Lueders, T., Manefield, M., Friedrich, M.W., and Murrell, J.C. (2007). DNA stable-isotope probing. Nat. Protoc. *2*, 860–866.

Neufeld, J.D., Chen, Y., Dumont, M.G., and Murrell, J.C. (2008). Marine methylotrophs revealed by stable-isotope probing, multiple displacement amplification and metagenomics. Environ. Microbiol. *10*, 1526–1535.

Öberg, G.M. (2003). The biogeochemistry of chlorine in soil. In natural production of organohalogen compounds, (Springer), pp. 43–62.

Ochsner, A.M., Sonntag, F., Buchhaupt, M., Schrader, J., and Vorholt, J.A. (2015). *Methylobacterium extorquens*: methylotrophy and biotechnological applications. Appl. Microbiol. Biotechnol. *99*, 517–534.

Oderbolz, D.C., Aksoyoglu, S., Keller, J., Barmpadimos, I., Steinbrecher, R., Skjøth, C.A., Plaß-Dülmer, C., and Prévôt, A.S.H. (2013). A comprehensive emission inventory of biogenic volatile organic compounds in Europe: improved seasonality and land-cover. Atmospheric Chem. Phys. *13*, 1689–1712.

O'Doherty, S., Simmonds, P., Cunnold, D., and Wang, H. (2001). *In situ* chloroform measurements at advanced global atmospheric gases experiment atmospheric research stations from 1994 to 1998. J. Geophys. Res-Atmos. *106*, 20429–20444.

Okubo, Y., Skovran, E., Guo, X., Sivam, D., and Lidstrom, M.E. (2007). Implementation of microarrays for *Methylobacterium extorquens* AM1. OMICS J. Integr. Biol. *11*, 325–340.

Patt, T.E., Cole, G.C., and Hanson, R.S. (1976). *Methylobacterium*, a new genus of facultatively methylotrophic bacteria. Int. J. Syst. Evol. Microbiol. *26*, 226–229.

Peano, C., Pietrelli, A., Consolandi, C., Rossi, E., Petiti, L., Tagliabue, L., De Bellis, G., and Landini, P. (2013). An efficient rRNA removal method for RNA sequencing in GC-rich bacteria. Microb. Inform. Exp. *3*, 1.

Peñuelas, J., Asensio, D., Tholl, D., Wenke, K., Rosenkranz, M., Piechulla, B., and Schnitzler, J.P. (2014). Biogenic volatile emissions from the soil: Biogenic volatile emissions from the soil. Plant Cell Environ. *37*, 1866–1891.

Peyraud, R., Kiefer, P., Christen, P., Massou, S., Portais, J.C., and Vorholt, J.A. (2009). Demonstration of the ethylmalonyl-CoA pathway by using ^{13}C metabolomics. Proc. Natl. Acad. Sci. *106*, 4846–4851.

Peyraud, R., Schneider, K., Kiefer, P., Massou, S., Vorholt, J.A., and Portais, J.C. (2011). Genome-scale reconstruction and system level investigation of the metabolic network of *Methylobacterium extorquens* AM1. BMC Syst. Biol. *5*, 1.

Peyraud, R., Kiefer, P., Christen, P., Portais, J.C., and Vorholt, J.A. (2012). Co-consumption of methanol and succinate by *Methylobacterium extorquens* AM1. PLoS ONE *7*, e48271.

Pham, V.H.T., and Kim, J. (2012). Cultivation of unculturable soil bacteria. Trends Biotechnol. *30*, 475–484.

Polerecky, L., Adam, B., Milucka, J., Musat, N., Vagner, T., and Kuypers, M.M.M. (2012). Look@NanoSIMS - a tool for the analysis of nanoSIMS data in environmental microbiology. Environ. Microbiol. 14, 1009–1023.

Prinn, R., Weiss, R., and Fraser, P. (2000). A history of chemically and radiatively important gases in air from ALE/GAGE/AGAGE. J. Geophys. Res. *105*, 17751–17792.

Prosser, J.I. (2015). Dispersing misconceptions and identifying opportunities for the use of 'omics' in soil microbial ecology. Nat. Rev. Microbiol. 13, 439–446.

Radajewski, S., Webster, G., Reay, D.S., Morris, S.A., Ineson, P., Nedwell, D.B., Prosser, J.I., and Murrell, J.C. (2002). Identification of active methylotroph populations in an acidic forest soil by stable-isotope probingc. Microbiology *148*, 2331–2342.

Rasche, M.E., Hyman, M.R., and Arp, D.J. (1991). Factors limiting aliphatic chlorocarbon degradation by *Nitrosomonas europaea*: cometabolic inactivation of ammonia monooxygenase and substrate specificity. Appl. Environ. Microbiol. *57*, 2986–2994.

Rappé, M.S., and Giovannoni, S.J. (2003). The uncultured microbial majority. Annu. Rev. Microbiol. *57*, 369–394.

Reaser, B.C., Yang, S., Fitz, B.D., Parsons, B.A., Lidstrom, M.E., and Synovec, R.E. (2016). Non-targeted determination of ^{13}C-labeling in the *Methylobacterium extorquens* AM1 metabolome using the two-dimensional mass cluster method and principal component analysis. J. Chromatogr. A *1432*, 111–121.

Redeker, K.R., Wang, N.Y., Low, J.C., McMillan, A., Tyler, S.C., and Cicerone, R.J. (2000). Emissions of methyl halides and methane from rice paddies. Science *290*, 966–969.

Redeker, K.R., Treseder, K.K., and Allen, M.F. (2004). Ectomycorrhizal fungi: a new source of atmospheric methyl halides? Glob. Change Biol. *10*, 1009–1016.

Redeker, K.R., and Kalin, R.M. (2012). Methyl chloride isotopic signatures from Irish forest soils and a comparison between abiotic and biogenic methyl halide soil fluxes. Glob. Change Biol. *18*, 1453–1467.

Rettedal, E.A., and Brözel, V.S. (2015). Characterizing the diversity of active bacteria in soil by comprehensive stable isotope probing of DNA and RNA with H_2 ^{18}O. MicrobiologyOpen *4*, 208–219.

Rhew, R.C., Østergaard, L., Saltzman, E.S., and Yanofsky, M.F. (2003). Genetic control of methyl halide production in *Arabidopsis*. Curr. Biol. *13*, 1809–1813.

Rhew, R.C., and Abel, T. (2007). Measuring simultaneous production and consumption fluxes of methyl chloride and methyl bromide in annual temperate grasslands. Environ. Sci. Technol. *41*, 7837–7843.

Rhew, R.C., Chen, C., Teh, Y.A., and Baldocchi, D. (2010). Gross fluxes of methyl chloride and methyl bromide in a California oak-savanna woodland. Atmos. Environ. *44*, 2054–2061.

Rigamonte, T.A., Pylro, V.S., and Duarte, G.F. (2010). The role of mycorrhization helper bacteria in the establishment and action of ectomycorrhizae associations. Braz. J. Microbiol. *41*, 832–840.

Rocca, J.D., Hall, E.K., Lennon, J.T., Evans, S.E., Waldrop, M.P., Cotner, J.B., Nemergut, D.R., Graham, E.B., and Wallenstein, M.D. (2015). Relationships between protein-encoding gene abundance and corresponding process are commonly assumed yet rarely observed. ISME J. *9*, 1693–1699.

Roche, S.D.L., and Leisinger, T. (1991). Identification of *dcmR*, the regulatory gene governing expression of dichloromethane dehalogenase in *Methylobacterium* sp. strain DM4. J. Bacteriol. *173*, 6714–6721.

Roselli, S. (2009). Génomique fonctionnelle de la dégradation microbienne du chlorométhane. Thèse de doctorat, Université de Strasbourg.

Roselli, S., Nadalig, T., Vuilleumier, S., and Bringel, F. (2013). The 380 kb pCMU01 plasmid encodes chloromethane utilization genes and redundant genes for vitamin B12- and tetrahydrofolate-dependent chloromethane metabolism in *Methylobacterium extorquens* CM4: A proteomic and bioinformatics study. PLoS ONE *8*, e56598.

Ruecker, A., Weigold, P., Behrens, S., Jochmann, M., Laaks, J., and Kappler, A. (2014). Predominance of biotic over abiotic formation of halogenated hydrocarbons in hypersaline sediments in Western Australia. Environ. Sci. Technol. *48*, 9170–9178.

Rüffer, M. (2013). Spatial distribution of *cmuA*-genotypes and CH_3Cl oxidation in a mixed hardwood forest soil (Steigerwald). Research proposal.

Sailaukhanuly, Y., Sárossy, Z., Carlsen, L., and Egsgaard, H. (2014). Mechanistic aspects of the nucleophilic substitution of pectin. On the formation of chloromethane. Chemosphere *111*, 575–579.

Saito, T., and Yokouchi, Y. (2008). Stable carbon isotope ratio of methyl chloride emitted from glasshouse-grown tropical plants and its implication for the global methyl chloride budget. Geophys. Res. Lett. *35*.

Saito, T., Yokouchi, Y., Kosugi, Y., Tani, M., Philip, E., and Okuda, T. (2008). Methyl chloride and isoprene emissions from tropical rain forest in Southeast Asia. Geophys. Res. Lett. *35*.

Schade, G., and Goldstein, A. (2001). Fluxes of oxygenated volatile organic compounds from a ponderosa pine plantation. J. Geophys. Res. *106*, 3111–3123.

Schaefer, J.K. (2002). *Leisingera methylohalidivorans* gen. nov., sp. nov., a marine methylotroph that grows on methyl bromide. Int. J. Syst. Evol. Microbiol. *52*, 851–859.

Schäfer, H., McDonald, I.R., Nightingale, P.D., and Murrell, J.C. (2005). Evidence for the presence of a CmuA methyltransferase pathway in novel marine methyl halide-oxidizing bacteria: novel marine methyl halide-oxidizing bacteria. Environ. Microbiol. *7*, 839–852.

Schäfer, H., Miller, L.G., Oremland, R.S., and Murrell, J.C. (2007). Bacterial cycling of methyl halides. Adv Appl Microbiol. *61*, 307-346.

Schauffler, S.M. (2003). Chlorine budget and partitioning during the stratospheric aerosol and gas experiment (SAGE) III ozone loss and validation experiment (SOLVE). J. Geophys. Res. *108*.

Schink, B., and Zeikus, J. (1980). Microbial methanol formation: a major end product of pectin metabolism. Curr. Microbiol. 6, 387–389.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., *et al*. (2009). Introducing mothur: Open-source, platform-independent, community-supported Software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75, 7537–7541.

Schmidt, S., Christen, P., Kiefer, P., and Vorholt, J.A. (2010). Functional investigation of methanol dehydrogenase-like protein XoxF in *Methylobacterium extorquens* AM1. Microbiology *156*, 2575–2586.

Schmidt, P.A., Bálint, M., Greshake, B., Bandow, C., Römbke, J., and Schmitt, I. (2013). Illumina metabarcoding of a soil fungal community. Soil Biol. Biochem. *65*, 128–132.

Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiß, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., *et al.* (2010). The primary transcriptome of the major human pathogen *Helicobacter pylori*. Nature *464*, 250–255.

Sharma, C.M., and Vogel, J. (2014). Differential RNA-seq: the approach behind and the biological insight gained. Curr. Opin. Microbiol. 19, 97–105.

Shorter, J., Kolb, C., and Crill, P. (1995). Rapid degradation of atmospheric methyl bromide in soil. Nature *377*, 717–719.

Shukla, A.K., Vishwakarma, P., Upadhyay, S.N., Tripathi, A.K., Prasana, H.C., and Dubey, S.K. (2009). Biodegradation of trichloroethylene (TCE) by methanotrophic community. Bioresour. Technol. *100*, 2469–2474.

Skovran, E., Palmer, A.D., Rountree, A.M., Good, N.M., and Lidstrom, M.E. (2011b). XoxF is required for expression of methanol dehydrogenase in *Methylobacterium extorquens* AM1. J. Bacteriol. *193*, 6032–6038.

Šmejkalová, H., Erb, T.J., and Fuchs, G. (2010). Methanol assimilation in *Methylobacterium extorquens* AM1: demonstration of all enzymes and their regulation. PLoS ONE *5*, e13001.

Söhngen, N.L. (1906). Uber bakterien welche methan ab kohlenstoffnahrung und energiequelle gerbrauchen. Zbl Bact Parasitenk *15*, 513–517.

Stacheter, A., Noll, M., Lee, C.K., Selzer, M., Glowik, B., Ebertsch, L., Mertel, R., Schulz, D., Lampert, N., Drake, H.L and Kolb, S. (2013). Methanol oxidation by temperate soils and environmental determinants of associated methylotrophs. ISME J. *7*, 1051–1064.

Studer, A., Vuilleumier, S., and Leisinger, T. (1999). Properties of the methylcobalamin: H4folate methyltransferase involved in chloromethane utilization by *Methylobacterium* sp. strain CM4. Eur. J. Biochem. *264*, 242–249.

Studer, A., Stupperich, E., Vuilleumier, S., and Leisinger, T. (2001). Chloromethane: tetrahydrofolate methyl transfer by two proteins from *Methylobacterium chloromethanicum* strain CM4. Eur. J. Biochem. FEBS *268*, 2931–2938.

Studer, A., McAnulla, C., Buchele, R., Leisinger, T., and Vuilleumier, S. (2002). Chloromethane-induced genes define a third C_1 utilization pathway in *Methylobacterium chloromethanicum* CM4. J. Bacteriol. *184*, 3476–3484.

Talia, P., Sede, S.M., Campos, E., Rorig, M., Principi, D., Tosto, D., Hopp, H.E., Grasso, D., and Cataldi, A. (2012). Biodiversity characterization of cellulolytic bacteria present on native Chaco soil by comparison of ribosomal RNA genes. Res. Microbiol. *163*, 221–232.

Tang, J.K.-H., You, L., Blankenship, R.E., and Tang, Y.J. (2012). Recent advances in mapping environmental microbial metabolisms through 13C isotopic fingerprints. J. R. Soc. Interface 9, 2767–2780.

Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. *10*, 512–526.

Taubert, M., Grob, C., Howat, A.M., Burns, O.J., Dixon, J.L., Chen, Y., and Murrell, J.C. (2015). XoxF encoding an alternative methanol dehydrogenase is widespread in coastal marine environments: *xoxF* in costal marine environments. Environ. Microbiol. *17*, 3937–3948.

Temkiv, T.Š., Finster, K., Hansen, B.M., Nielsen, N.W., and Karlson, U.G. (2012). The microbial diversity of a storm cloud as assessed by hailstones. FEMS Microbiol. Ecol. *81*, 684–695.

Traunecker, J., Preuss, A., and Diekert, G. (1991a). Isolation and characterization of a methyl chloride utilizing, strictly anaerobic bacterium. Arch. Microbiol. *156*, 416–421.

United Nations Environment Programm (UNEP) (2005). Production and consumption of ozone depleting substances under the Montreal protocol 1986-2004.

Van Aken, B. (2004). *Methylobacterium populi* sp. nov., a novel aerobic, pink-pigmented, facultatively methylotrophic, methane-utilizing bacterium isolated from poplar trees (Populus deltoidesxnigra DN34). Int. J. Syst. Evol. Microbiol. *54*, 1191–1196.

Van Assche, E., Van Puyvelde, S., Vanderleyden, J., and Steenackers, H.P. (2015). RNA-binding proteins involved in post-transcriptional regulation in bacteria. Front. Microbiol. *6*.

Van der Klei, I.J., Yurimoto, H., Sakai, Y., and Veenhuis, M. (2006). The significance of peroxisomes in methanol metabolism in methylotrophic yeast. Biochim. Biophys. Acta BBA - Mol. Cell Res. *1763*, 1453–1462.

VanDien, S.J., Marx, C.J., O'Brien, B.N., and Lidstrom, M.E. (2003). Genetic characterization of the carotenoid biosynthetic pathway in *Methylobacterium extorquens* AM1 and isolation of a colorless mutant. Appl. Environ. Microbiol. *69*, 7563–7566.

Van Dijk, E.L., Jaszczyszyn, Y., and Thermes, C. (2014). Library preparation methods for next-generation sequencing: Tone down the bias. Exp. Cell Res. *322*, 12–20.

Vannelli, T., Studer, A., Kertesz, M., and Leisinger, T. (1998). Chloromethane metabolism by *Methylobacterium* sp. Strain CM4. Appl. Environ. Microbiol. *64*, 1933–1936.

Vannelli, T., Messmer, M., Studer, A., Vuilleumier, S., and Leisinger, T. (1999). A corrinoid-dependent catabolic pathway for growth of a *Methylobacterium* strain with chloromethane. Proc. Natl. Acad. Sci. *96*, 4615–4620.

Vorholt, J.A., Chistoserdova, L., Stolyar, S.M., Thauer, R.K., and Lidstrom, M.E. (1999). Distribution of tetrahydromethanopterin-dependent enzymes in methylotrophic bacteria and phylogeny of methenyl tetrahydromethanopterin cyclohydrolases. J. Bacteriol. *181*, 5750–5757.

Vorholt, J.A. (2002). Cofactor-dependent pathways of formaldehyde oxidation in methylotrophic bacteria. Arch. Microbiol. *178*, 239–249.

Vorholt, J.A. (2012). Microbial life in the phyllosphere. Nat. Rev. Microbiol. *10*, 828–840.

Vos, M., Wolf, A.B., Jennings, S.J., and Kowalchuk, G.A. (2013). Micro-scale determinants of bacterial diversity in soil. FEMS Microbiol. Rev. *37*, 936–954.

Vu, H.N., Subuyuj, G.A., Vijayakumar, S., Good, N.M., Martinez-Gomez, N.C., and Skovran, E. (2016). lanthanide-dependent regulation of methanol oxidation systems in *Methylobacterium extorquens* AM1 and their contribution to methanol growth. J. Bacteriol. *198*, 1250–1259.

Vuilleumier, S., Ivoš, N., Dean, M., and Leisinger, T. (2001). Sequence variation in dichloromethane dehalogenases/glutathione S-transferases. Microbiology *147*, 611–619.

Vuilleumier, S. (2002). Coping with a halogenated one-carbon diet: aerobic dichloromethane-mineralising bacteria. In Biotechnology for the Environment, Focus on Biotechnology Serie, pp. 105–131.

Vuilleumier, S., Chistoserdova, L., Lee, M.-C., Bringel, F., Lajus, A., Zhou, Y., Gourion, B., Barbe, V., Chang, J., Cruveiller, S., *et al*. (2009). *Methylobacterium* genome sequences: a reference blueprint to investigate microbial metabolism of C_1 compounds from natural and industrial sources. PLoS ONE *4*, e5584.

Vuilleumier, S., Nadalig, T., Farhan Ul Haque, M., Magdelenat, G., Lajus, A., Roselli, S., Muller, E.E.L., Gruffaz, C., Barbe, V., Medigue, C., *et al*. (2011). Complete genome sequence of the chloromethane-degrading *Hyphomicrobium* sp. strain MC1. J. Bacteriol. *193*, 5035–5036.

Wagner, M., Nielsen, P.H., Loy, A., Nielsen, J.L., and Daims, H. (2006). Linking microbial community structure with function: fluorescence in situ hybridization-microautoradiography and isotope arrays. Curr. Opin. Biotechnol. *17*, 83–91.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.

Warneke, C., Karl, T., and Judmaier, H. (1999). Acetone, methanol, and other partially oxidized volatile organic emissions from dead plant matter by abiologicalp rocesses: Significance for atmospheric HOx chemistry. Global Biogeochemical Cycles *13*, 9–17.

Warwick, S., Bowen, T., McVeigh, H., and Embley, T.M. (1994). A phylogenetic analysis of the family *Pseudonocardiaceae* and the genera *Actinokineospora* and *Saccharothrix* with 16S rRNA sequences and a proposal to combine the genera *Amycolata* and *Pseudonocardia* in an emended genus *Pseudonocardia*. Int. J. Syst. Evol. Microbiol. *44*, 293–299.

Watling, R., and Harper, D.B. (1998). Chloromethane production by wood-rotting fungi and an estimate of the global flux to the atmosphere. Mycol. Res. *102*, 769–787.

White, R.H. (1982). Biosynthesis of methyl chloride in the fungus *Phellinus pomaceus*. Arch. Microbiol. *132*, 100–102.

Whittenbury, R., Phillips, K.C., and Wilkinson, J.F. (1970). Enrichment, isolation and some properties of methane-utilizing bacteria. Microbiology *61*, 205–218.

Wohlfahrt, G., Amelynck, C., Ammann, C., Arneth, A., Bamberger, I., Goldstein, A.H., Gu, L., Guenther, A., Hansel, A., Heinesch, B., *et al*. (2015). An ecosystem-scale perspective of the net land methanol flux: synthesis of micrometeorological flux measurements. Atmospheric Chem. Phys. *15*, 7413–7427.

Wishkerman, A., Gebhardt, S., McRoberts, C.W., Hamilton, J.T.G., Williams, J., and Keppler, F. (2008). Abiotic methyl bromide formation from vegetation, and its strong dependence on temperature. Environ. Sci. Technol. *42*, 6837–6842.

Wuosmaa, A., and Hager, L. (1990). Methyl chloride transferase: a carbocation route for biosynthesis of halometabolites. Science *249*, 160–162.

Wyborn, N.R., Mills, J., Williams, S.G., and Jones, C.W. (1996). Molecular characterisation of formamidase from Methylophilus methylotrophus. Eur. J. Biochem. 240, 314–322.

Xiao, X. (2008). Optimal estimation of the surface fluxes of chloromethanes using a 3-D global atmospheric chemical transport model. Massachusetts Institute of Technology.

Xie, S., Sun, W., Luo, C., and Cupples, A.M. (2011). Novel aerobic benzene degrading microorganisms identified in three soils by stable isotope probing. Biodegradation *22*, 71–81.

Yang, S., Hoggard, J.C., Lidstrom, M.E., and Synovec, R.E. (2013). Comprehensive discovery of 13C labeled metabolites in the bacterium *Methylobacterium extorquens* AM1 using gas chromatography–mass spectrometry. J. Chromatogr. A *1317*, 175–185.

Yoshida, Y., Wang, Y., Zeng, T., and Yantosca, R. (2004). A three-dimensional global model study of atmospheric methyl chloride budget and distributions. J. Geophys. Res. *109*.

Yoshida, Y., Wang, Y., Shim, C., Cunnold, D., Blake, D.R., and Dutton, G.S. (2006). Inverse modeling of the global methyl chloride sources. J. Geophys. Res. *111*.

Yokouchi, Y., Saito, T., Ishigaki, C., and Aramoto, M. (2007). Identification of methyl chloride-emitting plants and atmospheric measurements on a subtropical island. Chemosphere *69*, 549–553.

Yurimoto, H. (2009). Molecular basis of methanol-inducible gene expression and its application in the methylotrophic yeast *Candida boidinii*. Biosci. Biotechnol. Biochem. *73*, 793–800.

Zahn, K., Inui, M., and Yukawa, H. (2000). Divergent mechanisms of 5’23S rRNA IVS processing in the α-*proteobacteria*. Nucleic acids research *28*, 4623–4633.

APPENDICES

UNIVERSITE DE STRASBOURG

RESUME DE LA THESE DE DOCTORAT

Discipline : Sciences du vivant

Spécialité : Aspects moléculaires et cellulaires de la biologie

Présentée par : Pauline CHAIGNAUD

Cotutelle de thèse entre l'Université de Strasbourg et l'Université de Bayreuth (Allemagne)

Titre : Le rôle des bactéries dans le filtrage du chlorométhane, un gaz destructeur de la couche d'ozone – Des souches modèles aux communautés microbiennes de sols forestiers

Unité de Recherche : UMR7156 UNISTRA CNRS
Laboratoire de Génétique Moléculaire, Génomique, Microbiologie (GMGM)
Département "Micro-organismes, génomes, environnement"
Equipe "Adaptations et Interactions Microbiennes dans l'Environnement"
28 rue Goethe, 67083 Strasbourg

Directrice de thèse : Françoise BRINGEL, Directrice de Recherche au CNRS
Localisation: Laboratoire GMGM, 28 rue Goethe, 67000 Strasbourg

Co-Directeur de thèse : Steffen KOLB, Directeur d'institut Landscape Biogeochemistry (LBG)
Localisation: ZALF, Leibniz Centre for Agricultural Landscape Research, Müncheberg, Allemagne

ECOLES DOCTORALES :

☐ ED - Sciences de l'Homme et des sociétés	☐ ED 269 - Mathématiques, sciences de l'information et de l'ingénieur
☐ ED 99 – Humanités	☐ ED 270 – Théologie et sciences religieuses
☐ ED 101 – Droit, sciences politique et histoire	☐ ED 413 – Sciences de la terre, de l'univers et de l'environnement
☐ ED 182 – Physique et chimie physique	☒ ED 414 – Sciences de la vie et de la santé
☐ ED 221 – Augustin Cournot	
☐ ED 222 - Sciences chimiques	

1. Introduction

Le chlorométhane (CH_3Cl) est un composé organique volatile, responsable de 16 % de la dégradation de l'ozone stratosphérique, liée aux composés chlorés [1]. Ses sources et ses puits sont surtout naturels [2,3]. Les échanges de chlorométhane entre les écosystèmes terrestres et l'atmosphère peuvent être modulés par les microorganismes capables de dégrader le chlorométhane, notamment ceux du sol [4,5]. De fait, le sol forestier constitue un puits important estimé à 137 Gg. an^{-1} [6,7]. Cependant, le rôle de filtre biologique des bactéries dans les flux d'émission de chlorométhane dans l'atmosphère reste mal évalué dans le budget global du chlorométhane. A ce jour, la seule voie métabolique connue de dégradation du chlorométhane, appelée voie *cmu* pour « ChloroMethane Utilization », a été caractérisée en condition oxique chez *Methylobacterium extorquens* CM4 [8,9]. Cette bactérie est méthylotrophe, c'est-à-dire capable d'utiliser des composés réduits sans liaison carbone-carbone comme unique source de carbone et d'énergie tels que le chlorométhane et le méthanol (CH_3OH), qui est le substrat de référence de la méthylotrophie. Les gènes de la voie *cmu* sont portés par le plasmide pCMU01, essentiel à l'utilisation du chlorométhane, qui contient également des gènes impliqués dans le métabolisme des cofacteurs associés à la voie *cmu*, le tétrahydrofolate (H_4F) et un cofacteur corrinoïde [10]. Parmi les gène *cmu*, le gène *cmuA* codant la chlorométhane-méthyltransférase est essentiel à la croissance sur chlorométhane et conservé chez d'autres souches chlorométhane-dégradantes [11]. C'est pourquoi, le gène *cmuA* est utilisé comme gène fonctionnel marqueur de la dégradation bactérienne du chlorométhane et sa présence a été mise en évidence au sein de communautés d'*Alphaproteobacteria* chlorométhane-dégradantes d'environnements marins et de sols [4,11–14]. L'utilisation microbienne du chlorométhane comme unique source de carbone et d'énergie dans le sol est restreinte par le fait que les concentrations ambiantes sont faibles, de l'ordre de 1 pM de chlorométhane par g de sol [15]. Pour activement contribuer au filtrage du chlorométhane émis même en faible quantité, les bactéries chlorométhane-dégradante du sol pourraient en plus du chlorométhane utiliser un autre composé en C1 présent en plus fortes concentrations dans le sol, tel que méthanol dont la concentration est estimée à 1 $nM.g^{-1}$ [16]. Un des objectifs de mon projet de thèse était de tester cette hypothèse en caractérisant des communautés bactériennes méthylotrophes de sols forestiers capables d'utiliser le chlorométhane seul ou en présence de méthanol.

2. Projet de thèse

Mon projet de thèse s'est articulé autour de deux volets : de l'étude d'une souche modèle cultivée au laboratoire vers l'étude de communautés bactériennes chlorométhane-dégradantes en microcosmes de sols forestiers. Le premier volet a consisté à approfondir nos connaissances fondamentales de la réponse adaptative au niveau transcriptionnel de la souche modèle *M. extorquens* CM4 à la croissance sur le chlorométhane et le méthanol ajoutés comme unique source de carbone et d'énergie. Pour mieux cibler la réponse spécifique à la croissance sur le chlorométhane, une approche de transcriptomique comparative par **RNA sequencing** a été réalisée chez *Methylobacterium extorquens* CM4, mais également chez une souche proche phylogénétiquement, *Methylobacterium extorquens* DM4, capable d'utiliser le dichlorométhane (CH_2Cl_2) et non le chlorométhane comme unique source de carbone et d'énergie. Les transcriptomes des deux souches en condition de croissance sur des méthanes chlorés (chlorométhane ou dichlorométhane) ont été comparés à ceux obtenus en croissance sur le méthanol, et ceci en collaboration avec le Génoscope. Dans le deuxième volet de mes travaux, une approche de type **Stable Isotope Probing** (SIP) [17] a été appliquée à des microcosmes préparés à partir de sol forestier, incubés avec du méthanol et/ou du chlorométhane marqués avec du ^{13}C, afin d'être incorporés spécifiquement dans l'ADN de communautés actives dans l'assimilation de ces composés. L'ADN enrichi en ^{13}C (ADN « lourd ») a été séparé de l'ADN non marqué (« léger » portant du ^{12}C) par ultracentifugation. A partir de l'ADN « lourd », la diversité bactérienne est estimée par séquençage de produits PCR d'un marqueur taxonomique (gène codant l'ARNr 16S) et de marqueurs fonctionnels ciblant la voie *cmu* (gène *cmuA*), la dégradation du méthanol (gène codant une sous-unité d'une méthanol déshydrogénase) et le métabolisme méthylotrophe central (gène *mch*) [16].

3. Résultats

3.1. Réponse transcriptionnelle d'une bactérie modèle à la croissance sur le chlorométhane en condition contrôlée au laboratoire. Parmi les 6276 CDS du génome de *M. extorquens* CM4, 137 ont été détectés comme différentiellement abondants uniquement en chlorométhane, avec comme attendu, les gènes *cmu* et d'autres gènes impliqués dans le métabolisme et le transport de cofacteurs essentiels à l'utilisation du chlorométhane par la voie *cmu* portée par le plasmide pCMU01. De plus, des gènes communs aux génomes de *M. extorquens* (core génome) pourraient jouer un rôle potentiel dans l'expulsion de protons produits lors de la déshalogénation (gène *hppA* codant une pyrophosphatase translocatrice de protons), la

régénération du NADP (gène *pnt* codant une « membrane-bound transhydrogenase ») et le métabolisme C1 du H_4F (gènes *gcvPHT* du complexe de clivage de la glycine). L'ensemble de ces données de transcriptomique, en accord avec des travaux précédents [18], démontre le rôle central du plasmide pCMU01 porteur de gènes plus exprimés en condition de culture sur chlorométhane non seulement pour les gènes essentiels de la voie *cmu*, mais aussi d'une vingtaine de gènes ayant des homologues portés par le chromosome, qui eux ne sont pas différentiellement exprimés. Enfin, l'utilisation du chlorométhane semble peu corrélée à celle du dichlorométhane (12 gènes du core génome ont des transcrits plus abondants en culture sur méthanes chlorés comparé au méthanol). Un 1er article à soumettre sous peu rassemble les résultats du travail effectué à Strasbourg.

3.2. Caractérisation de la diversité des communautés bactériennes dégradant le chlorométhane dans des microcosmes de sol forestier. Les résultats du séquençage des fractions enrichies en ^{13}C ont permis de mettre en évidence la présence de nouvelles communautés bactériennes capables d'utiliser le méthanol et le chlorométhane. Des *Alphaproteobacteria* du genre *Methylovirgula* ont été détectées comme majoritaires dans la communauté bactérienne capable d'utiliser soit le chlorométhane, soit le méthanol comme source de carbone. L'ajout simultané de méthanol et de chlorométhane a permis la caractérisation d'une nouvelle famille de bactéries assimilatrices de chlorométhane apparentées aux *Actinomycetales* du genre *Actinothermus*. L'analyse des séquences du gène *cmuA* a révélé qu'une partie des OTUs (Operational Taxonomic Unit) identifiées sont proches phylogénétiquement du gène *cmuA* de *Methylobacterium extorquens* CM4, alors que d'autres sont affiliées à des souches non cultivables de l'environnement. Un 2e article en préparation rassemblera les résultats des travaux réalisés à Bayreuth.

4. Conclusions et perspectives

L'étude du transcriptome par du RNAseq, de deux souches phylogénétiquement proches, *Methylobacterium extorquens* CM4 et DM4, a permis une meilleure compréhension de l'implication des génomes communs et spécifiques dans l'adaptation liée à la croissance en présence de méthanol ou de chlorométhane, démontrant l'importance des génomes variables dans cette adaptation. Cependant, les éléments et les mécanismes de régulation liés au métabolisme des composés en C_1 restent encore inconnus. Dans le but de les identifier, des banques directionnelles pour l'identification des sites d'initiation de la transcription ont été préparées et sont actuellement en cours de séquençage. L'approche SIP a permis une première

évaluation de la diversité des communautés bactériennes chlorométhane-dégradantes, avec ou sans ajout de méthanol, de sols forestiers. Des bactéries du genre *Methylovirgula* ou *Acidothermus* ont été nouvellement identifiés comme actives dans l'assimilation du chlorométhane dans des microcosmes de sols forestiers. Des approches non culturales (métagénomique; marquage FISH pour « Fluorescence In Situ Hybridation »; PCR quantitative) couplées à des approches culturales permettraient à l'avenir de mieux estimer l'abondance, la diversité, les voies métaboliques impliquées dans le filtrage du chlorométhane par les communautés microbiennes du sol.

5. Bibliographie

1. Montzka S, Reimann S, O'Doherty S, Engel A, Krüger K, Sturges WT. Ozone-depleting substances (ODSs) and related chemicals. 2011; Available: http://oceanrep.geomar.de/10405/2/03-Chapter_1.pdf

2. Cox MJ, Schäfer H, Nightingale PD, McDonald IR, Murrell JC. Diversity of methyl halide-degrading microorganisms in oceanic and coastal waters. FEMS Microbiol Lett. 2012;334: 111–118. doi:10.1111/j.1574-6968.2012.02624.x

3. Keppler F, Harper DB, Röckmann T, Moore RM, Hamilton JTG. New insight into the atmospheric chloromethane budget gained using stable carbon isotope ratios. Atmospheric Chem Phys. 2005;5: 2403–2411.

4. Borodina E, Cox MJ, McDonald IR, Murrell JC. Use of DNA-stable isotope probing and functional gene probes to investigate the diversity of methyl chloride-utilizing bacteria in soil. Environ Microbiol. 2005;7: 1318–1328.

5. Clerbaux C, Cunnold DM, Anderson J, Engel A, Fraser PJ, Mahieu E, et al. Long-lived organic compounds (Chapter 1). Sci Assess Ozone Deplet 2006. 2007; Available: http://orbi.ulg.ac.be/handle/2268/34986

6. Harper DB. The global chloromethane cycle: biosynthesis, biodegradation and metabolic role (1982 to January 2000). Nat Prod Rep. 2000;17: 337–348. doi:10.1039/a809400d

7. Yoshida Y. A three-dimensional global model study of atmospheric methyl chloride budget and distributions. J Geophys Res. 2004;109. doi:10.1029/2004JD004951

8. Vannelli T, Studer A, Kertesz M, Leisinger T. Chloromethane metabolism by *Methylobacterium* sp. strain CM4. Appl Environ Microbiol. 1998;64: 1933–1936. doi:05/1998; 64(5):1933-6.

9. Vannelli T, Messmer M, Studer A, Vuilleumier S, Leisinger T. A corrinoid-dependent catabolic pathway for growth of a *Methylobacterium* strain with chloromethane. Proc Natl Acad Sci. 1999;96: 4615–4620. doi:10.1073/pnas.96.8.4615

10. Messmer M, Wohlfarth G, Diekert G. Methyl chloride metabolism of the strictly anaerobic, methyl chloride-utilizing homoacetogen strain MC. Arch Microbiol. 1993;160: 383–387.

11. Nadalig T, Farhan Ul Haque M, Roselli S, Schaller H, Bringel F, Vuilleumier S. Detection and isolation of chloromethane-degrading bacteria from the *Arabidopsis thaliana* phyllosphere, and characterization of chloromethane utilization genes: chloromethane-degrading bacteria from *A. thaliana*. FEMS Microbiol Ecol. 2011;77: 438–448. doi:10.1111/j.1574-6941.2011.01125.x

12. Miller LG, Warner KL, Baesman SM, Oremland RS, McDonald IR, Radajewski S, et al. Degradation of methyl bromide and methyl chloride in soil microcosms: Use of stable C isotope fractionation and stable isotope probing to identify reactions and the responsible microorganisms. Geochim Cosmochim Acta. 2004;68: 3271–3283. doi:10.1016/j.gca.2003.11.028

13. Nadalig T, Greule M, Bringel F, Keppler F, Vuilleumier S. Probing the diversity of chloromethane-degrading bacteria by comparative genomics and isotopic fractionation. Front Microbiol. 2014;5. doi:10.3389/fmicb.2014.00523

14. Schäfer H, McDonald IR, Nightingale PD, Murrell JC. Evidence for the presence of a CmuA methyltransferase pathway in novel marine methyl halide-oxidizing bacteria: Novel marine methyl halide-oxidizing bacteria. Environ Microbiol. 2005;7: 839–852. doi:10.1111/j.1462-2920.2005.00757.x

15. Harper DB, Hamilton JTG, Ducrocq V, Kennedy JT, Downey A, Kalin RM. The distinctive isotopic signature of plant-derived chloromethane: possible application in constraining the atmospheric chloromethane budget. Chemosphere. 2003;52: 433–436. doi:10.1016/S0045-6535(03)00206-6

16. Stacheter A, Noll M, Lee CK, Selzer M, Glowik B, Ebertsch L, et al. Methanol oxidation by temperate soils and environmental determinants of associated methylotrophs. ISME J. 2013;7: 1051–1064. doi:10.1038/ismej.2012.167

17. Neufeld JD, Dumont MG, Vohra J, Murrell JC. Methodological Considerations for the Use of Stable Isotope Probing in Microbial Ecology. Microb Ecol. 2007;53: 435–442. doi:10.1007/s00248-006-9125-x

18. Roselli S, Nadalig T, Vuilleumier S, Bringel F. The 380 kb pCMU01 plasmid encodes chloromethane utilization genes and redundant genes for vitamin B_{12}- and tetrahydrofolate-dependent chloromethane metabolism in *Methylobacterium extorquens* CM4: a proteomic and bioinformatics study. PLoS ONE. 2013;8: e56598. doi:10.1371/journal.pone.0056598

UNIVERSITÉ DE STRASBOURG

UNIVERSITÄT BAYREUTH

PhD thesis abstract and key words

Pauline CHAIGNAUD

Bacteria as chloromethane sinks – from model strains to forest soil communities

Chloromethane (CH_3Cl) is a volatile organic compound responsible for over 15% of stratospheric ozone degradation due to chlorinated compounds. It is mainly produced by living and decaying plants. Bacteria utilizing CH_3Cl as sole carbon and energy source for growth were shown to be involved in the filtering of CH_3Cl emissions to the atmosphere. This biological process remains to be quantified in the environment, especially for forest soil, a major CH_3Cl sink. The *cmuA* gene is used as a biomarker of bacterial CH_3Cl degradation in environmental studies. It encodes a CH_3Cl methyltransferase essential for bacterial growth by the *cmu* (chloromethane utilization) pathway for growth with CH_3Cl and the only one characterized so far. My thesis project had a double aim: i) In depth studies of CH_3Cl adaptation of a model methylotrophic bacterium, *Methylobacterium extorquens* strain CM4; ii) Exploration of bacterial CH_3Cl-utilizers in forest

An RNAseq study of strain CM4 has shown that growth with CH_3Cl leads to a difference of transcription of 137 genes in its 6.2 Mb genome compared to growth with methanol (CH_3OH). Among those, genes of the *cmu* pathway and other genes involved in the metabolism of essential cofactors for CH_3Cl utilization by this pathway, are all plasmid pCMU01-encoded. Paralogous genes located on the chromosome were not differentially expressed. On the other hand, other chromosomal genes potentially involved in extruding protons generated during CH_3Cl deshalogenation (*hppA*), $NADP^+$ regeneration (*pnt*), or in the cofactor tetrahydrofolate metabolism (*gcvPHT*) were differentially expressed.

The diversity of CH_3Cl-degrading bacteria in forest soil of the German natural park of Steigerwald was studied in microcosms using stable isotope probing. Microorganisms able to assimilate labeled [^{13}C]-CH_3Cl incorporate this heavy carbon isotope in their DNA. Sequence analysis of the PCR-amplified 16S RNA encoding gene from [^{13}C]-DNA fractions uncovered phylotypes of the genus *Methylovirgula* and of the order of the *Actinomycetales*, which were not associated with bacterial CH_3Cl degradation so far. In contrast, PCR-amplified sequences of *cmuA* and other genes of methylotrophic metabolism were closely related to known CH_3Cl-degrading isolates. These results suggest that bacteria containing genes of the *cmu* pathway acquired by horizontal gene transfer as well as bacteria lacking the *cmu* pathway contribute to biological filtering of CH_3Cl in forest soil. Future experiments coupling molecular and culture methods will aim to discover new CH_3Cl-degrading pathways and to characterize the abundance and diversity of CH_3Cl-degradation metabolism in soil and other environmental compartments.

Key words: chloromethane, bacterial dechlorination, methylotrophy, *Methylobacterium extorquens* CM4, bacterial communities, RNA sequencing, stable isotope probing, forest soil, microcosms

UNIVERSITÉ DE STRASBOURG

Pauline CHAIGNAUD

UNIVERSITÄT BAYREUTH

Le rôle des bactéries dans le filtrage du chlorométhane, un gaz destructeur de la couche d'ozone – des souches modèles aux communautés microbiennes de sols forestiers

Die Rolle von Bakterien als Filter für Chlormethan, ein Ozonschicht-schädigendes Gas – Von Modellorganismen zu mikrobiellen Gemeinschaften im Waldboden

Le chlorométhane (CH_3Cl) est un composé organique volatile responsable de plus de 15 % de la dégradation de l'ozone stratosphérique due aux composés chlorés. Il est produit majoritairement par les plantes vivantes ou en décomposition. Les bactéries capables d'utiliser le CH_3Cl comme source de carbone pour leur croissance peuvent jouer un rôle de filtre dans les émissions de CH_3Cl vers l'atmosphère. Ce processus biologique reste à quantifier dans l'environnement, notamment pour les sols forestiers considérés comme un puits majeur de ce composé. Dans les études environnementales, le gène *cmuA* est utilisé comme biomarqueur de la dégradation bactérienne du CH_3Cl. Il code une chlorométhane méthyltransférase essentielle à la croissance bactérienne avec le CH_3Cl par la voie *cmu* (pour chloromethane utilization), la seule caractérisée à ce jour. Mon projet de thèse avait un double objectif : i) l'approfondissement des connaissances de l'adaptation au CH_3Cl chez une bactérie méthylotrophe modèle, *Methylobacterium extorquens* CM4; ii) l'exploration de la diversité des bactéries CH_3Cl-dégradantes de sols forestiers.

L'étude RNAseq chez la souche CM4 a montré que la croissance avec le CH_3Cl s'accompagne de différences dans la transcription de 137 gènes de son génome (6.2 Mb) par rapport à sa croissance sur le méthanol (CH_3OH). Les gènes de la voie *cmu*, ainsi que d'autres gènes impliqués dans le métabolisme de cofacteurs essentiels à l'utilisation du CH_3Cl par cette voie et eux aussi portés par le plasmide pCMU01 de la souche, en font partie. Les paralogues de ces gènes localisés sur le chromosome ne sont quant à eux pas différentiellement exprimés. En revanche, d'autres gènes du chromosome, potentiellement impliqués dans l'excrétion de protons produits lors de la déshalogénation (*hppA*), la régénération du $NADP^+$ (*pnt*), ou le métabolisme du cofacteur tétrahydrofolate (gènes *gcvPHT*), le sont.

L'étude de la diversité des bactéries CH_3Cl-dégradantes de sol forestier de la réserve naturelle de Steigerwald (Allemagne) a été réalisée sur des microcosmes par une approche de « Stable Isotope Probing ». Les micro-organismes capables d'assimiler le CH_3Cl marqué au [^{13}C] incorporent cet isotope lourd du carbone dans leur ADN. L'analyse des séquences amplifiées par PCR des gènes codant l'ARN 16S des fractions d'ADN enrichies en [^{13}C] a permis de mettre en évidence de nouveaux phylotypes, du genre *Methylovirgula* et de l'ordre des *Actinomycetales*, distincts de ceux auxquelles les souches dégradant le CH_3Cl isolées jusqu'ici sont affiliées. En revanche, les séquences du gène *cmuA* et d'autres gènes du métabolisme méthylotrophe obtenues par PCR à partir de l'ADN enrichi en [^{13}C] sont très proches de celles des souches CH_3Cl-dégradantes connues. Les résultats obtenus suggèrent ainsi que des bactéries ayant acquis par transfert horizontal les gènes de dégradation de la voie *cmu* ou ne possédant pas de gène *cmuA* contribuent au filtrage biologique du CH_3Cl des sols forestiers. A l'avenir, le couplage de différentes méthodes moléculaires et des approches culturales visera à découvrir de nouvelles voies microbiennes de l'utilisation du CH_3Cl, et à caractériser l'abondance et la diversité des métabolismes impliqués dans la dégradation du CH_3Cl dans les sols et d'autres compartiments environnementaux.

Mots-clés : chlorométhane, déchloration bactérienne, méthylotrophie, *Methylobacterium extorquens* CM4, communautés bactériennes, RNA sequencing, stable isotope probing, sols forestiers, microcosmes

Chloromethane (CH_3Cl Chlormethan (CH_3Cl) ist eine volatile organische Verbindung (BVOC), die für mehr als 15% des stratosphärischen Ozonabbaus verantwortlich ist. Es wird vor allem von lebenden und toten Pflanzenteilen gebildet. Bakterien, die CH_3Cl als Kohlenstoff- und Energiequelle für ihr Wachstum nutzen, sind an der Reduktion der CH_3Cl -Emissionen in die Atmosphäre beteiligt. Dieser biologische Prozess wurde bislang in der Umwelt nicht quantifiziert, insbesondere in Waldböden, die eine wichtige biologische Senke darstellen. Das *cmuA*-Gen wird als Biomarker zur Detektion von bakterieller CH_3Cl -Abbau in der Umwelt eingesetzt. Es kodiert für eine CH_3Cl -Methyltransferase, die essentiell für Wachstums mittels des *cmu* (chloromethane utilization)-Stoffwechsel ist – der bislang einzig bekannte Abbau-Stoffwechselweg. Mein Promotionsprojekt verfolgte zwei Ziele: i) Hochauflösende Untersuchungen am Transkriptom zur Anpassung an die CH_3Cl -Nutzung im Model-Bakterium *Methylobacterium extorquens* Stamm CM4; ii) Erfassung von bakteriellen CH_3Cl -Nutzern in einem Waldboden.

In einer RNAseq-Studie am Stamm CM4 konnte gezeigt werden, dass Wachstum auf CH_3Cl zu unterschiedlichen Transkription von 137 Genen des 6.2 Mb-Genom im Vergleich mit Wachstum auf Methanol (CH_3OH) führte. Von diesen Genen sind alle Gene des *cmu*-Stoffwechselweges und die Gene zur Synthese essentieller Kofaktoren auf dem Plasmid pCMU01 kodiert. Paraloge, die sich auf dem Chromosom befinden, wurden nicht differentiell exprimiert. Andererseits wurden chromosomale Gene, die potentiell am Export von während der Dehalogenierung gebildeten Protonen (*hppA*), an der $NADP^+$-Regeneration (*pnt*) oder an der Synthese des Kofaktors Tetrathydrofolat (*gcvPHT*) beteiligt sind, differentiell exprimiert.

Im Zweiten Teil des Promotionsprojektes wurde die Biodiversität der CH_3Cl-abbauenden Bakterien in einem Waldboden im deutschen Naturpark Steigerwald mittels Stabiler-Isotopen-Beprobung in Mikrokosmen untersucht. Mikroorganismen, die [^{13}C]-CH_3Cl assimilierten, inkorporierten dieses stabile Isotop in ihrer DNA. Sequenz-Analysen markierter DNA anhand des Genmarkers 16S-rRNA-Gen ergaben Phylotypen der Gattung *Methylovirgula* und der Ordnung *Actinomycetales* - Bakterien, die bislang nicht mit CH_3Cl-Abbau in Zusammenhang gebracht werden konnten. Dahingegen waren die markierten Genotypen des Gens *cmuA* und anderer Gene aus dem C1-Stoffwechsel nah zu bekannten Genotypen bereits kultivierter Reinkulturen verwandt. Diese Ergebnisse legen nahe, dass Bakterien, die Gene des *cmu*-Stoffwechsels besitzen, diese durch horizontalen Gentransfer aquiriert haben bzw. dass es CH_3Cl-abbauende Bakterien gibt, die nicht den *cmu*-Stoffwechselweg besitzen. Zukünftige Untersuchungen, die molekulare und Kultivierungsmethoden kombinieren werden, sollen neue CH_3Cl-Abbau-Stoffwechselwege identifizieren und darüber hinaus die Abundanz und Diversität der CH_3Cl-Abbaustoffwechsel in Böden und anderen Umwelten aufklären.

Stichworte: Chlormethan, bakterielle Dechlorinierung, Methylotrophie, *Methylobacterium extorquens* CM4, bakterielle Gemeinschaften, RNA-Sequenzierung, Stabile-Isotopen-Beprobung, Waldboden, Mikrokosmen