## Title

Characterization of Caffeine Biodegradation and Transcriptional Regulation in Methylorubrum populi PINKEL



## Author

Subuyuj Hernandez, Gabriel Alberto

## Publication Date

2023

Peer reviewed|Thesis/dissertation

Characterization of Caffeine Biodegradation and Transcriptional Regulation in *Methylorubrum populi* PINKEL

By

Gabriel Alberto Subuyuj Hernandez

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Microbiology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Rebecca E. Parales, Chair

_____

Johan Leveau

_____

John C. Meeks

Committee in Charge

2023

## Abstract

A facultative pink-pigmented methylotroph, *Methylorubrum populi* PINKEL, was isolated from compost containing coffee grounds via a selective enrichment strategy using caffeine (1,3,7-trimethylxanthine) as the sole carbon, nitrogen, and energy source. In addition to caffeine, *M. populi* PINKEL was also able to grow on other methylxanthines, including theobromine (3,7-dimethylxanthine), paraxanthine (1,7-dimethylxanthine), 7-methylxanthine, and xanthine. Whole cell biodegradation assays in combination with HPLC analysis of spent media confirmed that *M. populi* PINKEL uses the *N*-demethylation pathway to degrade caffeine to a mixture of theobromine and paraxanthine. Both of these dimethylxanthines get converted to xanthine via two consecutive *N*-demethylation steps, with 7-methylxanthine as the intermediate.

*In silico* comparison of the *Pseudomonas putida* CBB5 *ndm* genes and the genome of *Methylorubrum populi* PINKEL helped identify genes that are predicted to encode enzymes of the *N*-demethylation pathway (amino acid percent identities ranging from 41% to 65%). The *ndm* genes in *M. populi* PINKEL are organized in two separate clusters: the first cluster is approximately 3 kb in length and contains the *ndmA* gene ($N_1$-demethylase; peg_1790); whereas the second cluster measures approximately 11.2 kb and contains the remaining four *ndm* genes: *ndmB* ($N_3$-demethylase; peg_1782), *ndmC* ($N_7$-demethylase; peg_1777), *ndmD* (reductase; peg_1781), and *ndmE* (peg_1776). Growth curve and HPLC analysis of each of the five Δ*ndm* mutants as well as the corresponding complemented strains helped determine not only which step of the pathway is affected by each individual *ndm* deletion, but that NdmA is responsible for the $N_1$ and $N_3$ demethylation of caffeine during the first step of the pathway.

Transcriptional promoter fusion assays identified the three promoter regions that participate in the transcriptional regulatory system of the five *ndm* genes, each upstream of the following three genes: *ndmA, ndmB,* and *ndmC*. The inability of a Δ*ndmR* (predicted to encode a LysR-type transcriptional regulator) mutant to upregulate expression from any of the three *ndm* promoter regions during inducing or un-inducing conditions indicates that the LysR protein positively regulates the *N*-demethylation pathway in *M. populi* PINKEL. Additionally, it was determined, that any of the intermediates of the *N*-demethylation pathway containing a methyl group can serve as an inducer of the pathway, although at different levels of induction.

This report describes research on the caffeine-degrading bacterium *Methylorubrum populi* strain PINKEL as a model to study caffeine biodegradation and the characterization of the transcriptional regulatory system for caffeine-degrading enzymes.

## Acknowledgments

A work of this magnitude requires assistance and support from many people. I am deeply grateful to all of those listed here and to anybody that I might have forgotten. First, I would like to thank my thesis advisor, Dr. Rebecca Parales, for accepting me into her lab and allowing me to research other aspects beyond chemotaxis. Her knowledge of microbiology, biochemistry, and molecular biology was a vast resource from which I was able to draw ideas countless times over the past six years. Becky was very patient with me and was very supportive in my goals beyond research.

Thank you, Juan Parales, for always coming back to the lab to help me with the HPLC machine and any other instrumentation help I needed despite already being retired. Thank you, Dr. Rebecca Schomer and Dr. Miguel Valderrama, for always bouncing ideas with me whenever I found speedbumps in my research. You both were always a source of encouragement. Thank you to all of the undergraduate students including Alycia Rasmussen, Mar De La Cruz, Mia Karlsson, and Anham Rafique, who helped me make significant progress in research and develop myself as a mentor. The truth is that without your help, I would not have finished my PhD in a timely manner.

I would like to thank John (Jack) Meeks and Johan Leveau for serving in my dissertation committee along with Becky Parales. Your advice and constant feedback during the annual dissertation meetings taught me how to ask questions and communicate scientific findings in an appropriate manner.

I would also like to thank Dr. Elizabeth Skovran. You were the first person to believe in me as a scientist. You gave me the opportunity to join your lab despite not

having any experience. You also played a crucial role in convincing me that I was PhD material, even when I could not see it.

Finally, I would like to express my appreciation to my family for their support. My parents, Jose and Maria Subuyuj, were very supportive in my journey as a scientist. Even though you did not understand what I was doing in the lab, you always encouraged me to work hard and never give up, even when things got hard. You both sacrificed many things and worked hard so that I could have the best education I could possibly have. Words can't express how much I love you and how much I appreciate all your hard work. Thank you to my brother Edwin Subuyuj for your constant support. You may not know this, but you set a good example for me, despite being younger. You were the first to start and finish graduate school and that really motivated me. I really admire you! Thank you to my mother-in-law (Kristen Criswell) and my father-in-law (Steven Criswell) for entrusting me with your daughter and always expressing how proud you are of me. I really hope I can make you proud of the way I treat and love your daughter. To the rest of the Subuyuj, Hernandez, Criswell, and Cox family, thank you so much for your constant words of encouragement.

I would also like to thank my wife, Samantha Subuyuj, and my daughters Evelyn Grace and Olivia Joy Subuyuj. I can't find the words to express how deeply fortunate I am to have you in my life. My baby girls were always willing to lend their daddy to the lab and always greeted my arrival home with a smile and hugs. My wonderful wife never complained about the long hours that I sometimes had to work. She has been my support for the last six years and I am looking forward to spending the rest of eternity with her. Last but not least, I thank Jesus Christ, my Lord and savior for all the wisdom

and guidance through this academic journey and always bringing the right people at the right time. Thank you to all who have helped me get to this point. It has certainly been worth the effort.

## Table of Contents

## List of Figures

## List of Tables

# Chapter 1

# Introduction

## Xanthine alkaloids: characteristics, natural distribution, and significance

### Characteristics of xanthine alkaloids

Methylxanthines and methyluric acids belong to a diverse family of compounds referred to as xanthine alkaloids (a subset of the purine alkaloid family). These compounds are naturally synthesized by nearly 100 plant species as secondary metabolites and are derived from purine nucleotides [Ashihara & Crozier, 1999]. Structurally, xanthine alkaloids are composed of a xanthine backbone, a heterocyclic organic compound built from coupled pyrimidinedione and imidazole rings [Talik et al., 2012], with different R-groups attached at the first, third, and seventh positions of the molecule (Figure 1.1A). In the case of caffeine, each of the three R-groups is a methyl group and as individual methyl groups are removed from the xanthine backbone, different methylxanthines are generated (Table 1.1). Methyluric acids have a similar structure to methylxanthines with the difference being the oxidized carbon located at the eighth position of the xanthine ring (Figure 1.1B).

Caffeine (1,3,7-trimethylxanthine), theobromine (3,7-dimethylxanthine), and theophylline (1,3-dimethylxanthine) are considered the most relevant methylxanthines as they are major components of widely consumed beverages and foods, such as coffee, tea, and chocolate [Franco et al., 2013]. Other naturally occurring methylxanthines that exist as transient byproducts of plant xanthine alkaloid metabolism include paraxanthine (1,7-dimethylxanthine), 1-methylxanthine, 3-methylxanthine and 7-methylxanthine [Algharrawi, K.H.R. et al., 2015]. The naturally occurring methyluric acids commonly found in tea plants include 1,3,7-trimethyluric acid, theacrine, methylliberine, and liberine (Table 1.1).

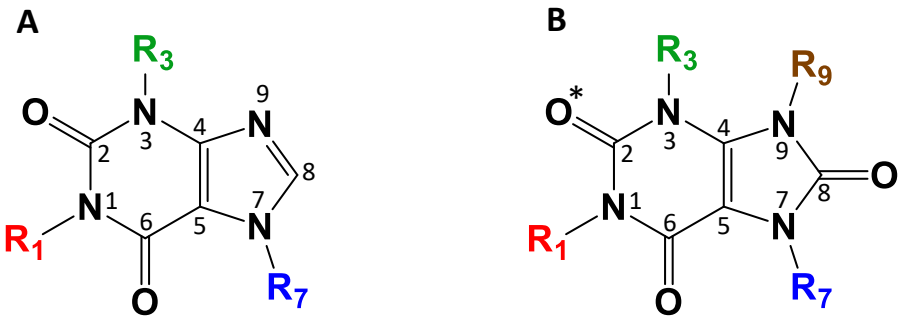



**Figure 1.1**: General structure of (A) alkylxanthines and (B) alkyluric acids. The asterisk indicates the oxygen that gets methylated in some methyluric acids.

**Table 1.1**: Naturally occurring Methylxanthines and methyluric acids.

| **methylxanthines** | **Trivial Name** | $R_1$ | $R_3$ | $R_7$ | | |
|---|---|---|---|---|---|---|
| Xanthine | - | H | H | H | | |
| 1-methylxanthine | - | $CH_3$ | H | H | | |
| 3-methylxanthine | - | H | $CH_3$ | H | | |
| 7-methylxanthine | Heteroxanthine | H | H | $CH_3$ | | |
| 3, 7-dimethylxanthine | Theobromine | H | $CH_3$ | $CH_3$ | | |
| 1, 7-dimethylxanthine | Paraxanthine | $CH_3$ | H | $CH_3$ | | |
| 1, 3-dimethylxanthine | Theophylline | $CH_3$ | $CH_3$ | H | | |
| 1, 3, 7-trimethylxanthine | Caffeine | $CH_3$ | $CH_3$ | $CH_3$ | | |
| **Methyluric acids** | **Trivial Name** | $R_1$ | $R_3$ | $R_7$ | $R_9$ | **O*** |
| Uric acid | - | H | H | H | H | - |
| 1, 3, 7-trimethyluric acid | - | $CH_3$ | $CH_3$ | $CH_3$ | H | - |
| 1, 3, 7, 9-tetramethyluric acid | Theacrine | $CH_3$ | $CH_3$ | $CH_3$ | $CH_3$ | - |
| O(2), 1, 9-trimethyluric acid | Liberine | $CH_3$ | H | H | $CH_3$ | $CH_3$ |
| O(2),1,7,9-tritramethyluric acid | Methylliberine | $CH_3$ | H | $CH_3$ | $CH_3$ | $CH_3$ |

## Natural distribution of xanthine alkaloids

Xanthine alkaloids are naturally produced by at least 100 plant species across 13 different orders within the plant kingdom (Figure 1.2) [Ashihara et al., 2020]. These plants grow better in areas of the world with tropical and subtropical climates [Baumann and Frishknecht, 1988] and include coffee (*Coffea* species), tea (*Camellia* species),

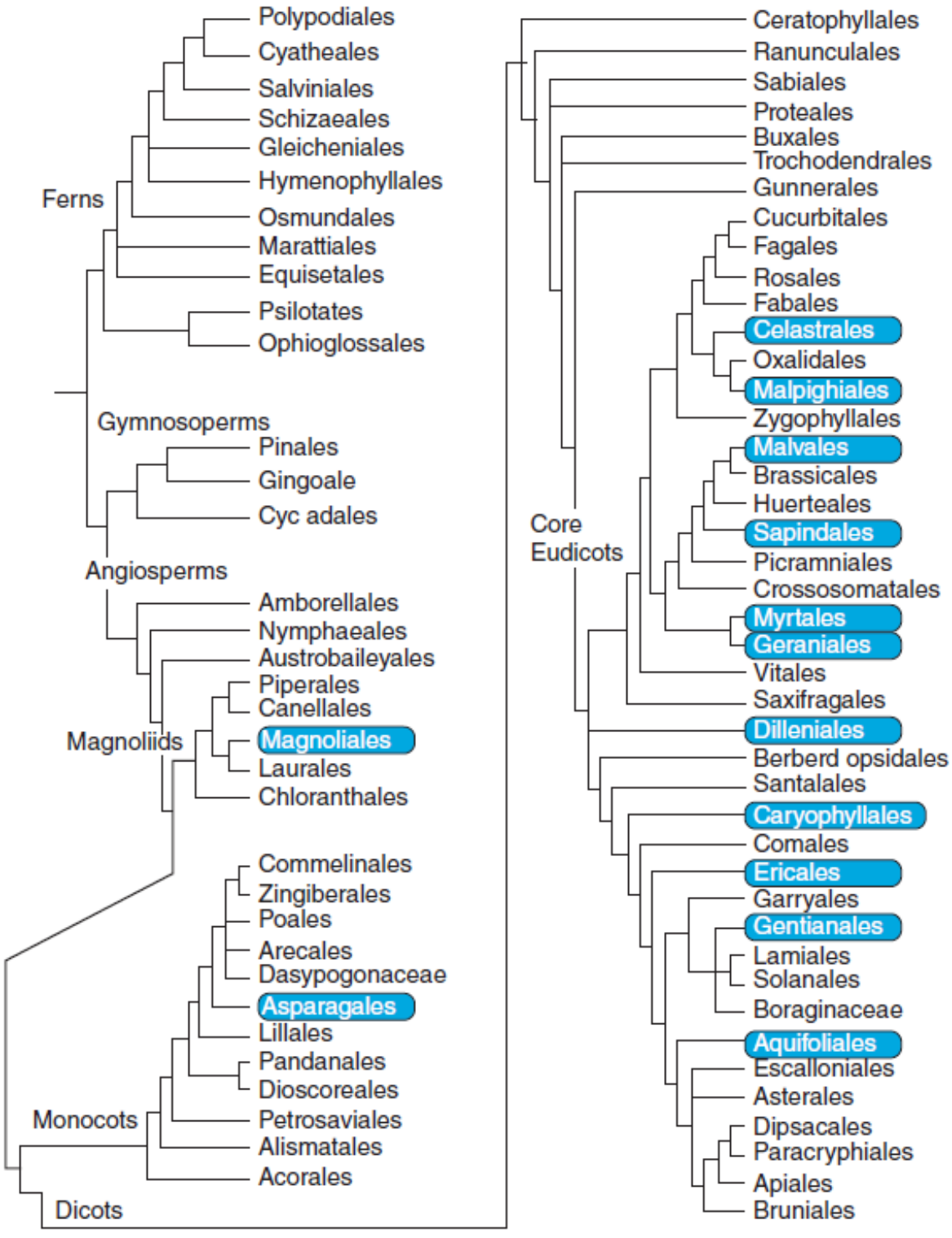



**Figure 1.2**: Phylogenetic relationship of xanthine alkaloid producing plant orders. The blue boxes show the orders with species that produce caffeine and related xanthine alkaloids.

Source: Ashihara et al., 2020

maté (*Ilex paraguariensis*), cocoa (*Theobroma cacao*), and guaraná (*Paullinaia cupana*) as summarized in Table 1.2 [Ashihara et al., 2008; Ashihara et al., 2020]. Caffeine and theobromine are the most abundant xanthine alkaloids observed in these plants, although paraxanthine, theophylline and methyluric acids such as theacrine are also present at lower concentrations [Ashihara et al., 2017]. Regardless of the xanthine alkaloid produced by the plant, synthesis occurs mainly in young leaves, seeds, and flowers [Ashihara et al., 2008].

**Table 1.2**: Occurrence of xanthine alkaloids in plants

| **Order** | ***Latin Name*** | **Common Name** | **Alkaloid** |
|---|---|---|---|
| Magnoliales | *Annona cherimola* | Cherimoya | Cf |
| Asparagales | *Scilla maritima* | Red squill | Cf |
| Celastrales | *Maytenus sp.* | Maytenus | Cf |
| Malpighiales | *Banisteriopsis caapi* | Caapi | Cf |
| Malvales | *Cola acuminata* | Kola nut | Cf>Tb |
| | *Herrania purpurea* | Monkey cocoa | Tc |
| | *Theobroma cacao* | Cacao (cocoa) | Tb>Cf |
| Sapindales | *Citrus paradisi* | Grapefruit | Cf>Tp>Tb>Px |
| | *Paullinaia cupana* | Guaraná | Cf>Tb>Tp |
| Myrtales | *Combretum jacquinii* | Red bushwillow | Cf |
| Geraniales | *Erodium cicutarium* | Pinweed | Cf |
| Dilleniales | *Davilla rugosa* | Fire vine | Cf |
| Caryophallales | *Cereus jamacaru* | Cactus | Cf |
| Ericales | *Camellia sinensis* | Tea | Cf |
| | *Camellia assamicaa* | Kucha | Tc>Cf |
| Gentinales | *Coffea arabica* | Coffee (arabica) | Cf |
| | *Coffea canephora* | Coffee (robusta) | Cf |
| Aquifoliales | *Ilex paraguariensis* | Maté | Cf |
| | *Villaresia mucronata* | Chilean citronella tree | Cf |

Cf = caffeine, Tb = theobromine, Tp = theophylline, Px = paraxanthine, Tc = theacrine

## Ecological roles of xanthine alkaloids

The physiological role of xanthine alkaloids *in planta* is still an ongoing area of study. Nonetheless, it seems that they do not have a role as nitrogen reserves since a great amount of xanthine alkaloids, which contain four nitrogen atoms per molecule, remain in the leaves after their natural detachment from the plant [Ashihara et al., 2011]. There are two hypotheses that have been proposed and tested concerning the ecological roles of xanthine alkaloids in plants: the chemical defense theory and the allelopathic theory.

The chemical defense theory proposes that caffeine is produced at high concentrations in new plant tissues, leaves, fruits, and flower buds to protect against phytopathogens and insects [Nathanson, 1984; Kim and Sano, 2008]. Work done by Hollingsworth et al. [2002] in which tomato plants were treated with a 1% caffeine solution showed that xanthine alkaloids act as natural pesticides and reduce feeding by tobacco hornworms. Similarly, treating cabbage leaves with caffeine killed or repelled snails and slugs by acting as a neurotoxin [Hollingsworth et al., 2003]. Aneja and Gianfagna [2001] also showed that caffeine is synthesized in cacao leaves in response to infections to inhibit the growth and proliferation of bacterial and fungal phytopathogens beyond the site of infection. The caffeine doses that affect the development of other tested organisms are summarized in Table 1.3.

The allelopathic theory proposes that caffeine is released into soil to inhibit the germination and growth of neighboring plants [Waller, 1989; Tanti et al., 2016]. When a seed first germinates, the primary root called the radicle grows downward to anchor the seedling to the soil. However, it has been reported that elongation of the root system

from the initial radicle of lettuce seedlings [Chou and Waller, 1980], rice seedlings [Smyth, 1992], and some leguminous plants [Rizvi et al., 1981] is affected by 50 to 90% in the presence of xanthine alkaloids. Additionally, Friedman and Waller [1983] reported an autotoxicity effect where the growth of young coffee plants is completely inhibited when planted in long established caffeine plantations where the concentration of caffeine in the soil is high.

**Table 1.3**: Caffeine doses affecting the growth and development of selected organisms

| Organism | Species | Dose (%) | Reference |
|---|---|---|---|
| Bacterium | *Pseudomonas syringae* | 0.04 | Kim & Sano, 2008 |
| Fungus | *Crinipellis perniciosa* | 0.05 | Aneja & Gianfagna, 2001 |
| | *Monacrosporium ambrosium* | 0.05–0.5 | Savitri et al., 1995 |
| | *Monacrosporium ambrosium* | 0.5–1.0 | Tsubouchi et al., 1985 |
| Insect (larvae) | *Pieris rapae* (small white butterfly) | 0.05 | Kim & Sano, 2008 |
| | *Manduca sexta* (tobacco horn moth) | 0.03–10 | Nathanson, 1984 |
| | *Vanessa cardui* (painted lady butterfly) | 0.1–0.3 | Nathanson, 1984 |
| | *Tenabrio* ssp (mealworm) | 0.1–0.3 | Nathanson, 1984 |
| | *Oncopeltus fasciatus* (milkweed bug nymph) | 0.3 | Nathanson, 1984 |
| | *Culex* spp. (mosquito) | 0.0007 | Nathanson, 1984 |
| | *Tribolium confusum* (confused flour beetle) | 0.2 | Nathanson, 1984 |
| | *Tribolium castaneum* (red flour beetle) | 0.2 | Nathanson, 1984 |
| Mollusk | *Zonitoides arboreus* (orchid snail) | 0.1 | Hollingsworth et al., 2002 |
| | *Veronicella cubensis* (slug) | 0.01–0.1 | Hollingsworth et al., 2002 |
| Plant | *Amaranthus spinosum* | 0.06–0.12 | Frischnecht & Baumann, 1985 |
| | *Lactuca sativa* (lettuce) | 0.03 | Chou & Waller, 1980 |
| | *Lactuca sativa* (lettuce) protoplasts | 0.01–0.02 | Chou & Waller, 1980 |

Source: Kim et al., 2010

## Caffeine metabolism: biosynthesis and degradation pathways

Biosynthesis of caffeine and other xanthine alkaloids

Xanthine alkaloids are plant derived secondary metabolites synthesized from purine nucleotides [Zulak et al., 2006]. These molecules also serve key cellular roles including precursors for nucleic acids, energy sources, components of major co-enzymes, activated intermediates in several biosynthetic pathways, and metabolic regulators [Henderson and Paterson, 1973; Stryer, 1996; Ashihara and Crozier, 1999]. The pathway(s) by which xanthine alkaloids are synthesized remained a topic of speculation until a series of studies on *in situ* metabolism of radiolabeled precursors and the characterization of key enzymes and genes helped elucidate the biosynthetic pathway in coffee (*Coffea arabica)* and tea (*Camellia sinensis*) [Ashihara et al., 1996a; Kato et al., 1999; Ashihara et al., 2011]. The biosynthetic pathway for xanthine alkaloids from purine nucleotides can be divided in two major stages: 1) the synthesis of xanthosine via *de novo* (synthesis from small molecules) and salvage (reutilization of purine nucleotides) pathways (Figure 1.3a) and 2) the conversion of xanthosine to caffeine (Figure 1.3b).

Production of xanthosine can be achieved by at least five different routes: the *de novo* purine biosynthesis route, the adenine monophosphate (AMP) route, the *S*-adenosyl-L-methionine (SAM) cycle route, the nicotinamide adenine dinucleotide (NAD) route, and the guanine monophosphate (GMP) route (Figure 1.3a).

i. *De novo* purine biosynthetic route: Inosine 5'-monophosphate (IMP) is considered the end product the *de novo* route (Figure 1.4). Plants produce IMP *de novo* by assimilating the following non-purine small molecules onto 5-phosphoribosyl-1-

pyrophosphate (PRPP) at specific sites: $HCO_3^-$ (C-6), 10-formyltetrahydrofolate (C-2 and C-8), glycine (C-4, C-5, and *N*-7), aspartate (*N*-1), and glutamine (*N*-3 and *N*-9) [Shemin and Rittenberg, 1947; Buchanan et al., 1948; Anderson and Gibbs, 1962; Proiser and Serenkov, 1963; Stryer, 1995]. Once IMP is generated, it is converted to xanthosine 5'-monophosphate (XMP) by IMP dehydrogenase and XMP is then converted to xanthosine by 5'-nucleotidase [Ashihara et al., 2008].

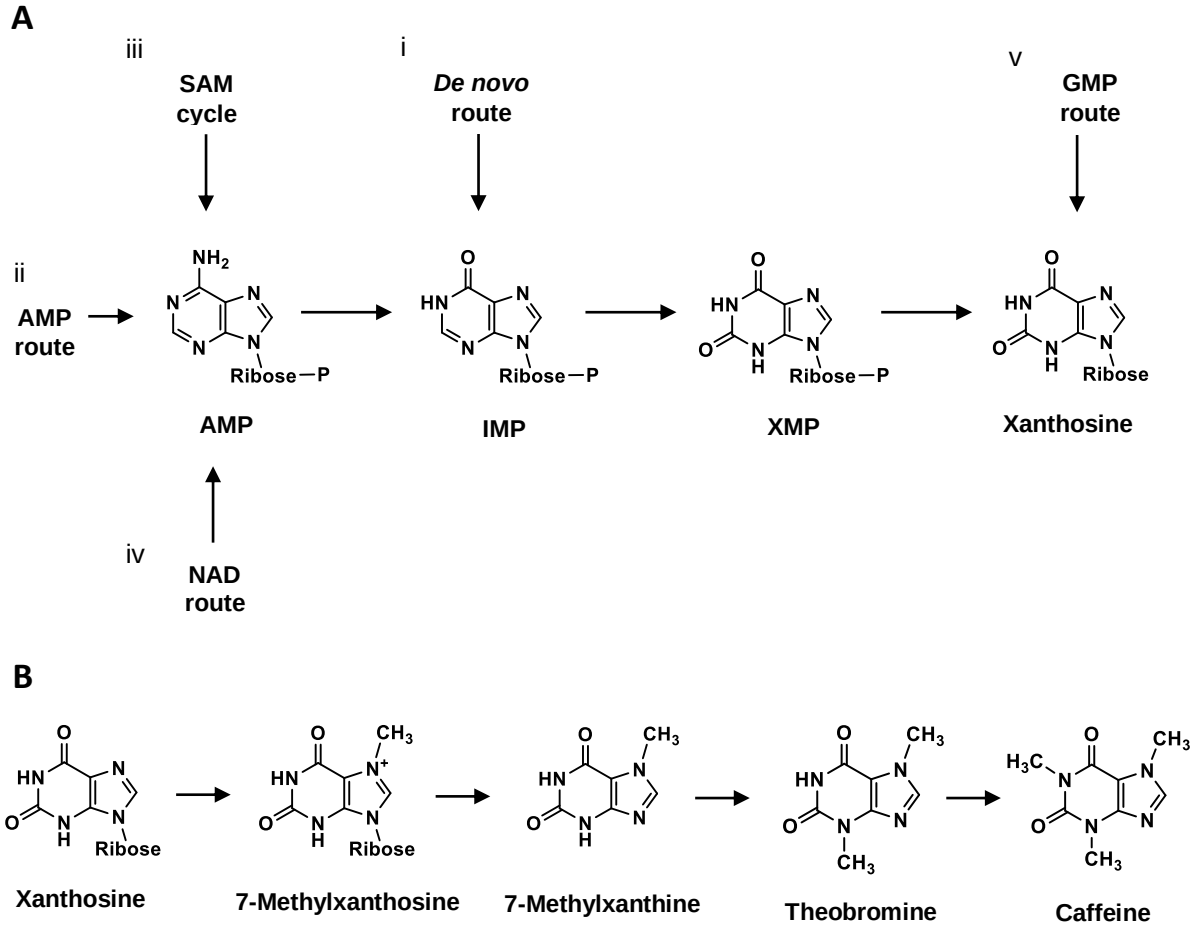



**Figure 1.3**: Xanthine alkaloid biosynthetic pathway (A) stage one and (B) stage two. Abbreviations are as follows: AMP, adenosine 5'-monophosphate; IMP, inosine 5'-monophosphate; XMP, xanthosine 5'-monophosphate; GMP, guanosine 5'-monophosphate; SAM, *S*-adenosyl-L-methionine; NAD, nicotinamide adenine diphosphate.

Source: Adapted from Ashihara et al., 2008

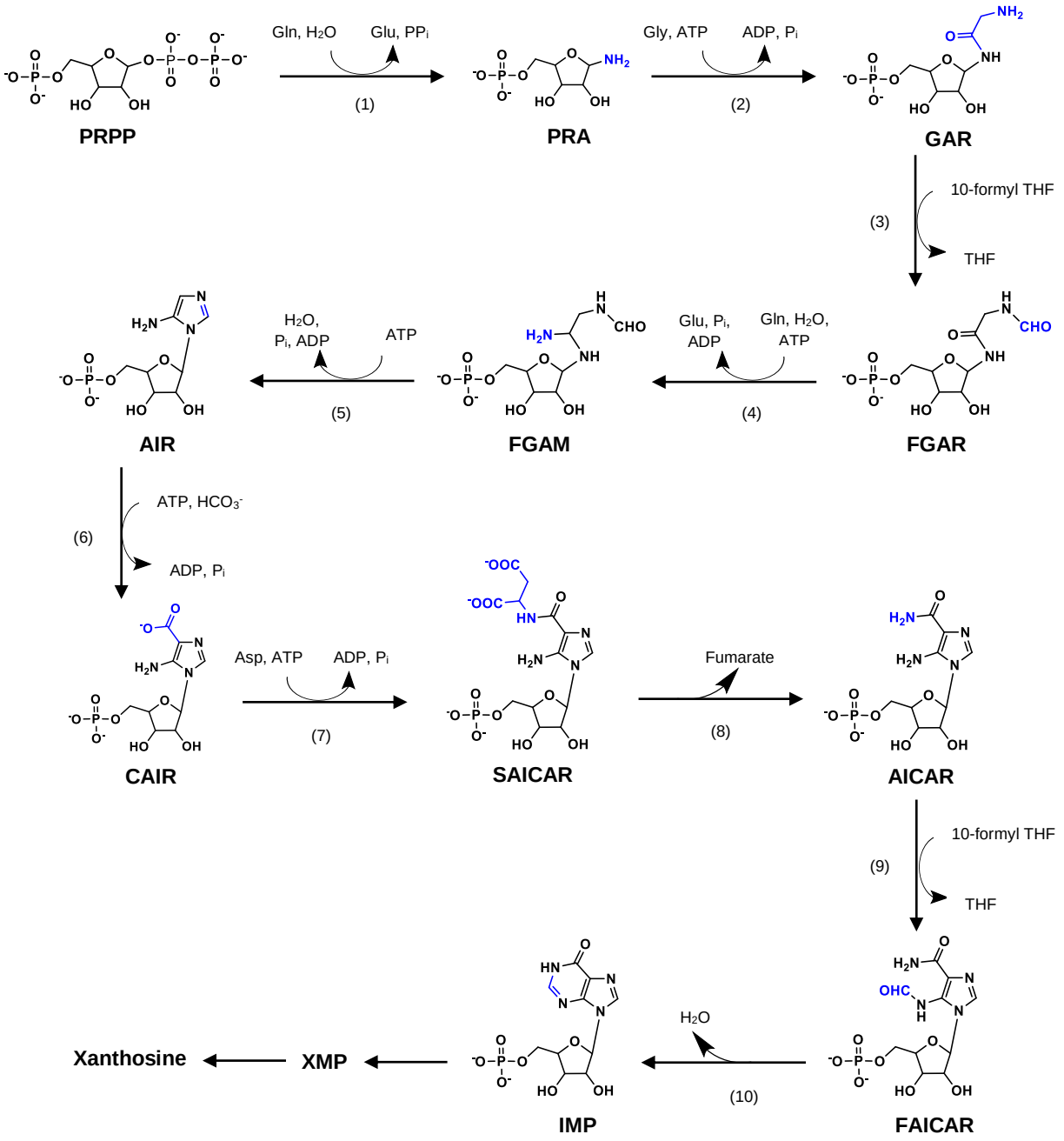



**Figure 1.4**: The *de novo* route to generate xanthosine for caffeine synthesis is initiated from 5-phosphoribosyl-1-pyrophosphate (PRPP). Abbreviations of metabolites are as follows: PRA, 5-phosphoribosyl amine; GAR, glycineamide ribonucleotide; FGAR, formylglycineamide ribonucleotide; FGAM, formylglycineamide ribonucleotide; AIR, 5-aminoimidazole ribonucleotide; CAIR, 5-aminoimidazole 4-carboxylate ribonucleotide; SAICAR, 5-aminoimidazole-4-N-succinocarboxyamide ribonucleotide; AICAR, 5-aminoimidazole-4-carboxyamide ribonucleotide; FAICAR, 5-formamidoimidazole-4-carboxyamide ribonucleotide; IMP, inosine 5′-monophosphate; XMP, xanthosine 5'-monophosphate; Asp, aspartate; Gln, glutamine; Gly, glycine; 10-THF, 10-formyltetrahydrofolate. The participating enzymes are (1) PRPP aminotransferase; (2) GAR synthetase; (3) GAR formyl transferase; (4) FGAM synthetase; (5) AIR synthetase; (6) AIR carboxylase; (7) SAICAR synthetase; (8) adenylosuccinate lyase; (9) AICAR formyl transferase; (10) IMP cyclohydrolase.

Source: adapted from Ashihara et al., 2017

ii. Adenosine monophosphate (AMP) route: Some of the xanthosine used for caffeine biosynthesis is produced via adenine salvage pathways (Figure 1.5) where adenine, adenosine, or deoxyadenosine serve as the initial substrates [Ashihara et al., 2018]. These purine bases are supplied by the intracellular breakdown of nucleic acids and nucleotides, as well as reactions that release adenine or adenosine [Ashihara et al., 2020]. When deoxyadenosine is the initial substrate, it is hydrolyzed to generate adenine by an adenosine nucleosidase. Then the 5-phosphoribosyl group from PRPP

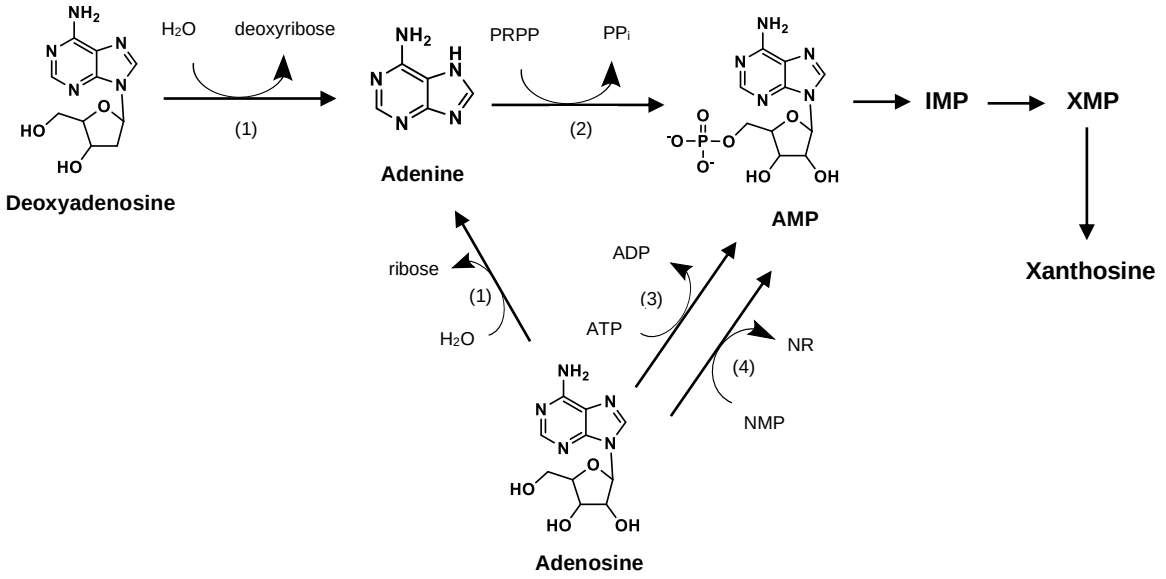



**Figure 1.5**: The adenosine 5'-monophosphate (AMP) route to generate xanthosine for caffeine synthesis. Abbreviations include IMP, inosine-5′-monophosphate; XMP, xanthosine 5'-monophosphate; PRPP, 5-phosphoribosyl-1-pyrophosphate. The participating enzymes are (1) nucleosidase; (2) adenine phosphoribosyltransferase; (3) adenosine kinase; (4) nucleoside monophosphate phosphotransferase.

Source: Adapted from Ashihara et al., 2020

is transferred to adenine by a phosphoribosyltransferase to generate AMP. On the other hand, adenosine can be directly converted to AMP by adenosine kinase and a nucleotide phosphotransferase, or via a two-step process in which adenosine is first converted to adenine by adenosine nucleosidase and then to AMP by adenine phosphoribosyltransferase. Once AMP is generated, it is converted to IMP by AMP deaminase. IMP produced via this route follows the same IMP → XMP → xanthosine pathway as described in the *de novo* route [Ashihara et al., 2008].

iii. *S*-adenosyl-L-methionine (SAM) cycle route: The second stage of the pathway (described in more detail below) involves three methylation steps in which SAM is the methyl donor. As the methyl groups are transferred from SAM to xanthine alkaloids by N-methyltransferases, *S*-adenosyl-L-homocysteine (SAH) is generated (Figure 1.6). SAH is then hydrolyzed to form L-homocysteine and adenosine by SAH hydrolase [Koshiishi, et al., 2001]. Adenosine is either converted to AMP via a two-step process in which adenine serves as the intermediate or directly to AMP by adenosine kinase [Koshiishi, et al., 2001]. Once AMP is produced, the AMP → IMP → XMP → xanthosine route is followed.

iv. Nicotinamide adenine dinucleotide (NAD) route: In this route, the enzyme NAD pyrophosphatase breaks down NAD to produce nicotinamide ribonucleotide and AMP (Figure 1.7) [Ashihara, 2015]. The generated AMP then follows the AMP → IMP → XMP → xanthosine route catalyzed by the same enzymes described in previous routes.

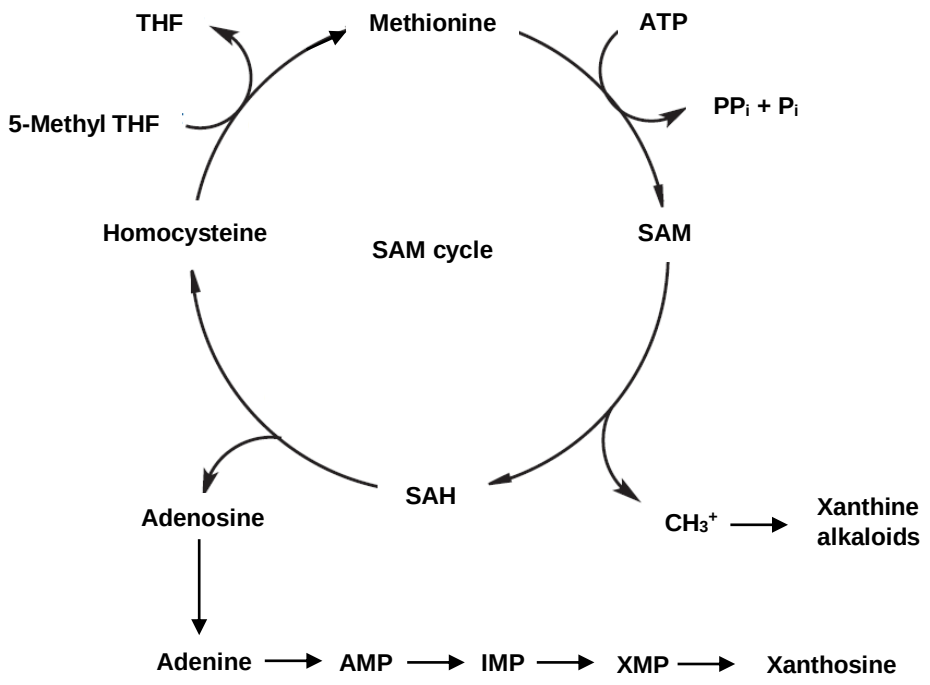



**Figure 1.6**: The *S*-adenosyl-L-methionine (SAM) cycle route to generate xanthosine for caffeine synthesis. Abbreviations include SAH, S-adenosyl-L-homocysteine; THF, tetrahydrofolate; AMP, adenosine 5'-monophosphate; IMP, inosine-5′-monophosphate; XMP, xanthosine 5'-monophosphate.

Source: Adapted from Ashihara et al., 2020

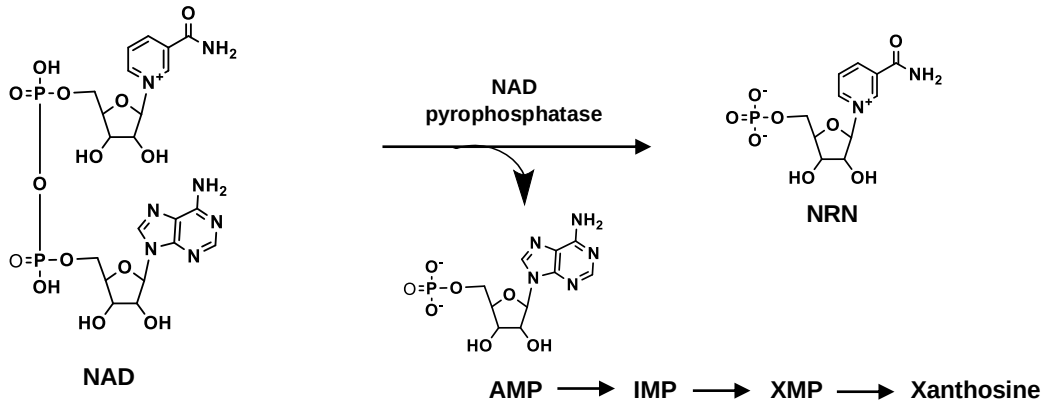



**Figure 1.7**: Nicotinamide adenine dinucleotide (NAD) route to generate xanthosine for caffeine synthesis. Abbreviations include AMP, adenosine 5'-monophosphate; IMP, inosine-5′-monophosphate; XMP, xanthosine 5'-monophosphate; NRN, nicotinamide ribonucleotide.

v. Guanosine monophosphate (GMP) route: A fraction of the xanthosine produced for caffeine biosynthesis comes from guanine salvage pathways (Figure 1.8). Here deoxyguanosine and guanine are converted to GMP by similar enzymatic reactions as seen in the AMP route (Figure 1.5). Once GMP is produced, it is converted to guanosine by a 5'-nucleotidase and then to xanthosine by guanosine deaminase [Negishi et al., 1994; Ashihara, 1997].

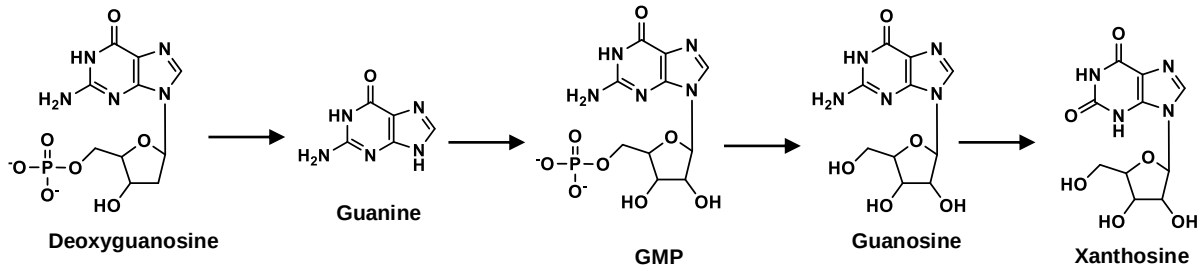



**Figure 1.8**: The guanosine 5'-monophosphate (GMP) route to generate xanthosine for caffeine synthesis.

The second stage of the pathway involves four steps to convert xanthosine to caffeine: three methylation steps and one nucleosidase reaction (Figure 1.3b). In the first step of this stage, xanthosine is converted to 7-methylxanthosine by a specific N-7 methyltransferase called 7-methylxanthine synthase [Negishi et al., 1985; Mizuno et al., 2003]. The ribose sugar is then cleaved from 7-methylxanthosine by methylxanthine nucleosidase to generate 7-methylxanthine. This molecule is further *N*-methylated by theobromine synthase at the N-3 position to generate theobromine [Negishi et al., 1988]. Caffeine is generated by a final N-1 methylation step catalyzed by caffeine synthase [Baumann et al., 1983; Fujimori et al., 1991]. There are other minor routes for caffeine synthesis that have been detected resulting in the transient production of 7-

methylxanthosine 5'-monophosphate (7-methyl-XMP), xanthine, 3-methylxanthine, paraxanthine, and theophylline [Ashihara and Crozier, 1999].

For most xanthine alkaloid-producing plants, caffeine seems to be the end product; however, there is a limited number of plant species, such as *Camellia assamica* var *kucha*, that oxidize caffeine to generate 1,3,7-trimethyluric acid as a starting point to generate other methyluric acids [Petermann and Baumann, 1983; Zheng et al., 2002]. Subsequent methylation and demethylation steps help generate the other three commonly observed methyluric acids in plants: theacrine, methylliberine and liberine. Although none of the enzymes that catalyze the biosynthesis of these methyluric acids have been identified, feeding experiments using $^{14}C$-labeled xanthine alkaloids suggest that caffeine is converted to the different methyluric acids following the pathway shown in Figure 1.9 [Zheng et al., 2002].

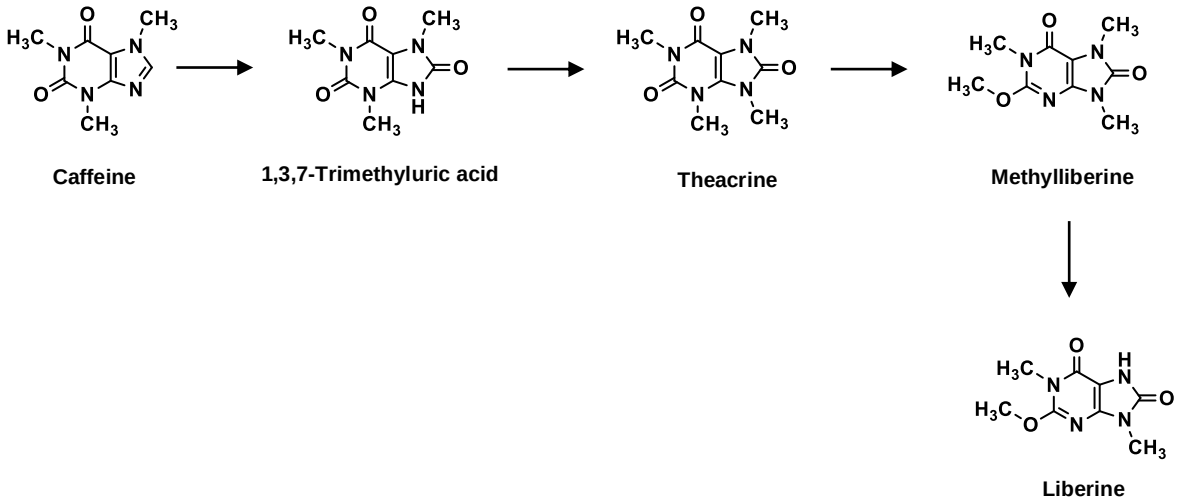



**Figure 1.9**: Suggested biosynthesis pathway of methyluric acids in plants. Enzymes participating in the metabolic steps have not been identified yet.

Source: Ashihara et al., 2020.

## Catabolism of methylxanthines in plants

Plants produce caffeine in young tissues as well as in actively growing parts such as leaves, buds, and fruits. Simultaneously, but at a much slower rate, caffeine can be degraded to xanthine via two pathways: the *N*-demethylation or the C-8 oxidation pathway [Ashihara et al., 2008]. The *N*-demethylation pathway (Figure 1.10a) is the predominant pathway for caffeine degradation in many xanthine alkaloid-producing plants (Table 1.2) in which caffeine is the major xanthine alkaloid generated. However, some plants, such as *Camellia assamicaa* (Kucha), follow the C-8 oxidation pathway (Figure 1.10b).

The plant *N*-demethylation pathway begins with the conversion of caffeine to theophylline by an $N_7$-demethylase. This step has been experimentally confirmed as the rate-limiting step for caffeine degradation and helps explain why caffeine is the major xanthine alkaloid that accumulates in different plant organs [Ito et al., 1997]. In the second step, the methyl group located at the first position gets removed from theophylline to generate 3-methylxanthine, which is further *N*-demethylated to produce xanthine in the final step [Peterman and Baumann, 1983; Suzuki and Waller, 1984; Ashihara et al., 1996b]. This *N*-demethylation pathway has alternative routes (dashed arrows in Figure 1.10a) that generate theobromine, 1-methylxanthine, and 7-methylxanthine as intermediates. Once xanthine is produced, it is oxidized to form uric acid, which is degraded to $CO_2$ and $NH_3$ via the purine catabolic pathway [Stasolla et al., 2003].

Some plants, such kucha and some *Coffea* species, follow the C-8 oxidation pathway in which the carbon located at the eighth position of caffeine is oxidized in the

first step to produce 1,3,7-trimethyluric acid (TMU). Subsequently, TMU is methylated at the N-9 position to generate theacrine, which is then converted all the way to liberine with methylliberine as an intermediate (Figure 1.10b) [Petermann et al., 1983].

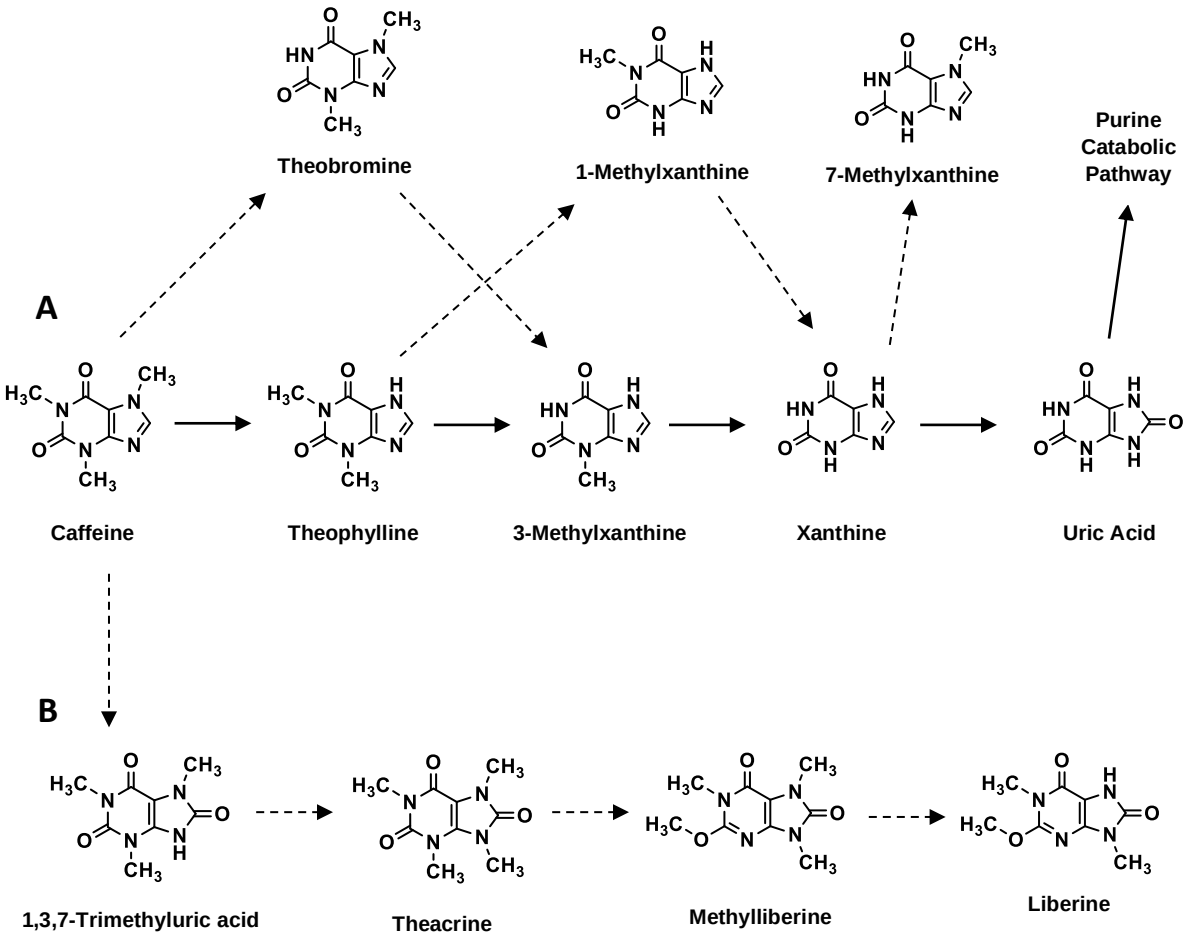



**Figure 1.10**: Caffeine catabolic pathway in plants. (A) *N*-demethylation pathway (B) C-8 oxidation pathway. Solid arrows represent the major degradation route, while dashed arrows represent alternative degradation steps.

Source: Adapted from Ashihara et al., 2008

## Catabolism of methylxanthines in mammals

Unlike plants, mammals, humans, primates, birds and many reptiles can only partially degrade xanthine alkaloids derived from caffeine and excrete urate or allantoin as an end product [Vogels et al., 1976]. Humans use hepatic cytochrome P450 enzymes such as CYP1A1, CYP1A2, CYP2E1 and CYP2A6 to catabolize caffeine [Campbell et al., 1987; Berthau et al., 1991]. These enzymes have been shown to catalyze *N*-demethylation and C-8 oxidation of caffeine and related methylxanthines [Tassaneeyakul et al., 1994]. More than 25 metabolites of caffeine have been identified in humans, including paraxanthine (80%), theobromine (10%), theophylline (4%) and others (6%) [Carrillo and Benitez, 2000]. While paraxanthine is the major metabolite in humans and rabbits, theophylline has been observed as the major metabolite in monkeys [Berthau et al., 1992]. The major catabolic pathway of caffeine in humans (Figure 1.11) involves two sequential *N*-demethylation steps at the N-3 and N-7 positions to generate paraxanthine and 1-methylxanthine, respectively [Carrillo et al., 2000]. There are other minor routes that result in the generation of theophylline, theobromine, 7-methylxanthine, and 3-methylxanthine. These methylxanthines are further oxidized to their corresponding methyluric acids to be excreted as part of urine (dashed arrows in Figure 1.11) [Carrillo and Benitez, 2000].

## Catabolism of methylxanthines in yeast and fungi

The wide consumption of caffeine and related xanthine alkaloids via various food products and pharmaceuticals has made such compounds ubiquitous in the environment. Thus, it comes as no surprise that caffeine-degrading microorganisms,

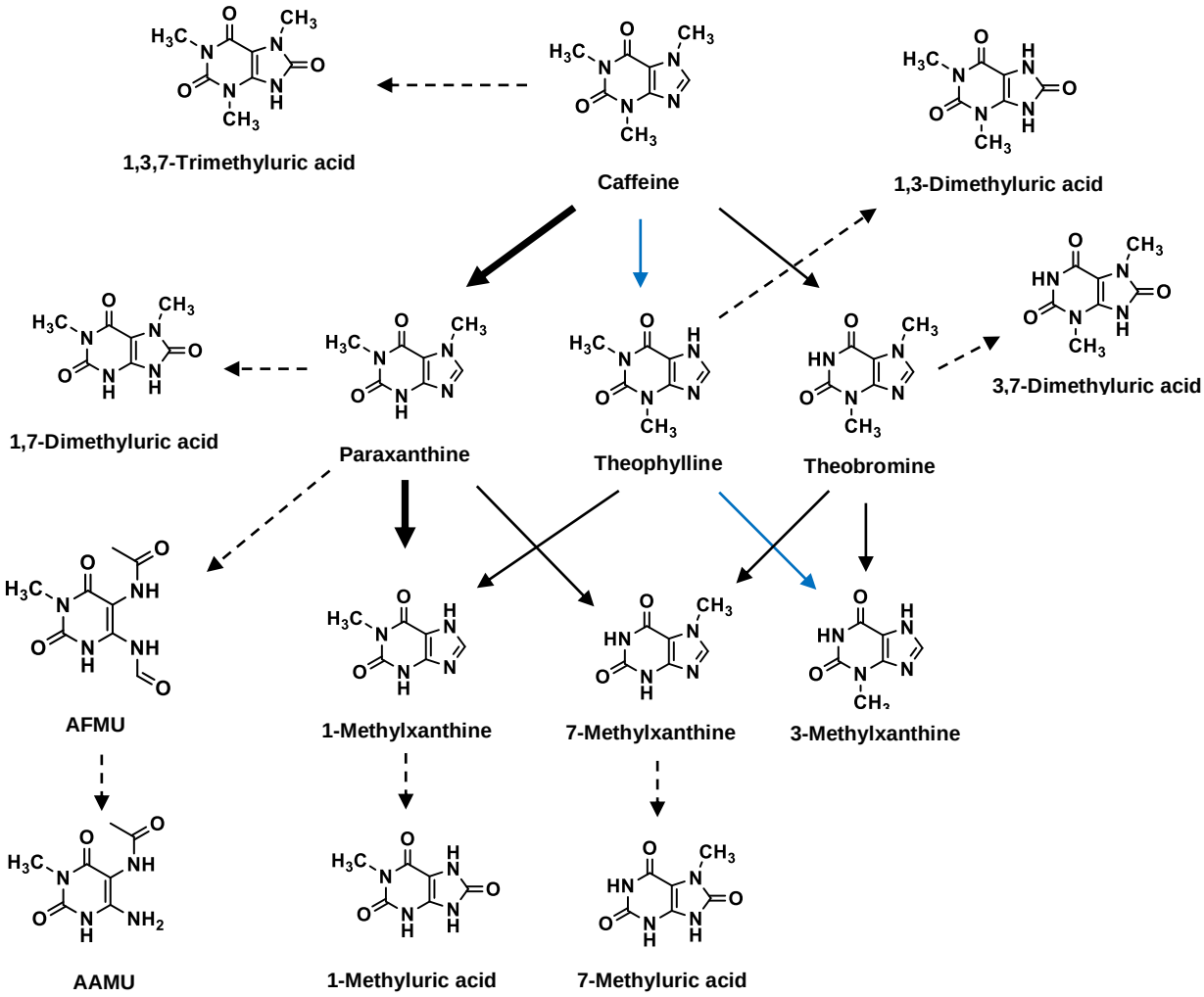



**Figure 1.11**: Caffeine catabolism in humans and fungi. Bold black arrows represent the major pathway in humans. Blue arrows show the major pathway in fungi. Other solid black arrows indicate minor *N*-demethylation steps in humans. Dashed arrows show C-8 oxidation of methylxanthines in humans. Abbreviations include AFMU, 6-formylamino-5-acetylamino-3-methyluracil; AAMU, 6-amino-5-acetylamino-3-methyluracil.

Source: Adapted from Lelo et al., 1986

such as yeast, fungi, and bacteria, have been isolated from the environment. Up until the 1970s, caffeine metabolism in microbes had not received much attention since caffeine was believed to be toxic to microbes [Gopishetty et al., 2011]. Yeast cells degrade caffeine in a similar way as humans, with paraxanthine as the first metabolite and requiring P450 enzymes to catalyze the *N*-demethylation of caffeine [Sauer et al.,

1982]. Although some fungal strains (Table1.4) capable of using caffeine as a sole nitrogen source have been isolated and studied over the past few decades, little is known about the enzymology of fungal caffeine metabolism [Gummadi et al., 2012]. Nonetheless, whole cell biodegradation assays in combination with HPLC analysis using fungal strains such as *Aspergillus* spp. and *Penicillium* spp. suggest an *N*-demethylation pathway, with theophylline as the first metabolite and 3-methylxanthine as the second metabolite (blue arrows in Figure 1.11) [Roussos et al., 1995; Hakil et al., 1998; Brand et al., 2000; Guitierrez-Sanchez et al., 2013; Nayak et al., 2013; Nanjundaiah et al., 2016; Zhou et al. 2018]. In some cases, trace amounts of theobromine, paraxanthine, 1-methylxanthine, and 7-methylxanthine were detected [Hakil et al., 1998].

**Table 1.4:** Caffeine-degrading fungal strains

| No. | Fungi | References |
|---|---|---|
| 1 | *Aspergillus* sp. V12A25 | (Roussos et al. 1995; Hakil et al., 1998) |
| 2 | *Penicillium* sp. V26A25 | |
| 3 | *Penicillium* sp. V33A25 | |
| 4 | *Aspergillus niger* C16A25 | |
| 5 | *Aspergillus fumigatus* C11B25 | |
| 6 | *Aspergillus niger* C17B25 | |
| 7 | *Aspergillus niger* C28B25 | |
| 8 | *Aspergillus tamarii* | (Gutierrez-Sanchez et al., 2013) |
| 9 | *Gliocladium roseum* | (Nayak et al., 2013) |
| 10 | *Aspergillus restrictus* | |
| 11 | *Chysosporium keratinophilum* | |
| 12 | *Fusarium solani* | |
| 13 | *Fusarium solani* | (Nanjundaiah et al., 2016) |
| 14 | *Aspergillus sydowii* NRRL250 | (Zhou et al., 2018) |

Source: Korekar et al., 2020.

## Catabolism of methylxanthines in bacteria

At least 71 caffeine-degrading bacterial strains belonging to 27 genera have been isolated over the past 50 years, with *Pseudomonas* species being the most common isolates identified [Vega et al., 2021]. Unlike plants and mammals [Mohanty et al., 2012], bacteria can use caffeine as a sole source of carbon, nitrogen, and energy for growth. Early studies using these caffeine-degrading bacteria revealed two catabolic pathways to degrade caffeine: the *N*-demethylation pathway and C-8 oxidation pathway [Summers et al., 2015]. However, the *N*-demethylation pathway is the most common pathway since it has been observed in more than 80% of the bacterial isolates in which caffeine degradation has been studied (Table 1.5). Even though caffeine-degrading bacteria have been studied for over 50 years, the nature of the genes and enzymes involved in the *N*-demethylation and C-8 oxidation catabolic pathways were only characterized in the past decade [Summers et al., 2015]. Nonetheless, the genes encoding proteins involved in chemotaxis and transcriptional regulation of these pathways remain unknown.

**Table 1.5**: List of isolated bacterial strains in which the caffeine catabolic pathway has been reported.

| Bacterial strain | Catabolic pathway | Reference |
|---|---|---|
| *Alcaligenes* sp. CF8 | C-8 oxidation | Mohapatra et al. 2006 |
| *Klebsiella – Rhodococcus* (mixture) | C-8 oxidation | Madyastha and Sridhar 1998 |
| *Methylorubrum populi* Pinkel | N-demethylation | Parales et al. 2019 |
| *Paraburkholderia caffeinilytica* CF1 | N-demethylation | Sun et al. 2020 |
| *Pseudomonas* sp. No. 6 | N-demethylation | Asano et al. 1993 |
| *Pseudomonas* sp. NCIM 5235 | N-demethylation | Dash and Gummadi 2006 |
| *Pseudomonas* sp. TH1 | N-demethylation | Topp et al. 2006 |
| *Pseudomonas* sp. CBB1 | C-8 oxidation | Yu et al. 2008 |
| *Pseudomonas* sp. CES | N-demethylation | Summers et al. 2020 |
| *Pseudomonas fluorescens* NRRL B-8052, NRRL B-8053 | N-demethylation | Haas and Stieglitz 1980 |
| *Pseudomonas fulva* | N-demethylation | Ceja-Navarro et al. 2015 |
| *Pseudomonas putida* 40 | N-demethylation | Woolfolk 1975 |
| *Pseudomonas putida* C1 | N-demethylation | Blecher 1976; Blecher and Lingens 1977; Hohnloser et al. 1980 |
| *Pseudomonas putida* NRRL B-8051 | N-demethylation | Haas and Stieglitz 1980 |
| *Pseudomonas putida* WS | N-demethylation | Glück and Lingens 1987, 1988 |
| *Pseudomonas putida* No. 352 | N-demethylation | Asano et al. 1993, 1994 |
| *Pseudomonas putida* IF-3 | N-demethylation | Koide et al. 1996 |
| *Pseudomonas putida* ATCC 700097 | N-demethylation | Ogunseitan 1996, 2002 |
| *Pseudomonas putida* L | N-demethylation | Yamaoka-Yano and Mazzafera 1999 |
| *Pseudomonas putida* KD6 | N-demethylation | Sideso et al. 2001 |
| *Pseudomonas putida* CBB5 | N-demethylation | Yu et al. 2009; Quandt et al. 2015 |
| *Serratia marcescens* | N-demethylation | Mazzafera et al. 1996 |

Source: Adapted from Vega et al., 2021

C-8 oxidation pathway in bacteria: The first step in the C-8 oxidation pathway (Figure 1.12) involves the oxidation of caffeine at the C-8 position to form 1,3,7-trimethyluric acid (TMU). So far, only three enzymes capable of catalyzing the C-8 oxidation of caffeine have been reported: two caffeine oxidases and one caffeine dehydrogenase. The first caffeine oxidase was isolated from a mixed culture of *Klebsiella* sp. and *Rhodococcus* sp. [Madyashta et al., 1999], while the second caffeine oxidase was isolated from a pure culture of *Alcaligenes* sp. CF8 [Mohapatra et al.,

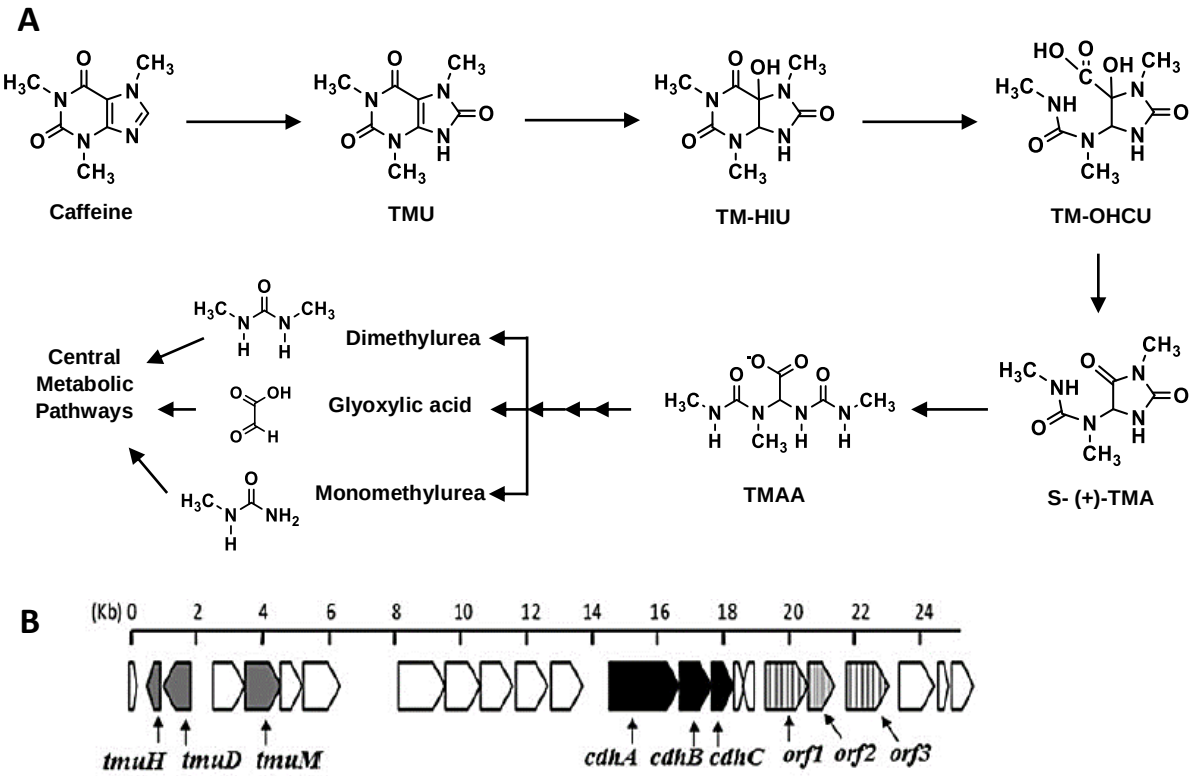



**Figure 1.12**: (A) Proposed C-8 oxidation pathway and (B) map of associated genes in *Pseudomonas putida* CBB1. Abbreviations include TMU, 1,3,7-trimethyluric acid; TM-HIU, 1,3,7-trimethyl-5-hydroxyisouric acid; TM-OHCU, 3,6,8-trimethyl-2-oxo-4-hydroxyl-4-carboxy-5-ureidoimidazoline; *S*-(+)-TMA, *S*-(+)-3,6,8-trimethylallantoin; TMAA, 1,6,8-trimethylallantoic acid. *cdhABC* encodes trimeric caffeine dehydrogenase; *tmuM* encodes trimethyluric acid monooxygenase; *tumH* encodes putative TM-HIU hydrolase; *tmuD* encodes putative TM-OHCU decarboxylase; *orf1* encodes a putative trimethylallantoinase; *orf2*, *orf3* encode putative gene products responsible for catabolism of 1,6,8-trimethylallantoic acid to simpler metabolites such as dimethylurea, monomethylurea and glyoxylic acid.

Source: Adapted from Summers et al., 2015

2006]. Both oxidases seem to have high specificity towards caffeine since only low activity was detected towards any other tested methylxanthine [Summers et al., 2015]. More recently, a quinone-dependent heterotrimeric caffeine dehydrogenase (Cdh) encoded by *cdhABC* was discovered in *Pseudomonas putida* CBB1 [Yu et al., 2008; Mohanty et al., 2012]. Once TMU is produced, it is oxidized to form 1,3,7-trimethyl-5-hydroxyisourate (TM-HIU) by a trimethyluric acid monooxygenase (TmuM), which is

then hydrolyzed to generate 3,6,8-trimethyl-2-oxo-4-hydroxy-4-carboxy-5 ureidoimidazoline (TM-OHCU) [Mohanty et al., 2012]. In the fourth step, a decarboxylase converts TM-OHCU to 3,6,8-trimethylallantoin (TMA). Degradation beyond TMA has not been fully characterized; however, the presence of a gene encoding a trimethylallantoinase homologous to allantoinases observed in uric acid degradation pathways suggest that TMA is hydrolyzed to 1,6,8-trimethylallantoic acid (TMAA) in the C-8 oxidation pathway [Summers et al., 2015]. Subsequently, TMAA can be catabolized to generate glyoxylic acid, dimethylurea, and monomethylurea to enter central metabolic pathways [Mohanty et al., 2013].

*N*-demethylation pathway in bacteria: In this pathway, the methyl groups in the caffeine molecule are removed via three sequential steps to form xanthine (Figure 1.13). For the major route (bold arrows in Figure 1.13), the methyl group located at the $N_1$ position is removed first to generate theobromine [Summers et al., 2011]. During the second step, the methyl group at the $N_3$ position is removed from theobromine to form 7-methylxanthine, which is further demethylated during the last step to yield xanthine [Summers et al., 2012]. Some bacterial strains have a minor route (dotted arrows in Figure 1.13) in which the methyl group at the $N_3$ position is removed from caffeine in the first step to form paraxanthine, which is then converted to 7-methylxanthine [Yamaoka and Mazzafera, 1999; Yu et al., 2009]. Each demethylation step produces one molecule of formaldehyde and one water molecule with the incorporation of molecular oxygen [Summers et al., 2012]. Once xanthine is generated, it is oxidized at the C-8 position to produce uric acid, which enters the purine catabolic pathway [Summers et al., 2015].

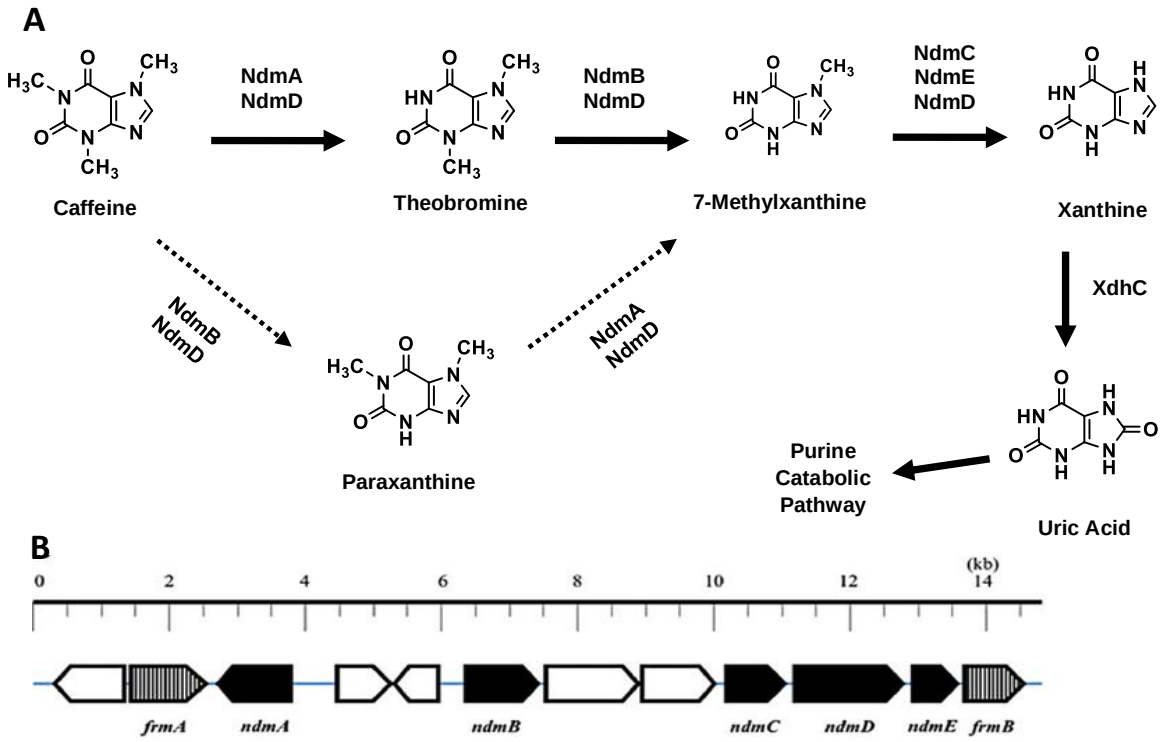



**Figure 1.13**: (A) *N*-demethylation pathway and (B) map of associated genes in *Pseudomonas putida* CBB5. Bold arrows represent the major pathway. Dotted arrows represent the alternative route.

Source: Adapted from Summers et al., 2015

During the past decade, the five proteins responsible for the three demethylation steps (NdmABCDE) were characterized in *Pseudomonas putida* CBB5 (Figure 1.13) [Summers et al., 2011; Summers et al., 2012; Summers et al., 2013]. The enzyme NdmA removes the methyl group at the $N_1$ position of methylxanthines, whereas NdmB specifically removes the methyl group at the $N_3$ position. Both enzymes are Rieske [2Fe-2S] monooxygenases with a non-heme ferrous iron at the active site, and both require the electron transfer protein NdmD for activity [Summers et al., 2012]. When *ndmD* is expressed with *ndmC* and *ndmE*, the three proteins form a complex that can remove the methyl group at the $N_7$ position of methylxanthines [Summers et al., 2013].

NdmC is a non-heme iron monooxygenase, but unlike NdmA and NdmB, it does not contain a Rieske [2Fe-2S] cluster. NdmD is a Rieske-type reductase and is required for all three demethylation steps to transfer electrons from NADH to power the demethylation reactions [Summers et al., 2011]. NdmE has a structural role in the formation of the NdmCDE complex that allows all three proteins to be soluble and active [Summers et al., 2012; Summers et al., 2014].

## Methylotrophy as a platform to study caffeine biodegradation

The organism used by the Parales laboratory to study caffeine degradation is *Methylorubrum populi* strain PINKEL, which is the only methylotrophic bacterium, to our knowledge, that has been obtained from caffeine enrichments. PINKEL was isolated from compost containing coffee grounds, following an enrichment strategy that used a minimal medium supplemented with 5 mM caffeine as the sole carbon, nitrogen, and energy source [Parales et al., 2019]. A pure, pink-pigmented culture was obtained after several passes on caffeine minimal agar plates. Phylogenetic analysis using 16S rRNA gene sequencing indicated that the closest relative to PINKEL is *Methylorubrum populi*. The draft genome of PINKEL (accession number WEKV00000000) has been assembled in 35 contigs with 5.6 mega base pairs. Approximately 5,681 protein-coding sequences were identified, including genes homologous to those in the *P. putida* N-demethylation pathway: *ndmA* ($N_1$-demethylase; peg_1790)*, ndmB* ($N_3$-demethylase; peg_1782)*, ndmD* (reductase; peg_1781) and *ndmCDE* ($N_7$-demethylase; peg_1777, 1781, 1776) [Parales et al., 2019]. Additionally, optimal growth conditions and methods for genetic manipulation (plasmids to knockout genes, transcriptional promoter fusion,

and complementation) to allow the study of caffeine degradation in PINKEL have been developed as part of this work.

Methylotrophic bacteria are Gram-negative, aerobic organisms found in various environments, including soil, water, and plants. By definition, these bacteria can use C1 compounds, which are reduced carbon sources lacking carbon-carbon bonds (i.e., methanol, methylamine, and formaldehyde) [Anthony, 1982]. PINKEL belongs to the genus *Methylorubrum,* which means it is also a facultative methylotroph capable of using non-C1 carbon compounds (i.e., succinate and caffeine). Thus, the ability of PINKEL to use the formaldehyde produced during caffeine catabolism as a carbon and energy source may provide a new caffeine-degrading microorganism with a unique metabolism that can potentially result in more effective caffeine consumption, in comparison to *Pseudomonas* strains currently used to study caffeine biodegradation. Further, PINKEL can serve as an additional tool to enable the identification of genes involved in the caffeine chemotactic response as well as the transcriptional regulation of caffeine catabolic genes.

## Significance and applications of xanthine alkaloids

There are many reasons for studying the biodegradation of xanthine alkaloids, including their potential applications in the food, medical, biotechnological, and environmental industries. Hence, a complete understanding, not only of the bacterial enzymes necessary for caffeine degradation, but how those genes are regulated and how bacteria sense xanthine alkaloids in the environment may contribute to these potential applications.

Food industry: Bio-decaffeination of coffee and tea

Caffeine, a significant component in the human diet, is widely consumed through beverages and other food products (Table 1.6) [Gummadi et al., 2012]. Although caffeine has various health benefits, it has also been reported to cause diarrhea in people with certain gastrointestinal sensitivities, such as irritable bowel syndrome [Koochakpoor et al., 2021]. Hence, certain members of the population prefer decaffeinated food and beverage alternatives. Currently, caffeine is removed via solvent extraction, water diffusion, or pressurized $CO_2$ extraction methods [Feldman and Katz, 1976; Udayasankar et al., 1986; Coultate, 2009]. The high temperatures and/or pressures at which these extraction methods are performed require large energy inputs and thus, have a high operational cost. Additionally, the solvent extraction methods use toxic compounds, such as dichloromethane, as part of the extraction process and may alter flavor in some food products [Mohapatra et al., 2006].

A greener and more environmentally friendly alternative to these chemical decaffeination methods involves the use of microbial cells and/or enzymes from caffeine-degrading microbes to catalyze bio-decaffeination [Gopishettey et al., 2011]. Although there is no current use of microbial cells for bio-decaffeination at an industrial scale, it has been shown that bacteria such as *Pseudomonas putida* CBB5 and CBB1 are capable of decaffeinating coffee and tea extracts [Gopishettey et al., 2012]. However, one of the major concerns in using bacteria for bio-decaffeination is the release on endotoxins during the process. An approach that is currently being explored to solve the endotoxin problem is to clone the caffeine-degrading genes into generally regarded as safe organisms (GRAS), such as *Saccharomyces cerevisiae*, to create an

enhanced caffeine-degrading strain [Summers et al., 2015]. Alternatively, other groups have been studying the feasibility of gene editing coffee plants such that caffeine synthesis is blocked [Leibrock et al., 2022]. Overall, using caffeine-degrading microbes and/or their enzymes provides a decaffeination method that is more eco-friendly without affecting flavor as with the current methods.

**Table 1.6**: Concentrations of caffeine in various common beverages and drugs

| Source of Caffeine | Serving size | mg/serving |
|---|---|---|
| **Coffee** | | |
| Decaffeinated | 10 oz | 4-15 |
| Instant | 10 oz | 9-216 |
| Plain, brewed | 10 oz | 128-250 |
| Espresso | 5 oz | 150-450 |
| **Tea** | | |
| Tea, brewed | 10 oz | 80-120 |
| Iced Tea | 12 oz | 65-75 |
| Green | 8 oz | 30-50 |
| Black | 8 oz | 25-110 |
| Yerba Mate | (8 oz) | 65-130 |
| **Chocolate** | | |
| Hot cocoa | 5 oz | 4 |
| Chocolate milk | 6 oz | 4 |
| Milk chocolate | 1.5 oz | 9 |
| Chocolate bar | 1.5 oz | 30 |
| **Soft Drinks** | | |
| Coca Cola | 12 oz | 35 |
| Pepsi | 12 oz | 40 |
| Dr. Pepper | 12 oz | 40 |
| Mountain Dew | 12 oz | 55 |
| **Energy Drinks** | | |
| Monster | 16 oz | 160 |
| Rockstar | 16 oz | 160 |
| Red Bull | 8.3 oz | 80 |
| Amp | 8.3 oz | 75 |
| **Drugs** | | |
| Anacin | 2 tablets | 32 |
| Excedrin | 1 tablet | 65 |

Source: Adapted from Carillo and Benitez 2000 and Gopishettey et al., 2011.

## Medical industry: Health and beauty applications

Xanthine derivatives, both natural and synthetic, have pharmaceutical and clinical applications due to their neurological, cardiac, and respiratory effects. Although the complete mechanism of action of methylxanthines is not fully understood, it is well known that these compounds exert their effects by acting as antagonists of $A_1$ and $A_2$ adenosine receptors and serve as non-selective phosphodiesterase inhibitors and activators of histone deacetylase [Gottwalt and Tadi, 2022]. These modes of action result in bronchial smooth muscle relaxation, pulmonary vasodilation, diuresis, as well as central nervous system and cardiac stimulation.

Caffeine is often used as an analgesic enhancer in cold, cough, headache, and asthma medications [Daly, 2007]. Since theophylline is a potent bronchodilator, it has been used to treat neonatal apnea, control asthma, and relieve bronchial spasms [Mazzafera, 2002]. On the other hand, theobromine is primarily used as a diuretic, myocardial stimulant, or vasodilator [Stavric, 1988]. Methyluric acids have been used in skin cosmetics and anti-dandruff preparations [Madyastha and Sridhar, 1999]. Synthetic xanthine derivatives such as doxofylline (commercially sold as Ansimar) and enprofylline have been studied for their bronchodilation activity but without the cardio stimulatory side effects observed for theophylline [Sankar 2008].

## Biotechnological industry: production of pharmaceuticals

Caffeine and some of its metabolites are highly valued chemicals for the production of pharmaceuticals that are used as active ingredients in asthma and heart medications [Dash and Gummadi, 2006; Daly, 2007; Sankar et al., 2008]. Unfortunately,

the chemical synthesis of xanthine-derived pharmaceuticals is expensive, requires multiple steps, and uses highly toxic and volatile reagents such as tetrahydrofuran (THF) and *N,N*-dimethylformamide (DMF) [Mock and Summers, 2023]. Additionally, selective alkylation is difficult to achieve because each nitrogen in the xanthine ring is chemically similar [Gopishetty et al., 2011]. Thus, the ability to harness bacterial enzymes that remove specific methyl groups from caffeine and replace them with alternative R-groups would yield a more effective strategy to produce medically-relevant xanthine derivatives. Mock and Summers [2023] successfully engineered *Escherichia coli* strains capable of producing different methylxanthines from caffeine by cloning different combinations of the five *ndm* genes from *P. putida* CBB5. However, the observed decrease in cell activity at concentrations of caffeine greater than 5 mM poses a limitation of the developed biocatalytic method. Thus, further optimization of the current method and/or the use of an alternative host more tolerant to high caffeine concentrations is needed.

Environmental industry: bioremediation and waste-water treatment

Despite its benefits, caffeine is considered an environmental pollutant since it impacts insect, plant, and microbial ecosystems [Nathanson, 1984; Rogers and Dernoncourt, 1998; Gokulakrishnan et al., 2005]. The worldwide consumption of caffeine and related xanthine alkaloids via food, beverages, and pharmaceuticals has resulted in increasing environmental contamination of soil and water systems [Li et al., 2020]. In fact, caffeine is currently used as an indicator of contamination by other pharmaceutically active compounds [He et al., 2018]. Hence, there is a great interest in

using caffeine-degrading bacteria as a bioremediation strategy to clean up contaminated environments.

## Dissertation overview

In this dissertation, I describe research on the caffeine-degrading bacterium *Methylorubrum populi* strain PINKEL as a model to study caffeine biodegradation and characterize the transcriptional regulatory system for caffeine-degrading enzymes. While the genes and enzymes for both the *N*-demethylation and C-8 oxidation pathways have been well characterized [Summers et al., 2015], little is known about how these genes are regulated at the transcriptional level or about the chemotaxis response towards caffeine and/or related xanthine alkaloids. In light of previously published results and preliminary data, my overarching hypotheses are:

1. PINKEL uses the *N*-demethylation pathway to degrade caffeine (Chapter 2), and
2. Expression of genes involved in this pathway is positively regulated by a specific transcriptional activator in response to caffeine or its metabolites (Chapter 3).

To test the first hypothesis, whole cell biodegradation assays in combination with HPLC analysis were used to identify the metabolites formed by PINKEL as it degrades caffeine and to corroborate that PINKEL follows the *N*-demethylation pathway. Additionally, the ability of PINKEL to grow on all *N*-demethylation metabolites was tested. Pathway mutants were constructed to verify that all predicted *ndm* genes are relevant for caffeine biodegradation, which offered insights into substrate specificity. To test the second hypothesis, four potential *ndm* promoter regions were cloned upstream

of a promoterless yellow fluorescent protein gene (*yfp)* and wild-type expression patterns were measured as an output of YFP via a transcriptional promoter fusion assay. A mutant strain lacking the putative LysR-type transcriptional regulator encoded in the vicinity of the predicted *ndm* genes was constructed and differential expression patterns were measured and compared to those of the wild-type strain. To determine the identity of the inducer, the *ndm* promoter fusion plasmids were mated into the pathway mutants and expression was measured in the presence and absence of each metabolite of the *N*-demethylation pathway. The confirmed *ndm* promoter regions were analyzed bioinformatically to predict the potential DNA binding site of the identified transcriptional regulator. Lastly, strains and methods to study chemotaxis responses towards methylxanthines and C1 compounds in PINKEL were developed.

## References


Anderson LE, & Gibbs M. (1962). The biosynthesis of caffeine in the coffee plant. *Journal of Biological Chemistry* 237(6): 1941-1944.

Aneja M, & Gianfagna T. (2001). Induction and accumulation of caffeine in young, actively growing leaves of cocoa (*Theobroma cacao* L.) by wounding or infection with *Crinipellis perniciosa*. *Physiology and Molecular Plant Pathology* 59: 13.

Anthony C. (1982). The biochemistry of methylotrophs. Academic Press, London, United Kingdom.

Ashihara H, Monteiro AM, Gillies FM, & Crozier A. (1996a). Biosynthesis of caffeine in leaves of coffee. *Plant Physiology* 111: 747.

Ashihara H, Monteiro AM, Moritz T, Gillies FM, & Crozier A. (1996b). Catabolism of caffeine and related purine alkaloids in leaves of *Coffea arabica* L. *Planta* 198: 334-339.

Ashihara H, Gillies FM, & Crozier A. (1997). Metabolism of caffeine and related purine alkaloids in leaves of tea (*Camellia sinensis* L.). *Plant Cell Physiology* 38: 413–419.

Ashihara H & Crozier A. (1999). Biosynthesis and metabolism of caffeine and related purine alkaloids in plants. *Advances in Botanical Research* 30, 117–205. doi: 10.1016/ s0065-2296(08)60228-1.

Ashihara H, Sano H, & Crozier, A. (2008). Caffeine and related purine alkaloids: Biosynthesis, catabolism, function, and genetic engineering. *Phytochemistry* 69(4): 841-856.

Ashihara H, Kato M, & Crozier A. (2011). Distribution, Biosynthesis and Catabolism of Methylxanthines in Plants. In: Methylxanthines. Handbook of Experimental Pharmacology, vol 200. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-13443-2_2.

Ashihara H. (2015). Plant biochemistry: Trigonelline biosynthesis in *Coffea arabica* and *Coffea canephora*. In: Coffee in Health and Disease Prevention (ed. V.R. Preedy), 19–28. San Diego: Academic Press.

Ashihara H, Mizuno K, Yokota T, & Crozier A. (2017). Xanthine Alkaloids: Occurrence, Biosynthesis, and Function in Plants. *Progress in Chemistry of Organic Natural Products* 105: 1-88. doi: 10.1007/978-3-319-49712-9_1. PMID: 28194561.

Ashihara H, Stasolla C, Fujimura T, & Crozier A. (2018). Purine salvage in plants. *Phytochemistry* 147: 89–124.

Ashihara H, Ludwig IA, & Crozier A. (2020). Occurrence of Purine Alkaloids. *In* Plant Nucleotide Metabolism - Biosynthesis, Degradation, and Alkaloid Formation (eds H. Ashihara, I.A. Ludwig & A. Crozier). https://doi.org/10.1002/9781119476139.ch14

Baumann TW, Koetz R, & Morath P. (1983). *N*-methyltransferase activities in suspension cultures of *Coffea arabica* L. *Plant Cell Reports* 2: 33-35.

Baumann TW & Frischknecht PM. (1988). Purines. *In* Phytochemicals in Plant Cell Cultures (F. Constabel & I.K. Vasil, eds.). https://doi.org/10.1016/B978-0-12-715005-5.50030-3.

Berthou F, Flinois J P, Ratanasavanh D, Beaune P, Riche C, & Guillouzo A. (1991). Evidence for the involvement of several cytochromes P-450 in the first steps of caffeine metabolism by human liver microsomes. *Drug Metabolism and Disposition* 19(3): 561-567.

Berthou F, Guillois B, Riche C, Dreano Y, Jacqz-Aigrain E, & Beaune PH. (1992). Interspecies variations in caffeine metabolism related to cytochrome P4501A enzymes. *Xenobiotica* 22(6): 671-680.

Brand D, Pandey A, Leon JR, Roussos S, & Soccol CR. (2000). Biological detoxification of coffee husk by filamentous fungi using a solid-state fermentation system. *Enzyme and Microbial Technology* 26: 127–133.

Buchanan JM, Sonne JC, & Delluva, AM. (1948). Biological precursors of uric acid. *Journal of Biological Chemistry* 173(1): 81-98.

Campbell ME, Grant DM, Inaba T, & Kalow W. (1987). Biotransformation of caffeine, paraxanthine, theophylline, and theobromine by polycyclic aromatic hydrocarbon-inducible cytochrome (s) P-450 in human liver microsomes. *Drug Metabolism and Disposition* 15(2): 237-249.

Carrillo JA, & Benitez J. (2000). Clinically significant pharmacokinetic interactions between dietary caffeine and medications. *Clinical Pharmacokinetics* 39(2): 127-153.

Chou CH, & Waller GR. (1980). Possible allelopathetic constituents of *Coffea arabica*. *Journal of Chemical Ecology* 6: 643-54.

Coultate TP. (2009). Food: the chemistry of its components. The Royal Society of Chemistry, Cambridge, UK.

Dash SS & Gummadi SN. (2006). Catabolic pathways and biotechnological applications of microbial caffeine degradation. *Biotechnology Letters* 28: 1993-2002.

Daly JW. 2007. Caffeine analogs: biomedical impact. *Cellular and Molecular Life Sciences* 64: 2153-2169.

Feldman JR, & Katz SM. (1976). Caffeine. In *Encyclopedia of Chemical Processing and Design: Blowers to Calcination.* Edited by J. J. McKetta. Marcel Dekker, New York, NY, p. 424-440.

Franco R, Onatibia-Astibia A. & Martinez-Pinilla E. (2013). Health benefits of methylxanthines in cacao and chocolate. *Nutrients* 5: 4159-4173. Doi10.3390/nu5104159.

Friedman J, & Waller G. (1983). Caffeine hazards and their prevention in germinating seeds of coffee (Coffea arabica L.). *Journal of Chemical Ecology* 9: 1099.

Frischknecht PM, & Baumann TW. (1985). Stress induced formation of purine alkaloids in plant tissue culture of *Coffea arabica*. *Phytochemistry* 24: 2255-7.

Fujimori N, Suzuki T, & Ashihara H. (1991). Seasonal variations in biosynthetic capacity for the synthesis of caffeine in tea leaves. *Phytochemistry* 30(7): 2245-2248.

Gottwalt B, & Tadi P. (2022). Methylxanthines. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK559165/.

Gummadi SN, Bhavya B, & Ashock N. (2012). Physiology, biochemistry and possible applications of microbial caffeine degradation. *Applications in Microbiology Biotechnology* 93: 545–554.

Gutiérrez-Sánchez G, Roussos S, & Augur C. (2013). Effect of caffeine concentration on biomass production, caffeine degradation, and morphology of *Aspergillus tamarii*. *Folia Microbiologica* (Praha) 58: 195– 200. https://doi.org/10.1007/s12223-012-0197-3.

Gokulakrishnan S, Chandraraj K & Gummadi SN. (2005). Microbial and enzymatic methods for the removal of caffeine. *Enzyme and Microbial Technology* 37: 225-232.

Gopishetty, SR, Louie TM, Yu CL, & Subramanian MV. (2011). Microbial degradation of caffeine, methylxanthines, and its biotechnological applications, p 44–67. *In* Thatoi HN, Mishra BB (ed), *Microbial Biotechnology Methods and Applications*. Narosa Publishing House Pvt, Ltd, New Delhi, India.

Gummadi SN, Bhavya B & Ashok N. (2012) Physiology, biochemistry and possible applications of microbial caffeine degradation. *Applications in Microbiology Biotechnology* 93: 545–554. https://doi.org/10.1007/s00253 011-3737-x.

Hakil M, Denis S, Viniegra-González G, & Augur C. (1998). Degradation and product analysis of caffeine and related dimethylxanthines by filamentous fungi. *Enzyme and Microbial Technology* 22: 355–359.

He Y, Nurul S, Schmitt H, Sutton NB, Murk TAJ, Blokland MH, Rijnaarts HHM, & Langenhoff AAM. (2018). Evaluation of attenuation of pharmaceuticals, toxic potency, and antibiotic resistance genes in constructed wetlands treating wastewater effluents. *Science of the Total Environment* 631-632: 1572-1581. doi: 10.1016/j.scitotenv.2018.03.083. Epub 2018 Mar 28. PMID: 29727981.

Henderson JF, & Paterson ARP. (1973). Nucleotide metabolism: An introduction. Academic Press, New York, NY.

Hollingsworth RG, Armstrong JW, & Campbell E. (2002). Pest control caffeine as a repellent for slugs and snails. *Nature* 417: 915.

Hollingsworth RG, Armstrong JW, & Campbell E. (2003). Caffeine as a novel toxicant for slugs and snails. *Annals of Applied Biology* 142: 91.

Ito E, Crozier A, & Ashihara H. (1997). Theophylline metabolism in higher plants. *Biochimica et Biophysica Acta*, General Subjects, 1336: 323–330.

Kato M, Mizuno K, Fujimura T, Iwama M, Irie M, Crozier A, & Ashihara H. (1999). Purification and characterization of caffeine synthase from tea leaves. *Plant Physiology* 120: 579.

Kim YS & Sano H. (2008). Pathogen resistance of transgenic tobacco plants producing caffeine. *Phytochemistry* 68: 882-888.

Kim YS, Choi YE, & Sano H. (2010). Plant vaccination: stimulation of defense system by caffeine production in planta. *Plant Signal and Behavior* 5: 489.

Koochakpoor G, Salari-Moghaddam A, Keshteli AH, Esmaillzadeh A, & Adibi P. (2021). Association of coffee and caffeine intake with irritable bowel syndrome in adults. *Frontiers in Nutrition*. 15; 8:632469. doi: 10.3389/fnut.2021.632469. PMID: 34211993; PMCID: PMC8241212.

Korekar G, Kumar A. & Ugale C. (2020). Occurrence, fate, persistence, and remediation of caffeine: a review. *Environmental Science and Pollution Research* 27: 34715–34733. https://doi.org/10.1007/s11356-019-06998-8.

Koshiishi C, Kato A, Yama S, Crozier A, & Ashihara H. (2001). A new caffeine biosynthetic pathway in tea leaves: utilization of adenosine released from the *S*-adenosyl-L-methionine cycle. *FEBS Letters* 499: 50–54.

Leibrock NV, Santegoets J, Mooijman PJW, Yusuf F, Zuijdgeest XCL, Zutt EA, Jacobs JGM, & Schaart JG. (2022). The biological feasibility and social context of gene-edited, caffeine-free coffee. *Food Science Biotechnology* 31(6): 635-655. doi: 10.1007/s10068-022-01082-3. PMID: 35646415; PMCID: PMC9133285.

Li S, Wen J, He B, Wang J, Hu X, & Liu J. (2020). Occurrence of caffeine in the freshwater environment: Implications for ecopharmacovigilance. *Environmental Pollution* 263(Pt B):114371. doi: 10.1016/j.envpol.2020.114371. Epub 2020 Mar 17. PMID: 32217417.

Madyastha KM, & Sridhar GR. (1999). Highly efficient C-8 oxidation of substituted xanthines with substitution at the 1-, 3-, and 7-positions using biocatalysts. *Journal of the Chemical Society, Perkin Transactions 1* 1999(6): 677-680.

Mazzafera, P. (2002). Degradation of caffeine by microorganisms and potential use of decaffeinated coffee husk and pulp in animal feeding. *Scientia Agricola* 59(4): 815-821.

Mizuno K, Kato M, Irino F, Yoneyama N, Fujimura T, & Ashihara H. (2003). The first committed step reaction of caffeine biosynthesis: 7-methylxanthosine synthase is closely homologous to caffeine synthases in coffee (Coffea arabica L.). *FEBS Letters* 547(1-3): 56-60.

Mock MB, & Summers RM. (2023). Mixed culture biocatalytic production of the high-value biochemical 7-methylxanthine. *Journal of Biological Engineering* 17: 2. https://doi.org/10.1186/s13036 022-00316-6

Mohanty SK, Yu CL, Das S, Louie TM, Gakhar L, & Subramanian M. (2012). Delineation of the caffeine C-8 oxidation pathway in *Pseudomonas* sp. strain CBB1 via characterization of a new trimethyluric acid monooxygenase and genes involved in trimethyluric acid metabolism. *Journal of Bacteriology* 194: 3872–3882.

Mohanty SK. (2013) A. Genetic characterization of the caffeine C-8 oxidation pathway in *Pseudomonas* sp. CBB1. B. Validation of caffeine dehydrogenase as a suitable enzyme for a rapid caffeine diagnostic test. PhD diss., University of Iowa, 2013.

Mohapatra BR, Harris N, Nordin R, & Mazumder A. (2006). Purification and characterization of a novel caffeine oxidase from *Alcaligenes* species. *Journal of Biotechnology* 125: 319-327.

Nanjundaiah S, Bhatt P, Rastogi NK, & Thakur MS. (2016). Response surface optimization for decaffeination and theophylline production by *Fusarium solani*. *Applied Biochemistry and Biotechnology* 178:58–75. https://doi.org/10.1007/s12010-015-1858-x.

Nathanson JA. (1984). Caffeine and related methylxanthines: Possible naturally occurring pesticides. *Science* 226 (4671): 184-187.

Nayak V, Pai PV, Pai A, Pai, S, Shushma, YD, & VRao CV. (2013). A comparative study of caffeine degradation by four different fungi. *Bioremediation Journal* 17: 79–85. https://doi.org/10.1080/10889868.2012.751960.

Negishi O, Ozawa T, & Imagawa H. (1985). Methylation of xanthosine by tea-leaf extracts and caffeine biosynthesis. *Agricultural and Biological Chemistry* 49: 887-890.

Negishi O, Ozawa T, & Imagawa H. (1988). N-methyl nucleosidase from tea leaves. *Agricultural and Biological Chemistry* 52(1): 169-175.

Negishi O, Ozawa T, & Imagawa H. (1994). Guanosine deaminase and guanine deaminase from tea leaves. *Bioscience, Biotechnology, and Biochemistry* 58: 1277.

Parales RE, Sharma G, Zhang X, Subuyuj GA, Langner JT, Wright ME, Ditty JL, & Dawson SC. (2019). Draft genome sequence of the caffeine-degrading methylotroph *Methylorubrum populi* Pinkel. *Microbiology Resource Announcements* 8: e01300-19.

Petermann J, & Baumann TW. (1983). Metabolic relations between methylxanthines and methyluric acids in *Coffea*. *Plant Physiology* 73: 961.

Proiser E, & Serenkov, GP. (1963). Caffeine biosynthesis in tea leaves. *Biokhimiya* 28: 857-861.

Rizvi SJH, Mukerji D, & Mathur SN. (1981). Selective phyto-toxicity of 1,3,7-trimethylxanthine between *Phaseolus mungo* and some weeds. *Agricultural and Biological Chemistry* 45: 1255.

Roca A, Rodriguez-Herva JJ, Duque E & Ramos JL. (2008). Physiological responses of *Pseudomonas putida* to formaldehyde during detoxification. *Microbial Biotechnology* 1 (2): 158-169.

Rogers PJ & Dernoncourt C. (1998). Regular caffeine consumption: a balance of adverse and beneficial effects for mood and psychomotor performance. *Pharmacology Biochemistry and Behavior* 59:1039-45.

Roussos S, Angeles AM, Refugio Trejo-Hernández M, Gaime Perraud I, Favela E, Ramakrishna M, Raimbault M, & Viniegra-González G. (1995). Biotechnological management of coffee pulp—isolation, screening, characterization, selection of caffeine-degrading fungi and natural microflora present in coffee pulp and husk. *Applied Microbiology and Biotechnology* 42: 756–762.

Sankar J, Lodha R & Kabra SG. (2008). Doxofylline: The next generation methylxanthine. *Indian Journal of Pediatrics* 68 (3): 251-254.

Sauer M, Kappeli O, & Fiechter A. (1982). Comparison of the cytochrome P450-containing monooxygenases originating from two different yeasts. In *Cytochrome P-450, Biochemistry, Biophysics, and Environmental Implications* pp. 453-457. Edited by E. Hietanen, M. Laitinen & O. Hanninen. Amsterdam.

Savitri Kumar N, Hewavitharange P, & Adikaram NKB. (1995). Attack on tea by *Xyleborus fornicatus*: inhibition of the symbiote, *Monacrosporium ambrosium*, by caffeine. *Phytochemistry* 40: 1113-6.

Shemin D, & Rittenberg D. (1947). On the utilization of glycine for uric acid synthesis in man. *Journal of Biological Chemistry* 167(3): 875-876.

Smyth D. (1992). Effect of methylxanthine treatment on rice seedling growth. *Journal of Plant Growth Regulation* 11: 125.

Stasolla C., Katahira R, Thorpe TA, & Ashihara H. (2003). Purine and pyrimidine nucleotide metabolism in higher plants. *Journal of Plant Physiology* 160: 1271–1295.

Stavric B. (1988). Methylxanthines: Toxicity to humans. 2. Caffeine. *Food and Chemical Toxicology* 26(7): 645-662.

Stryer L. (1995). Biochemistry, 4th ed. WH Freeman and Co., New York, NY.

Summers RM, Louie TM, Yu CL & Subramanian M. (2011). Characterization of a broad specificity non-haem iron *N*-demethylase from *Pseudomonas putida* CBB5 capable of utilizing several purine alkaloids as a sole carbon and nitrogen source. *Microbiology* 157: 583-592.

Summers RM, Louie TM, Yu CL, Gakhar L, Louie KC & Subramanian M. (2012). Novel, highly specific *N*-demethylases enable bacteria to live on caffeine and related purine alkaloids. *Journal of Bacteriology* 194: 2041-2049.

Summers RM, Seffernick JL, Quandt EM, Yu CL, Barrick JE & Subramanian M. (2013). Caffeine junkie: an unprecedented glutathione *S*-transferase-dependent oxygenase required for caffeine degradation by *Pseudomonas putida* CBB5. *Journal of Bacteriology* 195: 3933-3939.

Summers R, Gopishetty S, Mohanty S, & Subramanian M. (2014). New genetic insights to consider coffee waste as feedstock for fuel, feed, and chemicals. *Central European Journal of Chemistry* 12: 1271-1279.

Summers RM, Mohanty SK, Gopishetty S & Subramanian M. (2015). Genetic characterization of caffeine degradation by bacteria and its potential applications. *Microbial Biotechnology* 8: 369-378.

Suzuki T, & Waller GR. (1984). Biosynthesis and biodegradation of caffeine, theobromine, and theophylline in Coffea arabica L. fruits. *Journal of Agricultural and Food Chemistry* 32: 845-848.

Talik P, Krzek J, Ekiert RJ. (2012). Analytical techniques used for determination of methylxanthines and their analogues. *Recent advances, Separation and Purification Reviews* 41, 1–61. doi: 10.1080/15422119.2011.569047.

Tanti AJ, Bhatacharyya PN, Sandilya SP, & Dutta P. (2016). Allelopathic potential of caffeine as growth and germination inhibitor to popular tea weed, *Borreria hispida*. *Current Life Sciences* 2 (4): 114-117.

Tassaneeyakul W, Birkett DJ, McManus ME, Veronese ME, Andersson T, Tukey RH & Miners JO. (1994). Caffeine metabolism by human hepatic cytochromes P450: Contributions of 1A2, 2E1 and 3A isoforms. *Biochemical Pharmacology* 47(10): 1767-1776.

Tsubouchi H, Terada H, Yamamoto K, Hisada K, Sakabe Y. (1985). Caffeine degradation and increased ochratoxin A production by toxigenic strains of *Aspergillus ochraceus* isolated from green coffee beans. *Mycopathologia* 1985; 90:181-6.

Udayasankar K, Manohar B, & Chokkalingam A. (1986). A note on supercritical carbon dioxide decaffeination of coffee. *Journal of Food Science and Technology* **23**(6): 326–328.

Vega FE, Emche S, Shao J, Simpkins A, Summers RM, Mock MB, Ebert D, Infante F, Aoki S and Maul JE. (2021). Cultivation and Genome Sequencing of Bacteria Isolated from the Coffee Berry Borer (*Hypothenemus hampei*), With Emphasis on the Role of Caffeine Degradation. *Frontiers in Microbiology* 12:644768. doi: 10.3389/fmicb.2021.644768.

Vogels GD, & C Van der Drift. **(**1976). Degradation of purines and pyrimidines by microorganisms. *Bacteriology Reviews* 40: 403–468.

Waller GR. (1989). Biochemical frontiers in allelopathy. *Biologia Plantarum* 31 (6): 418-447.

Yamaoka-Yano DM & Mazzafera P. (1999). Catabolism of caffeine and purification of a xanthine oxidase responsible for methyluric acids production in *Pseudomonas putida* L. *Revista De Microbiologia* 30: 62-70.

Yu CL, Kale Y, Gopishetty S, Louie TM, & Subramanian M. (2008) A novel caffeine dehydrogenase in *Pseudomonas* sp. strain CBB1 oxidizes caffeine to trimethyluric acid. *Journal of Bacteriology* 190: 772–776.

Yu CL, Louie TM, Summers R, Kale Y, Gopishetty S & Subramanian M. (2009). Two distinct pathways for metabolism of theophylline and caffeine are expressed in *Pseudomonas putida* CBB5. *Journal of Bacteriology* 191 (14): 4624-4632.

Zavialov IA, Dahanukar VH, Nguyen H, Orr C & Andrews DR. (2004). New and practical method for synthesis of 1-and 1,3-substituted xanthines. *Organic Letters* 6 (13): 2237-2240.

Zheng XQ, Ye CX, Kato M, Crozier A, & Ashihara H. (2002). Theacrine (1,3,7,9 tetramethyluric acid) synthesis in leaves of a Chinese tea, kucha (*Camellia assamica var. kucha*). *Phytochemistry* 60: 129.

Zhou B, Ma C, Wang H, & Xia T. (2018). Biodegradation of caffeine by whole cells of tea-derived fungi *Aspergillus sydowii, Aspergillus niger* and optimization for caffeine degradation. *BMC Microbiology* 18: 53. https://doi.org/10.1186/s12866-018-1194-8

Zulak KG, Liscome DK, Ashihara H, & Facchini, PJ. (2006). Alkaloids, in Plant Secondary Metabolites: Occurrence, Structure and Role in the Human Diet (eds A. Crozier, M.N. Clifford & H. Ashihara), Blackwell, Oxford, pp. 102–136.

# Chapter 2

## Caffeine catabolic pathway in *Methylorubrum populi* PINKEL

## Introduction

Caffeine and all the metabolites of its catabolism belong to a diverse class of compounds referred to as xanthine alkaloids. Structurally, alkylated xanthines are composed of a xanthine ring with different R-groups attached at the 1st, 3rd, and 7th position of the ring (Figure 1.1 and Table 1.1). In the case of caffeine, each of the three R-groups is a methyl group and as the caffeine molecule is either oxidized or as individual methyl groups are removed, the different metabolites of caffeine biodegradation are generated.

There are many reasons for studying caffeine biodegradation, including its potential applications in the food, bioremediation, and pharmaceutical industries. Caffeine, a significant component in the human diet, is widely consumed through beverages and other food products [Gummadi et al., 2012]. Additionally, caffeine and some of its metabolites are highly valued chemicals to produce pharmaceuticals that are used as active ingredients in asthma and heart medications [Dash and Gummadi, 2006; Daly, 2007; Sankar et al., 2008]. Unfortunately, the chemical synthesis of xanthine-derived pharmaceuticals is expensive, requires multiple steps, and uses highly toxic and volatile reagents. Additionally, selective alkylation is difficult to achieve because each nitrogen in the xanthine ring is chemically similar [Zavialov et al., 2004]. Thus, the ability to harness bacterial enzymes that remove specific methyl groups from caffeine to be replaced with alternative R-groups would yield a more effective strategy to produce medically relevant xanthine derivatives. Despite its benefits, caffeine is also considered an environmental pollutant since it negatively impacts insect, plant, and microbial ecosystems [Nathanson, 1984; Rogers and Dernoncourt, 1998;

Gokulakrishnan et al., 2005]. Hence, there is a great interest in using caffeine-degrading bacteria as a bioremediation strategy.

To advance these potential applications, researchers have been seeking to understand the enzymes involved in caffeine catabolism, in efforts to engineer bacteria for bio-decaffeination, bioremediation, and chemical production. To date, about 71 bacterial strains capable of degrading caffeine have been reported, most of which belong to the genus *Pseudomonas*, primarily *Pseudomonas putida* [Vega et al., 2021]. Early studies using these caffeine-degrading bacteria revealed two caffeine catabolic pathways: the *N*-demethylation pathway and the C-8 oxidation pathway [Summers et al., 2015]. The *N*-demethylation pathway appears to be the most common pathway since it has been observed in more than 80% of the isolates in which caffeine degradation has been studied. In *Pseudomonas putida* CBB5, the *N*-demethylation pathway involves three demethylation steps to convert caffeine (1,3,7-trimethylxanthine) to xanthine and requires five genes: *ndmA, ndmB, ndmC, ndmD and ndmE* (Figure 1.13). The first step is catalyzed by NdmA, which removes the methyl group at the $N_1$ position to convert caffeine to theobromine, which is further demethylated by NdmB at the $N_3$ position to form 7-methylxanthine. Both of these demethylation steps require the aid of NdmD to acquire and transfer electrons from NADH to power the demethylation reactions, which generate formaldehyde as a byproduct. The last methyl group is removed from 7-methylxanthine by the NdmCDE protein complex to generate xanthine. There is a minor route in which the $N_3$ methyl group is removed in the first step, forming paraxanthine, which is then converted by a second demethylation reaction to 7-methylxanthine; however, only trace amounts of paraxanthine are produced in

comparison to theobromine. For most *Pseudomonas putida* strains, formaldehyde is transported out of the cell or oxidized to carbon dioxide to avoid the toxic effects of this highly reactive aldehyde [Roca et al., 2008].

Recently, the three enzymes responsible for the three demethylation steps in *Pseudomonas putida* CBB5 were characterized as Rieske [2Fe-2S] monooxygenases with a non-heme ferrous iron at the active site. Members of this family of enzymes have either two or three separate components that form a soluble electron transport chain to harness the electrons provided by NADH and power the incorporation of oxygen at the catalytic site. In the case of *N*-demethylases, there are two components: a reductase and an oxygenase (Figure 2.1). The reductase has an NADH binding domain, where the electrons are initially harnessed and transferred to the FAD located at the FAD binding domain. The electrons are then shuttled to the Rieske [2Fe-2S] cluster domain in the oxygenase component, where one iron is coordinated by two histidine residues and the second iron is coordinated by two cysteine residues [Ferraro et al., 2005]. The electrons are finally transported to the mononuclear iron for catalysis of the corresponding substrate. In *P. putida* CBB5, NdmA, NdmB and NdmC have been identified as oxygenase components, whereas NdmD has been identified as iron-sulfur flavoprotein reductase. NdmE has been shown to be required for structural support when the NdmCDE complex is formed for the removal of the $N_7$ methyl group of methylxanthines [Summers et al., 2012; Summers et al., 2014].

The organism used by the Parales Lab to study caffeine degradation is *Methylorubrum populi* strain PINKEL, which is the only methylotrophic bacterium, to our knowledge, that has been isolated from caffeine enrichments [Parales et al., 2019].

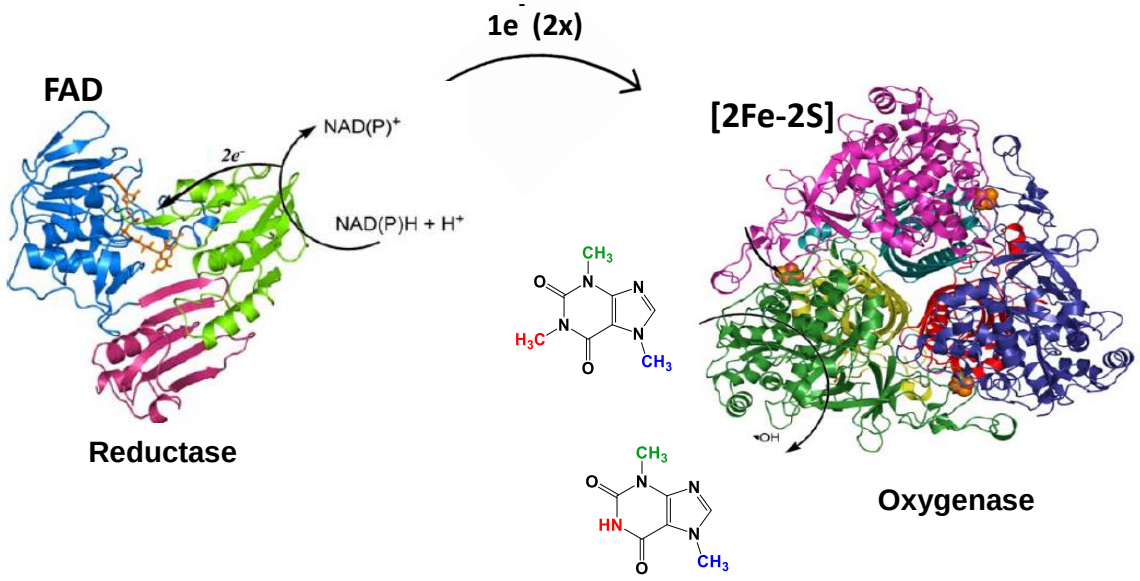



**Figure 2.1**: Two component Rieske [2Fe-2S] non-heme iron monooxygenases. The first component is a reductase protein that harnesses the electrons from NADH, which are then passed to the FAD domain. The electrons are shuttled to the Rieske [2Fe-2S] cluster domain in the oxygenase component. The electrons are transported to the mononuclear iron where the reaction occurs.

Source: Adapted from Ferraro et al., 2005

Methylotrophic bacteria are Gram-negative, aerobic organisms found in various environments, including soil, water, and plants. By definition, these bacteria can use C1 compounds, which are reduced carbon sources containing no carbon-carbon bonds (i.e., methanol, methylamine, and formaldehyde) [Anthony, 1982]. PINKEL belongs to the genus *Methylorubrum,* suggesting that it is a facultative methylotroph capable of using non-C1 carbon compounds (i.e., succinate and caffeine). Thus, the ability of PINKEL to use the formaldehyde produced during caffeine catabolism as a carbon and energy source provides this newly isolated caffeine-degrading microorganism with a unique metabolism that can potentially result in more effective caffeine consumption, in comparison to *Pseudomonas* strains currently used to study caffeine biodegradation.

In this study, the pathway and genes required by PINKEL to metabolize caffeine were examined. Whole cell biodegradation assays in combination with high performance liquid chromatography (HPLC) analysis using wild-type and pathway mutant strains were used to assess metabolite formation, determine the pathway followed by PINKEL to degrade caffeine, and to verify if the predicted genes encode proteins required for caffeine biodegradation.

## Materials and Methods

### Chemical reagents

All the xanthine alkaloids used in this study, including caffeine (1,3,7-trimethylxanthine), theobromine (3,7-dimethylxanthine), paraxanthine (1,7-dimethylxanthine), 7-methylxanthine, xanthine, and uric acid were purchased from Sigma-Aldrich (St. Louis, MO). Unless otherwise stated, all reagents used to make growth media, buffers, agarose gels, and HPLC-grade chemicals were bought from Fisher Scientific (Pittsburg, PA). Restriction endonucleases, Antarctic Phosphatase, T4 DNA ligase, NEB builder HIFI DNA assembly cloning kit, and the Quick-load purple DNA ladder (50 bp and 1 kb) were obtained from New England Biolabs (NEB, Ipswitch, MA). The GoTaq Green (2x) Master Mix was purchased from Promega (Madison, WI). The custom-made DNA oligonucleotides used in this study were bought and analyzed using the Integrated DNA Technologies (IDT) platform (Coralville, IA).

## Methylxanthine stock solutions

It has been well documented that some of the methylxanthines, such as theobromine and xanthine, have poor solubility in aqueous solutions [Bruns and Fergus, 1989]. To overcome this problem and increase the solubility of methylxanthines in aqueous solutions, water was heated to 100°C before adding the corresponding methylxanthine. This method allowed us to make the final concentrations needed for theobromine, paraxanthine, and 7-methylxanthine. To be consistent, the same method was used to dissolve caffeine even though dissolving caffeine in aqueous solutions is not as difficult. Since xanthine does not dissolve, even after boiling the water, growth of cultures on xanthine was tested by providing solid crystals of xanthine as the sole carbon source on solid agar medium. Because theobromine precipitates out of solution within 3 days, all methylxanthine stock solutions were freshly made the day they were used. To ensure that methylxanthines did not degrade as a result of the water boiling procedure, samples from the stock solutions were analyzed via HPLC. No additional metabolites were detected as a result of heating.

## Bacterial strains, primers, and plasmids

The bacterial strains, plasmids, and primers used in this study are listed in Tables 2.1, 2.2, and 2.3, respectively. *Escherichia coli* DH5α was used for cloning and plasmid propagation. *Escherichia coli* S17-1 was used to mate the engineered plasmids into *Methylorubrum populi* PINKEL. Prior to performing any experiments, *M. populi*

**Table 2.1:** Bacterial strains used in this study

| Strain | Description* | Reference or source |
|---|---|---|
| *E. coli* strains | | |
| DH5α | Cloning host | Life Technologies, Gaithersburg, MD |
| NEB 5α | Cloning host | New England Biolabs, Ipswitch, MA |
| S17-1 | Helper strain for conjugation | Simon et al., 1983 |
| *M. populi* strains | | |
| PINKEL | Wild-type caffeine degrading strain | This Study |
| PINKELsw | PINKEL strain swarmed for a highly motile cell population | This Study |
| PINKELnc | PINKELsw strain with Δ*celA* gene cluster deletion (peg_1742 - peg_1745) | This Study |
| GS162 | PINKELnc Δ*ndmA* (peg_1790) | This Study |
| GS163 | PINKELnc Δ*ndmB* (peg_1782) | This Study |
| GS164 | PINKELnc Δ*ndmC* (peg_1777) | This Study |
| GS165 | PINKELnc Δ*ndmD* (peg_1781) | This Study |
| GS166 | PINKELnc Δ*ndmE* (peg_1776) | This Study |
| GS238 | PINKELnc Δ*ndmA* complemented with pGS1 ($Kan^R$) | This Study |
| GS239 | PINKELnc Δ*ndmB* complemented with pGS2 ($Kan^R$) | This Study |
| GS240 | PINKELnc Δ*ndmC* complemented with pGS3 ($Kan^R$) | This Study |
| GS241 | PINKELnc Δ*ndmD* complemented with pGS4 ($Kan^R$) | This Study |
| GS242 | PINKELnc Δ*ndmE* complemented with pGS5 ($Kan^R$) | This Study |
| GS243 | PINKELnc Δ*ndmA* complemented with pLC290 ($Kan^R$) | This Study |
| GS244 | PINKELnc Δ*ndmB* complemented with pLC290 ($Kan^R$) | This Study |
| GS245 | PINKELnc Δ*ndmC* complemented with pLC290 ($Kan^R$) | This Study |
| GS246 | PINKELnc Δ*ndmD* complemented with pLC290 ($Kan^R$) | This Study |
| GS247 | PINKELnc Δ*ndmE* complemented with pLC290 ($Kan^R$) | This Study |

All *M. populi* PINKEL strains used in this study are isogenic derivatives of the swarmed *celA* deletion mutant (PINKELnc).

* $Kan^R$, Kanamycin resistant; sw, swarmed; nc, non-clumping

**Table 2.2:** Plasmids strains used in this study

| Plasmid | Description* | Reference or source |
|---|---|---|
| pHV2 | Modified pCM184 suicide vector expressing *sacB* ($Kan^R$, $Tc^R$, $Amp^R$) | Roszczenko-Jasińska et al., 2020 |
| pLC290 | Expression vector using a cumate inducible promoter ($Kan^R$) | Chubiz et a., 2013 |
| pCM157 | Modified pCM62 vector expressing cre ($Tc^R$). Used to delete kanamycin insertion cassette. | Marx and Lidstrom, 2002 |
| pGS1 | *ndmA* gene cloned into pLC290 ($Kan^R$) | This Study |
| pGS2 | *ndmB gene* cloned into pLC290 ($Kan^R$) | This Study |
| pGS3 | *ndmC gene* cloned into pLC290 ($Kan^R$) | This Study |
| pGS4 | *ndmD gene* cloned into pLC290 ($Kan^R$) | This Study |
| pGS5 | *ndmE* gene cloned into pLC290 ($Kan^R$) | This Study |
| pGS6 | *celABCD* gene cluster (peg_1742 - peg_1745) 767 bp upstream PCR fragment cloned into MfeI - NdeI sites and 964 bp downstream PCR fragment cloned into AgeI - SacI sites of pHV2 ($Kan^R$, $Tc^R$, $Amp^R$) | This Study |
| pGS7 | *ndmA* gene (peg_1790) 629 bp upstream PCR fragment cloned into MfeI - NcoI sites and 878 bp downstream PCR fragment cloned into AgeI - SacI sites of pHV2 ($Kan^R$, $Tc^R$, $Amp^R$) | This Study |
| pGS8 | *ndmB* gene (peg_1782) 565 bp upstream PCR fragment cloned into MfeI - NcoI sites and 749 bp downstream PCR fragment cloned into AgeI - SacI sites of pHV2 ($Kan^R$, $Tc^R$, $Amp^R$) | This Study |
| pGS9 | *ndmC* gene (peg_1777) 639 bp upstream PCR fragment cloned into MfeI - NcoI sites and 598 bp downstream PCR fragment cloned into AgeI - SacI sites of pHV2 ($Kan^R$, $Tc^R$, $Amp^R$) | This Study |
| pGS10 | *ndmD* gene (peg_1781) 523 bp upstream PCR fragment cloned into MfeI - NcoI sites and 1008 bp downstream PCR fragment cloned into AgeI - SacI sites of pHV2 ($Kan^R$, $Tc^R$, $Amp^R$) | This Study |
| pGS11 | *ndmE* gene (peg_1776) 732 bp upstream PCR fragment cloned into MfeI - NcoI sites and 802 bp downstream PCR fragment cloned into AgeI - SacI sites of pHV2 ($Kan^R$, $Tc^R$, $Amp^R$) | This Study |

* $Kan^R$, Kanamycin resistant; $Tc^R$, tetracycline resistant; $Amp^R$, ampicillin resistant

**Table 2.3:** Primers used in this study

| Primer name | *Sequence (5' - 3') |
|---|---|
| 1742_MfeI_UF | atatcaattgtcctactcgaccaagttcaacgc |
| 1742_NdeI_UR | tatacatatgaggatgtagcggagcaccacg |
| 1745_AgeI_DF | atataccggtctacgcctcgcagaagcagacc |
| 1745_SacI_DR | atatgagctcttcgagcacgtcgatgaacagca |
| 1790_MfeI_UF | atatcaattggtgtcatgcgaagggattggaag |
| 1790_NcoI_UR | atatccatggtgccagaaatgccgaag |
| 1790_AgeI_DF | atataccggttcgatgcagtaccgcaagtggtt |
| 1790_SacI_DR | tatagagctcagacgaccctctcctggttcgat |
| 1782_MfeI_UF | atatcaattggacaggcccatctcagagcgaaa |
| 1782_NcoI_UR | atatccatggccgaacccacacgatgtcgtac |
| 1782_AgeI_DF | atataccggtcaatatcggaaatggctgcgtga |
| 1782_SacI_DR | atatgagctcgctgctggtcgataggtcaacga |
| 1777_MfeI_UF | atatcaattgttccgatgaggacgatgggtttc |
| 1777_NcoI_UR | atatccatggtgtcgctcaaggtcgtgaagatg |
| 1777_AgeI_DF | tataaccggtagacctacggcgtctcccct |
| 1777_SacI_DR | atatgagctcccttgtagggcgtcaggtcgatg |
| 1781_MfeI_UF | atatcaattgtcaggaccatcggagggacttc |
| 1781_NcoI_UR | atatccatgggccatcctcagggactagct |
| 1781_AgeI_DF | atataccggtggcggagaggaaggccaataaga |
| 1781_SacI_DR | atatgagctccgcgctcgatcagaagagtgacc |
| 1776_MfeI_UF | tatacaattgaagtccgctacggccacatcttc |
| 1776_NcoI_UR | atatccatgggcactcgctcgacagcacgtagt |
| 1776_AgeI_DF | atataccggtgcatcgacctgacgccctacaag |
| 1776_SacI_DR | tatagagctcagccctccatctcgcggtaattc |
| 1790C_KpnI_G290fw | ctggtctgtttgtggtacctgtaaccgcgatcaagaattgc |
| 1790C_MluI_G290rv | gctcgcatgcactagtacgcgtcaacgcagctggtctttggttc |
| 1782C_KpnI_290fw | atatggtacccagcgtcgagtctagcgtc |
| 1782C_AgeI_290rv | atagaccggtagcgcgtccaagtcgattctc |
| 1777C_KpnI_G290fw | ctggtctgtttgtggtacctcaattcacccgctctcaatggt |
| 1777C_MluI_G290rv | gctcgcatgcactagtacgcgtttccgcccctatctcaagcc |
| 1781C_KpnI_G290fw | ctggtctgtttgtggtacccttcgaggcgaggagcatgag |
| 1781C_MluI_G290rv | gctcgcatgcactagtacgcgtctggatctcggctgtttcgac |
| 1776C_KpnI_290fw | atatggtaccgtctcccctctggcttgagata |
| 1776C_AgeI_290rv | atataccggttgccagccatcgataatcctacc |

* Underlined sequences indicate restriction sites. The PCR product generated using primers with no underlined sequences were cloned using Gibson cloning.

PINKEL was swarmed several times in minimal salts base (MSB, see below) soft agar medium containing 0.25 mM caffeine as a carbon source, 0.1X Hutner's mineral base (trace elements solution) and 0.3% noble agar (Figure 2.2) [Hardwood et al., 1994; Ditty and Parales, 2015]. This strategy allowed the enrichment of a highly motile population of PINKEL cells (PINKELsw for swarmed PINKEL). One difficulty with working with certain pink-pigmented methylotrophic bacteria is their tendency to form clumps under certain growth conditions. This makes measuring the growth of cultures difficult. Additionally, performing chemotaxis assays would not be possible since studying the swimming behavior of bacteria requires free swimming cells. Other researchers have found that deleting three genes suggested to be involved in cellulose synthesis reduced or eliminated the clumping phenotype in *Methylobacterium extorquens* AM1 and PA1 [Delaney et al., 2013]. To determine if the same non-clumping phenotype was replicable in PINKEL, we deleted the gene cluster containing the three homologous genes that encode the cellulose synthase: *celA* (peg_1742), *celB* (peg_1743), *celC* (peg_1744) and *celD* (peg_1745). Deletion of this gene cluster was done in cells obtained from the highly motile PINKEL population (PINKELsw). A reduction in the amount of clumping was observed and a freezer stock of the resulting strain was saved and used as the wild-type strain (PINKELnc for non-clumping PINKEL) for all the experiments performed in this study. All mutants generated in this study were derivatives of this strain as well.

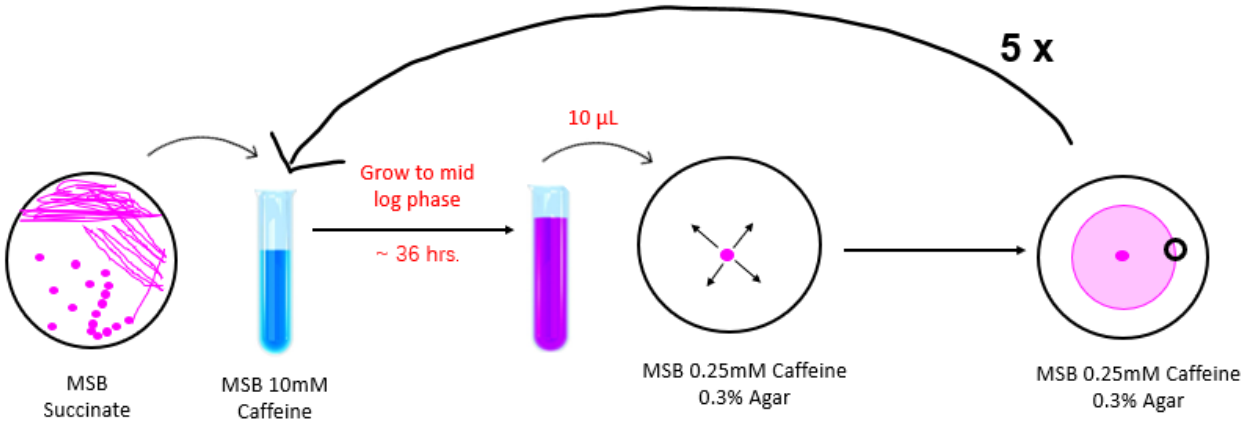



**Figure 2.2**: Growth in soft agar swim plates to enrich for a population of highly motile *M. populi* PINKEL cells. A seed culture of PINKEL was grown in 10 mL of MSB minimal medium containing 10 mM caffeine. Once the culture reached late exponential phase ($OD_{600}$ ~ 0.9) 10 µL was pipetted into the center of an MSB soft agar (0.3% Noble agar) Petri plate containing a low concentration of caffeine (0.25 mM). A chemotactic response is visualized as a sharp pink ring of growth that formed as cells multiplied in number and swam away from the middle following the gradient that was generated as the cells metabolized the caffeine. Once the cells moved 20 mm away from the point of inoculation (in ~ 10 days), the wide end of a sterile P1000 pipette tip was used to obtain a plug of agar containing cells from the edge of outer ring. The plug was immediately inoculated into 10 mL of MSB containing 10 mM caffeine. This enrichment process was repeated five times after which the populations swam 20 mm in 2 days rather than 10 days.

## Culture media and growth conditions

*Escherichia coli* strains were grown aerobically in Lysogeny broth (LB) or on LB agar plates at 37°C (5 g Bacto tryptone, 2.5 g yeast extract, 2.5 g NaCl with or without 15 g Bacto Agar per 1 L distilled water). *Methylorubrum populi* PINKEL was grown aerobically at 30°C in a phosphate buffered minimal medium (MSB) [Stanier et al., 1966] with the following modifications: 1. The Hutner's base (trace elements solution) stock solution was filter sterilized and added to the medium after autoclaving and 2. The ammonium sulfate solution was omitted when testing whether PINKEL could use methylxanthines as sole nitrogen sources. MSB was supplemented with 10 mM

succinate, 124 mM methanol, 5- or 10-mM caffeine, 5 mM theobromine, 5 mM paraxanthine, 5 mM 7-methylxanthine or a combination of 5 mM succinate and 5 mM methylxanthine or increasing concentrations of formaldehyde as indicated in the text or figure legends. To select and maintain plasmids in *M. populi* and *E. coli* strains, the following concentrations of antibiotics were added when appropriate: 10 µg/ml tetracycline, 50 µg/mL kanamycin, or 100 µg/ml ampicillin. For complementation studies, cumate at a final concentration of 30 µg/mL was added to the medium to induce expression of the gene cloned downstream of a cumate-inducible promoter in the complementation plasmid pLC290 (described in more detail below). During the procedures to generate kanamycin insertion mutants, 10% sucrose was added to counter-select against cells containing a single crossover (procedure described in more detail below) [Chubiz et al., 2013]. Liquid cultures were incubated at 200 rpm in a shaking incubator.

## Cloning and DNA manipulations

All molecular biology kits mentioned below were purchased from Thermo Fisher Scientific (Chicago IL). Genomic DNA from *M. populi* PINKEL was isolated using the PureLink Genomic DNA Mini Kit, and plasmids were purified using a GeneJet plasmid Miniprep Kit. Manipulations of DNA fragments, plasmids, and transformation of *E. coli* strains were carried out by standard methods [Sambrook et al., 1989]. All PCR amplifications were performed using Pfu polymerase in Pfu reaction buffer (200 mM Tris-Cl (pH 8.8), 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 1% Triton X-100, 1 mg/ml bovine serum albumin, 20 mM $Mg_2SO_4$) under the following conditions: 95°C denaturation, 65-

60°C annealing (decreasing by 1°C for the first 5 cycles), 72°C elongation, with an elongation time of 1 min/kb of PCR product. All PCRs were set-up for a total of 30 cycles with the initial denaturation and final extension periods each being 10 minutes. Since the genome of PINKEL has a high G+C content, a final concentration of 3% DMSO was added to each PCR as described for other pink-pigmented methylotrophic bacteria [Vu et al., 2021]. PCR products were purified using a GeneJet Gel Extraction Kit and amplicons digested with restriction endonucleases were purified using a GeneJet PCR Purification Kit. Restriction endonucleases used to digest each PCR product are specified in the primer list (Table 2). Some amplicons were cloned into their corresponding plasmid following the recommendations stated in the NEB builder HIFI DNA assembly cloning kit manual. Gene manipulations were corroborated via standard colony PCR procedures using the Promega GoTaq Green (2x) Master Mix. The sequences of cloned PCR products were verified by fluorescent automated DNA sequencing using the University of California Davis DNA Sequencing Facility with an Applied Biosystems 3730 automated sequencer.

## Construction of mutant strains

The method used to generate all markerless gene deletion mutants listed in Table 1 consisted of two stages (Figure 2.3): 1. Generation of a kanamycin insertion mutant via an allelic exchange mechanism, and 2. Deletion of the kanamycin resistance cassette introduced during the first stage. The allelic exchange suicide plasmid pHV2 (Figure 2.4 A) [Roszczenko-Jasinska et al., 2020] was used during the first stage to generate the kanamycin insertion mutants (Figure 2.3 A). This vector has a *sacB* gene

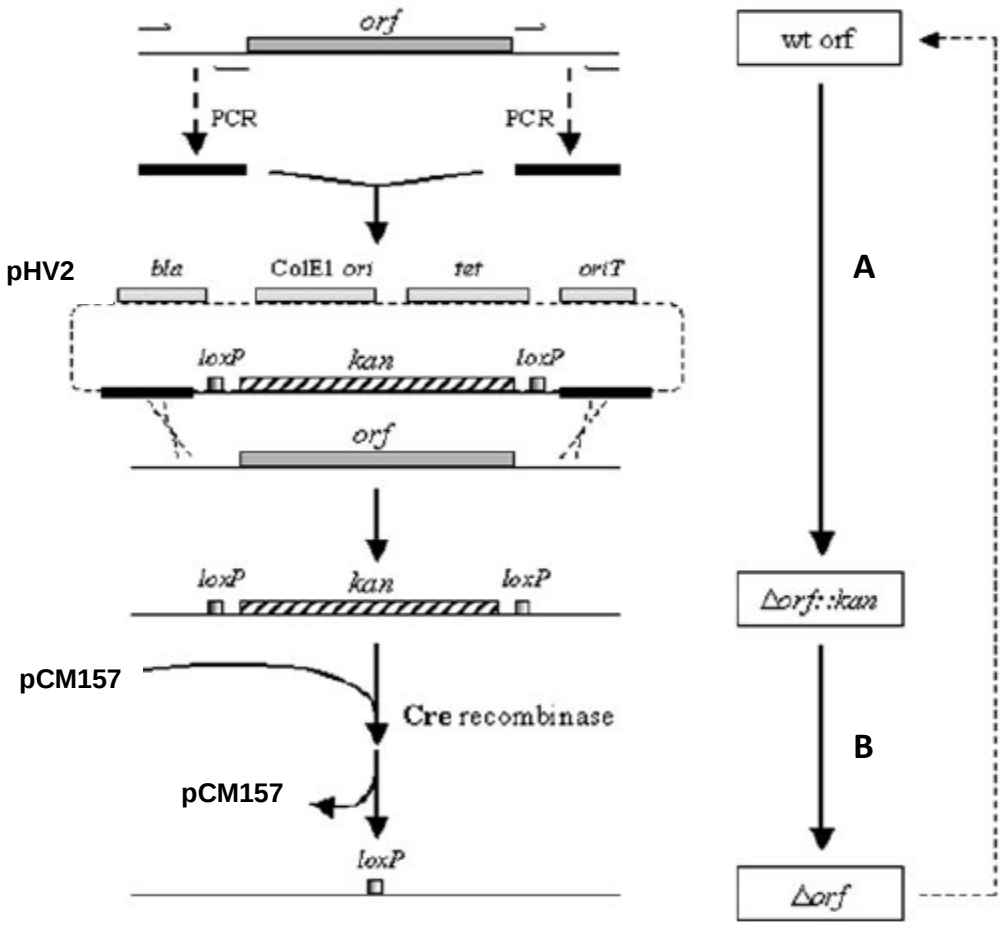



**Figure 2.3**: Strategy to generate markerless gene deletions in *M. populi*. (A) DNA fragments upstream and downstream of the target gene are amplified by PCR and cloned into pHV2. Allelic exchange leads to a kanamycin insertion mutant (B), which can then be rendered unmarked through the introduction of the *cre* expression plasmid pCM157. The process can then be repeated for a second target gene to generate a strain bearing multiple genetic manipulations.

Source: adapted from Marx and Lidstrom, 2002

to allow for counter-selection against single crossovers and contains a kanamycin resistance cassette flanked by two *loxP* sites, which are flanked by two multiple cloning sites. The upstream and downstream regions of each target gene were amplified via PCR, cloned into pHV2, and the resulting plasmid was used to transform *E. coli* S17-1. The steps necessary to mate the engineered plasmid into PINKEL and select for cells

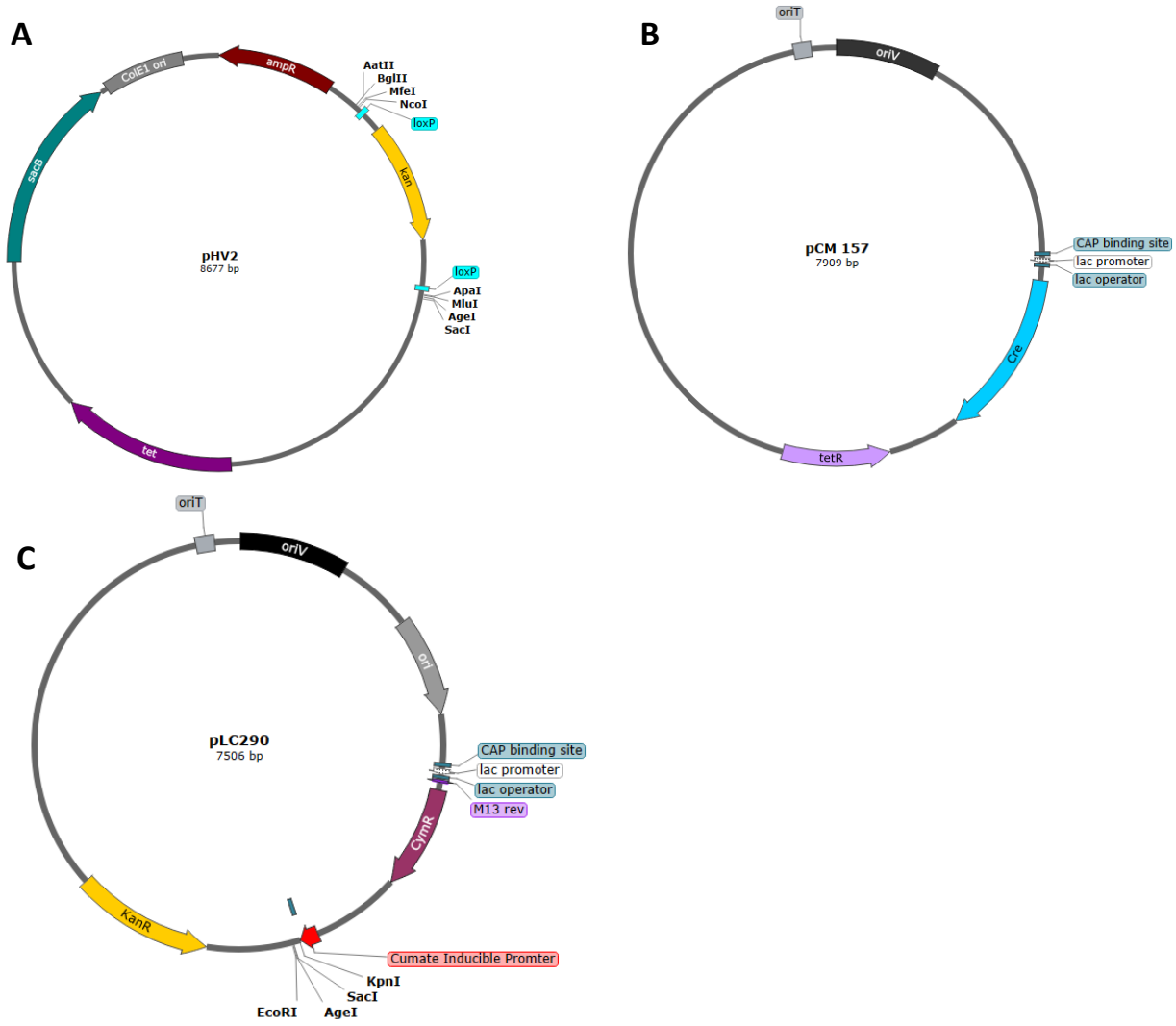



**Figure 2.4**: Vector maps of plasmids used in this study. (A) pHV2, (B) pCM157, and (C) pLC290

that underwent allelic exchange that resulted in a kanamycin insertion mutant (Figure 2.5) are similar to those previously described [Marx and Lidstrom, 2002; Vu et al., 2021] with the following modifications: 1. A final concentration of 10% sucrose was used to select against cells containing single crossovers and 2. Methanol was used as the sole carbon source to select for PINKEL cells after the biparental mating step instead of selecting for rifamycin resistance. To remove the kanamycin resistance cassette and generate makerless deletion mutants during the second stage (Figure 2.3 B), the

tetracycline resistance plasmid pCM157 (Figure 2.4 B) was used [Marx and Lidstrom, 2002]. This vector encodes the Cre recombinase, which is necessary to catalyze the recombination of the *loxP* sites flanking the kanamycin resistance cassette. This recombination leads to the unmarked removal of the kanamycin resistance cassette. The steps necessary to select for gene deletion strains are summarized in Figure 2.6.

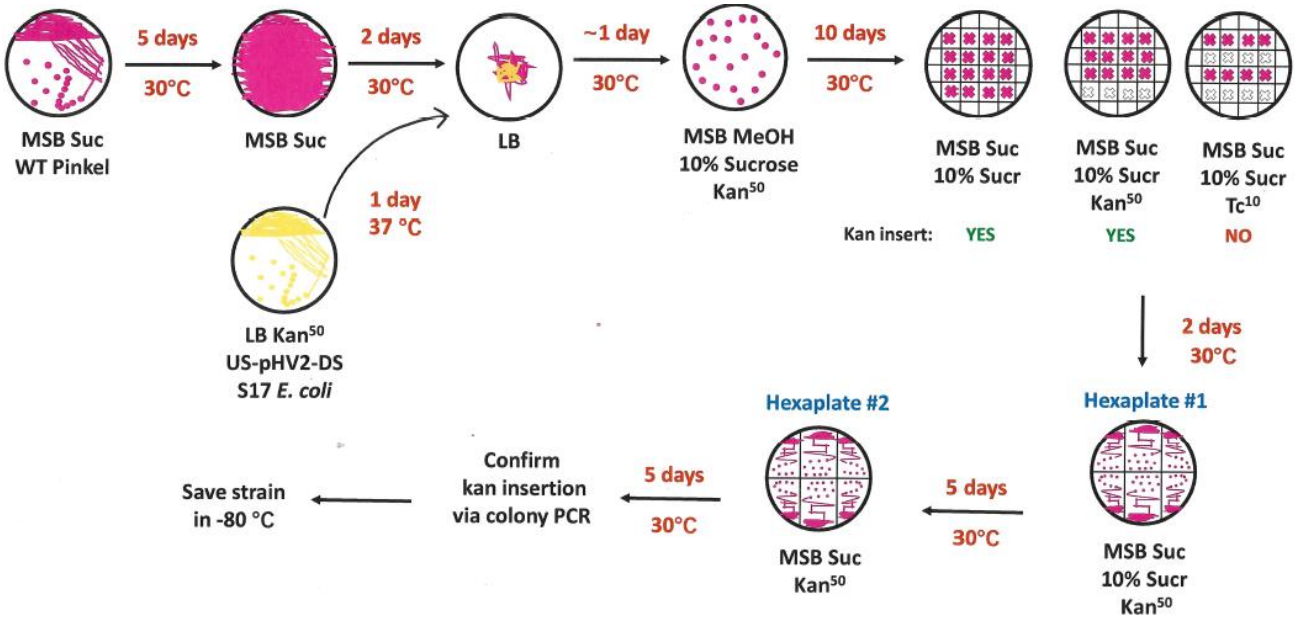



**Figure 2.5**: Method for mating pHV2 plasmids into *M. populi* to select for cells that underwent allelic exchange leading to kanamycin insertion mutants. A few swabs of *M. populi* cells from a biomass plate were mated with one colony of *E. coli* S17-1 [containing an upstream-pHV2-downstream (US-pHV2-DS) plasmid engineered to knock out a target gene] on an LB agar plate. The cells were allowed to mate for 1 day (~24 h) at 30°C. An MSB agar plate containing methanol (1% MeOH), 10% sucrose, and 50 µg/ml kanamycin ($kan^{50}$) was used to select for *M. populi* cells with a kanamycin insertion mutation. Candidate colonies were patched onto MSB 10 mM succinate 10% sucrose, MSB 10 mM succinate 10% sucrose 50 µg/ml kanamycin, and MSB 10 mM succinate 10% sucrose 10 µg/ml tetracycline plates to screen for double crossovers and then passed onto fresh MSB succinate 50 µg/ml plates until a pure culture was obtained. Diagnostic colony PCR was performed to verify that *M. populi* cells had a kanamycin cassette inserted within the target gene before saving a freezer stock.

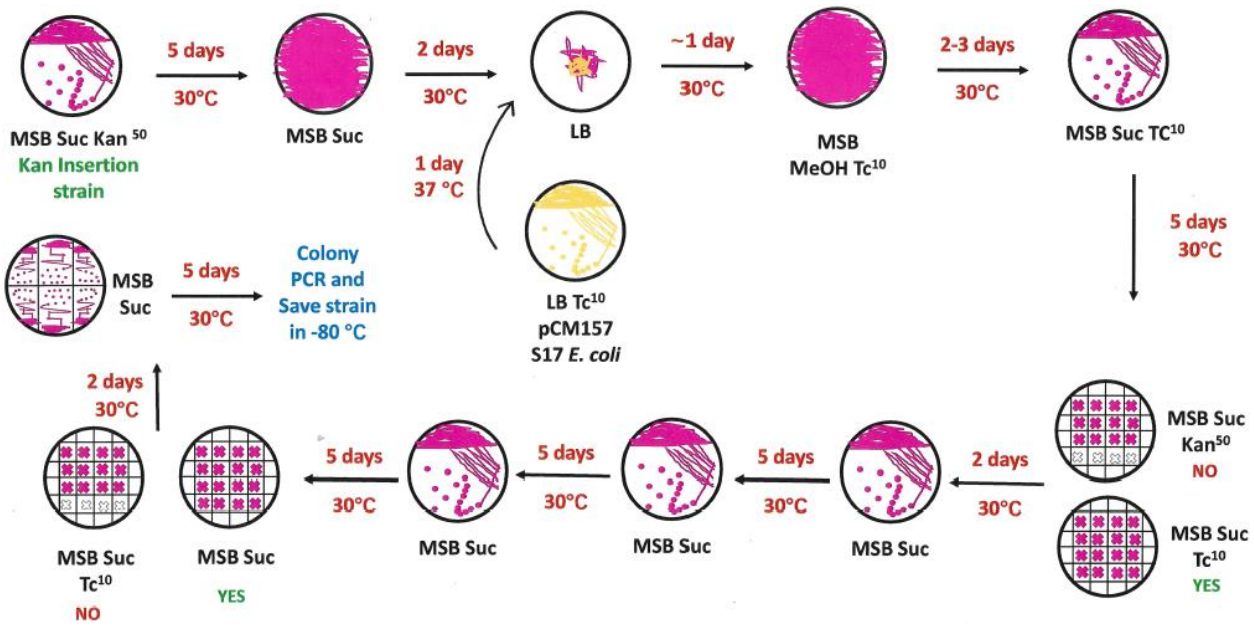



**Figure 2.6**: Method for resolving the insertion of the kanamycin resistance cassette using pCM157 to achieve makerless deletions. Mating between the *M. populi* strains bearing a kanamycin resistance cassette insertion and *E. coli* S17-1 containing the pCM157 plasmid were done on LB agar plates. The cells were allowed to mate for 1 day at 30°C. An MSB agar plate containing methanol (1% MeOH) and 10 µg/ml tetracycline ($Tc^{10}$) was used to select for *M. populi* cells that received the pCM157 plasmid. This plasmid carries the gene encoding the Cre recombinase, which is necessary to catalyze the recombination of the *loxP* sites flanking the kanamycin resistance cassette. The recombination leads to the unmarked removal of the kanamycin resistance cassette. Colonies were patched onto MSB succinate plates containing either kanamycin or tetracycline to screen for colonies that had lost the kanamycin resistance cassette. Kanamycin sensitive, tetracycline resistant cultures were passed several times on MSB succinate plates with no antibiotic to cure the pCM157 plasmid. Curing of the plasmid was confirmed by patching colonies onto MSB succinate plates with and without tetracycline. Tetracycline-sensitive cultures were streaked onto fresh MSB succinate plates and diagnostic colony PCR was performed to confirm that markerless gene deletion was achieved before saving strains as freezer stocks.

## Construction of complementation strains

To complement mutants with a functional copy of a deleted gene, the vector pLC290 was used (Figure 2.4 C) [Chubiz et al., 2013]. This plasmid carries a kanamycin resistance gene and has a multiple cloning site immediately downstream of a cumate-inducible promoter. The KpnI and MluI restriction sites were used to linearize the

pLC290 plasmid and an NEB builder HIFI DNA assembly cloning kit was used to clone the PCR-amplified genes used for complementation (complete sequences of these genes are provided in Appendix A). All PCR products were sequenced before cloning to ensure there were no point mutations present. The protocol shown in Figure 2.7 was used to introduce the complementation plasmids into *M. populi* mutants and kanamycin was used to maintain the plasmid. Empty vector (pLC290 with no cloned gene) was mated into wild-type PINKEL to serve as a control.

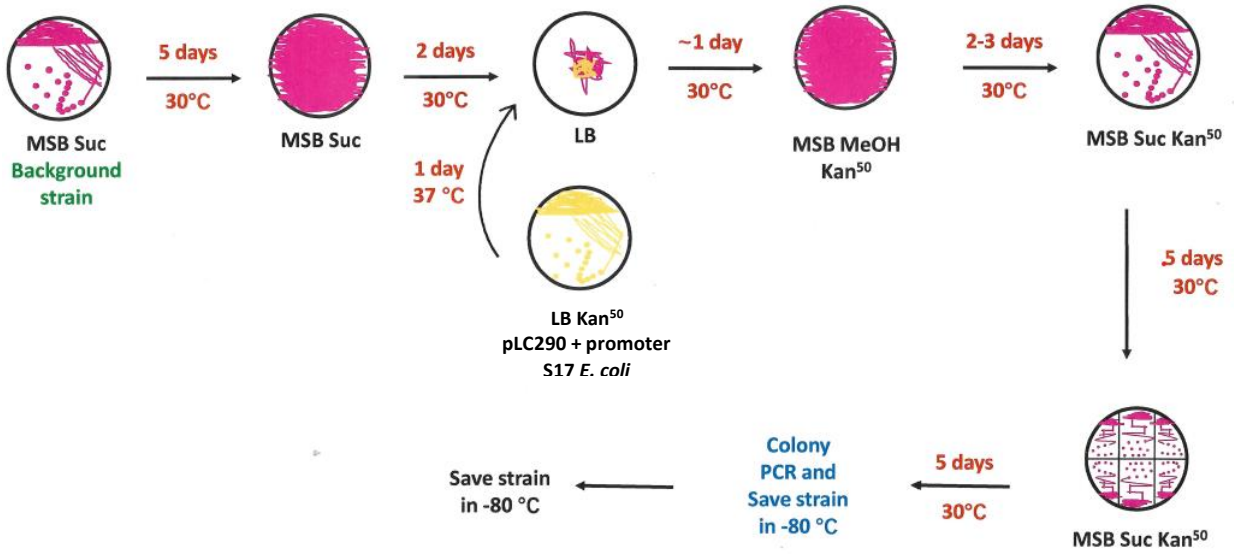



**Figure 2.7**: Method for mating pLC290 complementation plasmids into the desired PINKEL background strain. A few swabs of PINKEL cells from a biomass plate were inoculated together with one colony of *E. coli* S17-1 containing pLC290 carrying the cloned gene of interest on a LB agar plate. The cells were allowed to mate for 1 day (~24 h) at 30°C. An MSB agar plate containing methanol (1% MeOH) and 50 ug/ml Kanamycin ($Kan^{50}$) was used to select for *M. populi* cells that received the complementation plasmid. Isolated colonies were passed onto a fresh MSB MeOH $Kan^{50}$ plate and then onto an MSB succinate (Suc) $Kan^{50}$ plate until a pure culture was obtained. Diagnostic colony PCR was performed to verify that *M. populi* cells had the complementation plasmid before saving a freezer stock.

## Determination of growth rates

Growth experiments were conducted at 30°C in borosilicate glass tubes (20 x 150 mm) using a minimum of three biological replicates. Tubes were placed in test tube racks at an approximate angle of 60° to ensure maximum aeration. The cultures were shaken at 30 ºC, 200 rpm using a shaking incubator. To obtain a seed culture, *M. populi* cells were grown to late exponential phase in 6 ml of MSB medium containing 10 mM succinate as the carbon source. Approximately 200 µl of the seed culture was transferred into 10 ml of MSB plus 5 mM of the test methylxanthine. Optical density at 600 nm ($OD_{600}$) was recorded every three hours using a Spectronic 20D spectrophotometer containing an adapter to read absorbance in 20 x 150 mm tubes. Measurements were taken until three consecutive measurements in the stationary phase were recorded.

## Whole cell biodegradation assay

Degradation of caffeine and related methylxanthines by *M. populi* was investigated by adding these compounds to MSB medium as growth substrates and monitoring metabolite formation (Figure 2.8). A seed culture of *M. populi* was first grown as described above. The seed culture was then used to inoculate 50 mL of MSB medium supplemented with the test methylxanthine to an initial $OD_{600}$ of approximately 0.03-0.05. All cultures were incubated at 30°C in a shaking incubator set at 200 rpm. Growth of *M. populi* was monitored periodically by measuring the $OD_{600}$ of each culture and 500 µl aliquots were taken every 3 hours. The collected samples were immediately centrifuged (10,000 rpm for 5 minutes at 4°C) and the supernatant (spent medium) was

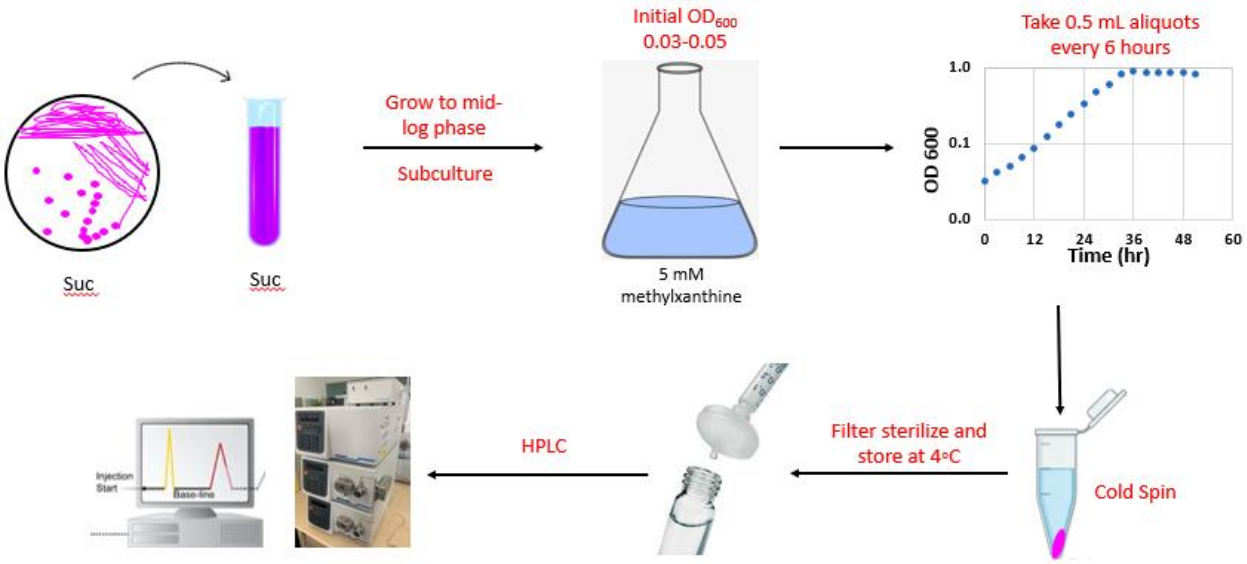



**Figure 2.8**: Whole cell biodegradation assay workflow to monitor the degradation of methylxanthines and formation of the corresponding metabolites.

filtered into a glass vial using 0.2 µm Millex hydrophilic PTFE syringe filters (Millipore Sigma, St. Louis, MO). The filtered spent medium was then analyzed via high performance liquid chromatography (HPLC) to monitor the degradation of methylxanthines and formation of the corresponding metabolites. Identities of the metabolites were established by comparing the retention times and absorption spectra of the metabolites with those of methylxanthine standards.

## Analysis of metabolites

Identification and quantification of caffeine and related methylxanthines were performed with a System Gold HPLC equipped with a photodiode array detector and 32 Karat Software version 5 (Beckman Coulter, Fullerton, CA). Test compounds were separated using a C18 column (4.5 x 150 mm) and detection at 271 nm. The mobile

phase used was composed of water, methanol, and acetic acid (750 ml water, 250 ml methanol, 100 µl acetic acid) at a flow rate of 0.5 ml per minute. A calibration curve for each methylxanthine was generated in order to determine the concentration of each detected metabolite based on the percent area under each peak.

## Results

### *M. populi* PINKEL carries genes encoding predicted *N*-demethylation proteins with the expected conserved domains

*In silico* comparison of the *Pseudomonas putida* CBB5 *ndm* genes and the genome of *Methylorubrum populi* PINKEL identified genes that are predicted to encode enzymes of the *N*-demethylation pathway with amino acid percent identities ranging from 41% to 65% (Figure 2.9). In contrast to *P. putida* CBB5, which has a single gene cluster encoding all the enzymes necessary to convert caffeine to xanthine, the *ndm* genes in PINKEL are located in two separate clusters. The first cluster is approximately 3 kb in length and contains the *ndmA* gene ($N_1$-demethylase; peg_1790). The second cluster measures approximately 11.2 kb and contains the remaining four *ndm* genes: *ndmB* ($N_3$-demethylase; peg_1782), *ndmC* ($N_7$-demethylase; peg_1777), *ndmD* (reductase; peg_1781), and *ndmE* (peg_1776).

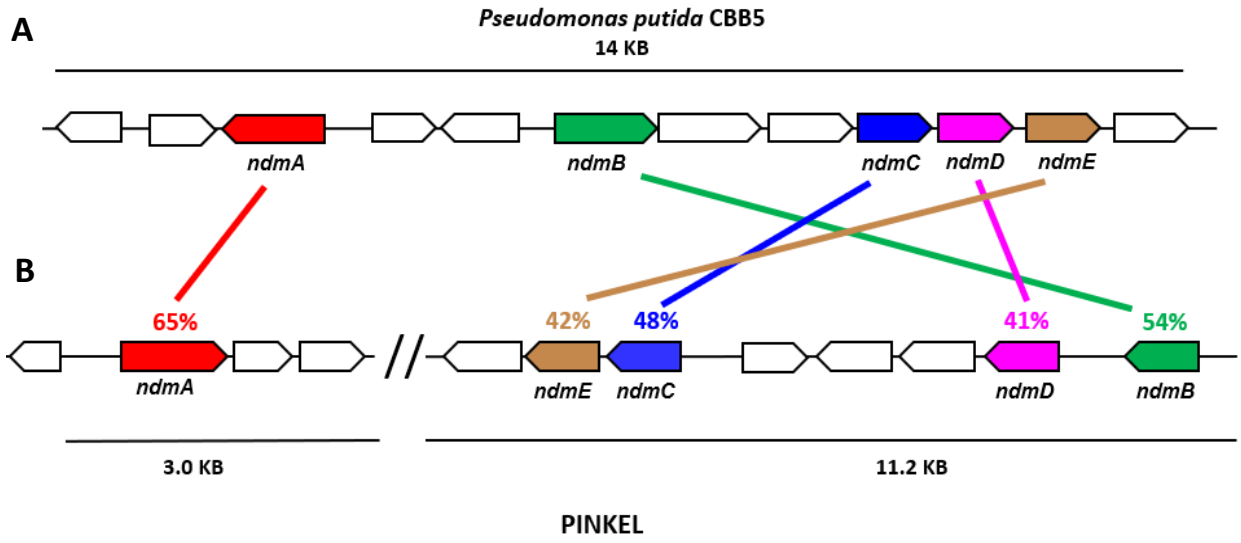



**Figure 2.9**: Genetic organization of *ndm* genes in **(A)** *Pseudomonas putida* CBB5 and **(B)** *Methylorubrum populi* PINKEL. Amino acid percent identities are indicated.

To determine if the genes that were found could encode *N*-demethylases, the amino acid sequences were analyzed for the conserved domains expected in Rieske-type monooxygenases (Figure 2.10). Both NdmA and NdmB have an N-terminal Rieske-type [2Fe-2S] cluster domain with the following conserved residue motif: $CXHX_{16}CX_2H$. Additionally, both have a C-terminal mononuclear iron binding domain with the following conserved residue motif: $DX_2HX_4H$. Although NdmC does not have a Rieske-type [2Fe-2S] cluster domain, it does have the conserved C-terminal mononuclear iron binding domain. NdmD has four conserved domains: 1. a Rieske-type [2Fe-2S] cluster domain located at the N-terminus, 2. an FAD binding domain, 3. an NADH binding domain, and 4. a plant-type [2Fe-2S] cluster domain at the C-terminus. On the other hand, NdmE has two glutathione-S-transferase (GST) domains.

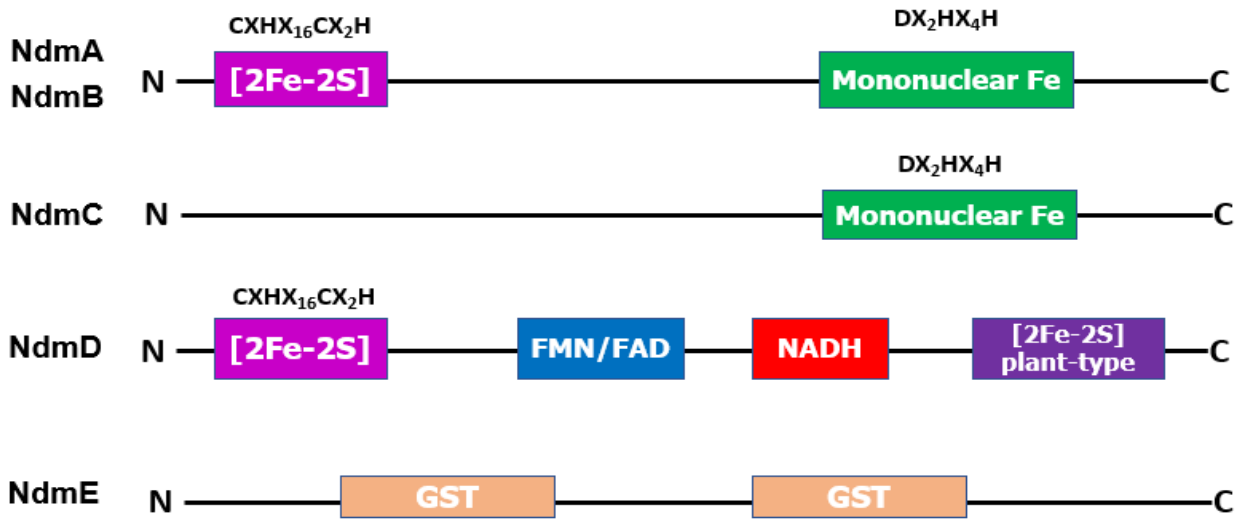



**Figure 2.10**: Conserved domains found in each of the five *N*-demethylase proteins from *M. populi* PINKEL.

## *M. populi* PINKEL can degrade caffeine faster than *Pseudomonas putida* CBB5 and grows on all *N*-demethylation pathway metabolites

To assess the growth of *M. populi* PINKEL in the presence of caffeine and determine how its growth compares with published data for *P. putida* CBB5, a minimal medium containing 10 mM caffeine as the sole carbon and nitrogen source was used. *M. populi* PINKEL reached a maximum cell density ($OD_{600}$) of approximately 1.2 after 36 hours of incubation and consumed 100% of the supplied caffeine within approximately 39 hours (Figure 2.11). In contrast, under similar growth conditions *P. putida* CBB5 was reported to reach a maximum $OD_{600}$ of 0.4 after 72 hours of incubation and took 8 days to completely consume 10 mM caffeine [Yu et al., 2009]. To assess whether the toxicity of the formaldehyde produced during each demethylation step could explain the differences in growth and caffeine consumption rate between the two strains, PINKEL was grown in the presence of increasing concentrations of formaldehyde (Figure 2.12).

PINKEL was able to grow at a similar growth rate in minimal methanol medium containing up to 8 mM formaldehyde (Figure 2.12). However, there was a longer lag phase when PINKEL was grown in the presence of formaldehyde concentrations greater than 2 mM. PINKEL grew in the presence of 10 mM formaldehyde, but at a slower rate; higher formaldehyde concentrations completely inhibited growth.

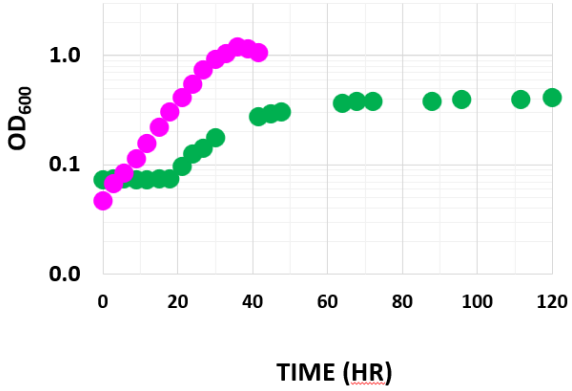
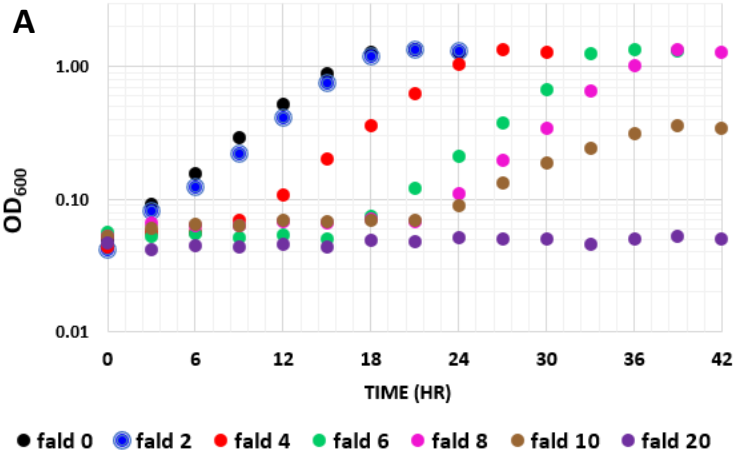



**Figure 2.11**: Growth of *M. populi* PINKEL (pink circles) and *P. putida* CBB5 (green circles) on 10 mM caffeine.



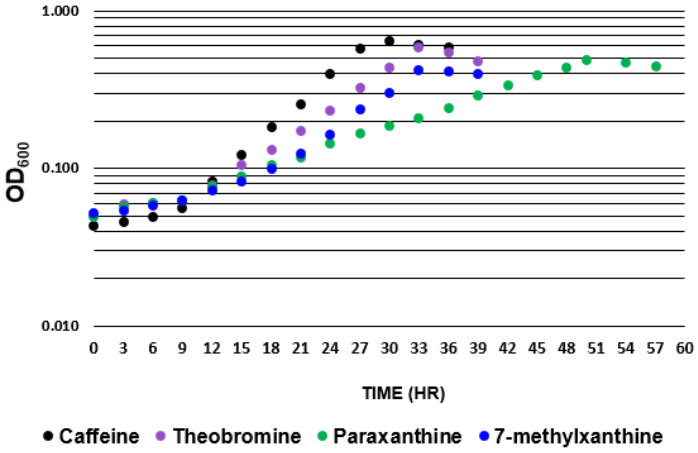

**B**

| Formaldehyde concentration (mM) | Growth Rate ($hr^{-1}$) | Doubling Time (hr) |
|---|---|---|
| 0 | 0.193 | 3.6 |
| 2 | 0.185 | 3.7 |
| 4 | 0.194 | 3.6 |
| 6 | 0.191 | 3.6 |
| 8 | 0.198 | 3.5 |
| 10 | 0.111 | 6.2 |
| 20 | No Growth | No Growth |

**Figure 2.12**: **(A)** Growth of the wild-type *M. populi* PINKEL strain on methanol plus increasing concentrations of formaldehyde. The different colored circles represent growth on the corresponding mM concentrations of formaldehyde **(B)** The growth rate and doubling time for each growth condition is also provided.

As an initial test to determine if *M. populi* PINKEL uses the *N*-demethylation pathway for caffeine degradation, the wild-type strain was grown in the presence of 5 mM of each of the methylxanthines of the *N*-demethylation pathway. In addition to caffeine, PINKEL grew on theobromine, paraxanthine, and 7-methylxanthine, although at slower growth rates (Figure 2.13). Due to their high insolubility in aqueous media, growth of PINKEL on xanthine and uric acid was only observed on plates when crystals of these carbon sources were added to the surface of a minimal agar plate containing no other carbon source.

**A**




**B**

| Methylxanthine | Growth Rate ($hr^{-1}$) | Doubling Time (hr) |
|---|---|---|
| Caffeine | 0.129 | 5.4 |
| Theobromine | 0.094 | 7.3 |
| Paraxanthine | 0.049 | 14.0 |
| 7-Methylxanthine | 0.096 | 7.2 |

**Figure 2.13**: (A) Growth of the wild-type *M. populi* PINKEL strain on 5 mM caffeine (black circles), theobromine (purple circles), paraxanthine (green circles), and 7-methylxanthine (blue circles). **(B)** Growth rates and doubling times for PINKEL grown with methylxanthines.

## Deletion of each *ndm* gene affects the growth of *M. populi* PINKEL on caffeine

To evaluate the requirement of the *ndm* gene products for growth on caffeine, each predicted *ndm* gene was individually deleted and mutant strains were grown in the presence of 5 mM caffeine. Deleting *ndmA, ndmB* or *ndmD* completely eliminated the

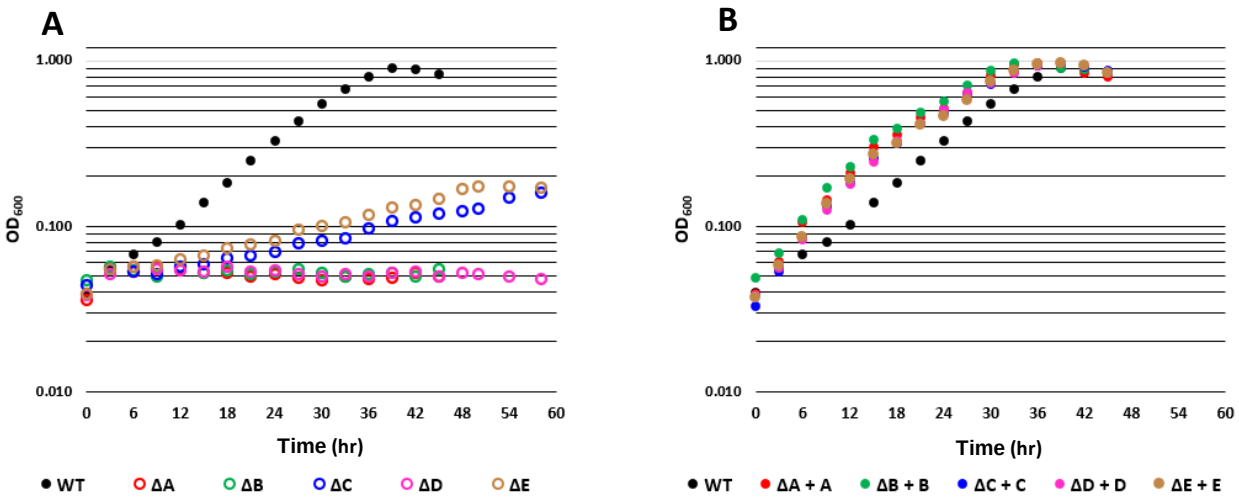



**Figure 2.14**: (A) Growth of the different Δ*ndm* mutants and **(B)** the Δ*ndm* mutant strains complemented with *ndm* genes in minimal medium containing 5 mM caffeine. The solid circles represent either wild-type *M. populi* PINKEL or the complemented deletion mutants. Open circles represent the deletion mutants. The wild-type strain was complemented with the empty vector. Kanamycin was added to medium used to grow complemented strains to select for plasmid maintenance.

ability of *M. populi* PINKEL to grow on caffeine (Figure 2.14A). On the other hand, the *ndmC* and *ndmE* deletion mutants had significantly lower growth rates and growth yields (maximum $OD_{600}$) on caffeine compared to the wild-type strain (Figure 2.14A). To ensure that the growth phenotypes were not due to polar effects or a spontaneous mutation elsewhere in the genome, each mutant was complemented with a functional copy of the corresponding deleted gene. As shown in Figure 2.14B, complementation with the relevant *ndm* gene restored the ability of each Δ*ndm* mutant to grow on caffeine.

Caffeine degradation intermediates correspond to *N*-demethylation metabolites

In addition to evaluating the ability of wild-type *M.* populi PINKEL to grow on methylxanthines, aliquots of spent medium were collected as the cells grew and were analyzed for metabolite formation via HPLC. When caffeine was the initial substrate, five metabolites were detected. Based on the retention times and absorption spectra of each metabolite as compared to methylxanthine standards, the metabolites were identified as theobromine, paraxanthine, 7-methylxanthine, xanthine, and uric acid (Figure 2.15A). When theobromine or paraxanthine were provided as the initial substrate, only 7-methylxanthine, xanthine, and uric acid were detected as metabolites (Figure 2.15B and Figure 2.15C). Xanthine and uric acid were the only metabolites detected in methylxanthine-grown cultures (Figure 2.15D).

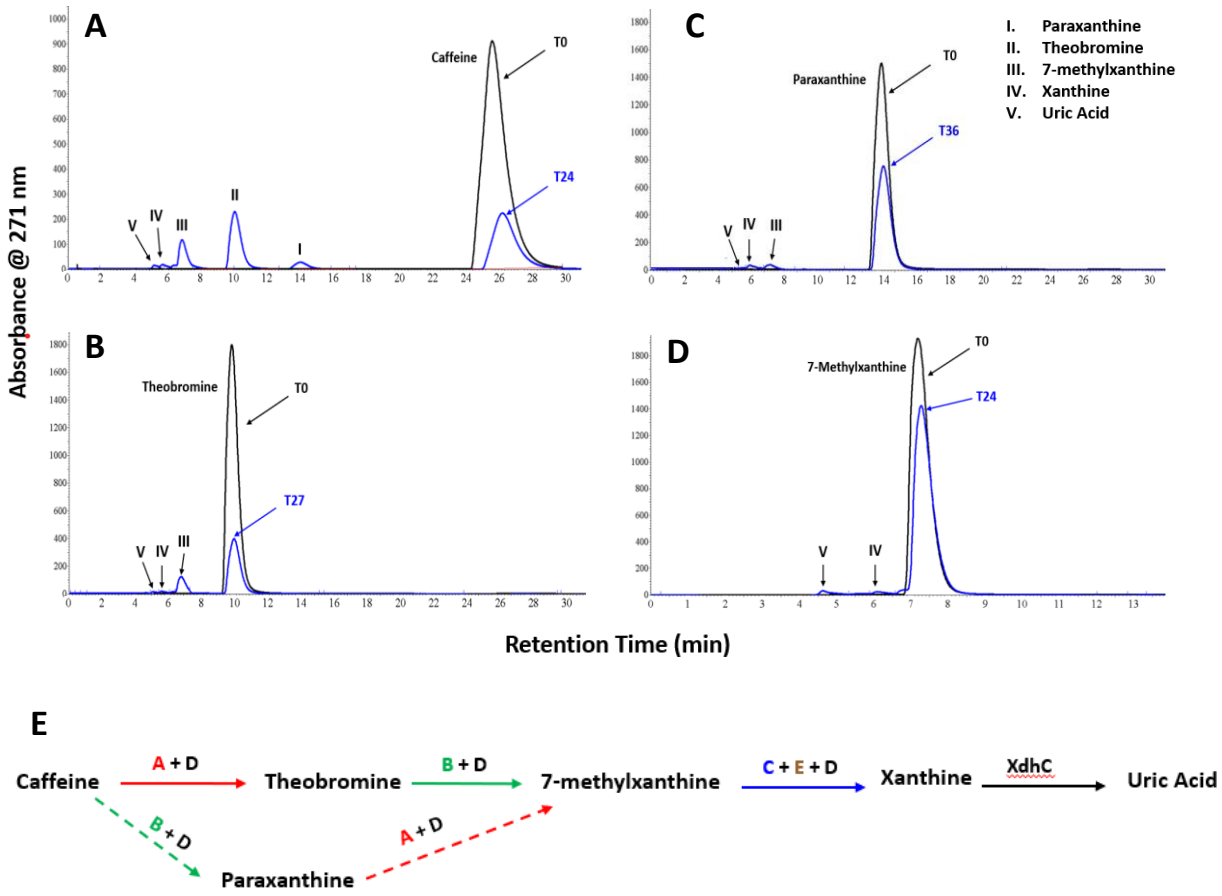



**Figure 2.15**: HPLC chromatograms showing the metabolites formed when wild-type *M. populi* PINKEL was grown on 5 mM **(A)** caffeine, **(B)** theobromine, **(C)** paraxanthine, or **(D)** 7-methylxanthine. Black peaks represent the samples collected at T0 (the beginning of the experiment) and blue peaks represent the samples collected at the indicated hours after the experiment started. **(E)** A map of the *N*-demethylation pathway is provided as a reference. Solid arrows represent the major pathway, and dashed arrows represent the minor pathway. Letters above each arrow indicate the *N*-demethylase proteins that were proposed to catalyze each step in *P. putida* CBB5 [Summers et al., 2015].

Each identified *N*-demethylase plays a role in the *N*-demethylation pathway

To determine which step of the pathway is catalyzed by each *N*-demethylase, the accumulation of metabolites was assessed using each of the five Δ*ndm* mutant strains as they grew in the presence of the methylxanthine intermediates of the pathway. No

growth was observed, and no metabolites were detected when caffeine or paraxanthine were provided as the initial substrate to a Δ*ndmA* strain (Figure 2.16A and Figure 2.16B). However, when theobromine was included in the growth medium as the initial substrate, 7-methylxanthine, xanthine, and uric acid were detected as metabolites (Figure 2.16C). Theobromine and trace amounts of paraxanthine, 7-methylxanthine, xanthine, and uric acid were produced as metabolites when caffeine was the initial substrate for an Δ*ndmB* strain (Figure 2.17A). No metabolites were detected when theobromine was the initial substrate (Figure 2.17B); however, the corresponding downstream metabolites were detected when 7-methylxanthine was provided (Figure 2.17C). Theobromine, paraxanthine, and 7-methylxanthine were detected when the Δ*ndmC* and Δ*ndmE* mutants were grown on caffeine (Figure 2.18). No growth or metabolites were detected when the Δ*ndmD* mutant was grown with any of the methylxanthines (Figure 2.19A through Figure 2.19C).

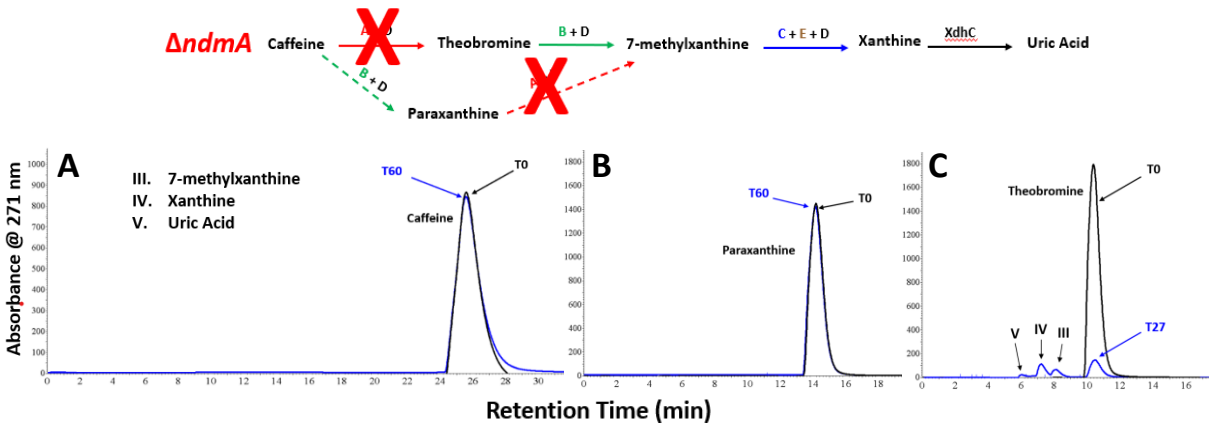



**Figure 2.16**: HPLC chromatograms showing the metabolites formed when the Δ*ndmA* mutant was grown on 5 mM **(A)** caffeine, **(B)** paraxanthine, or **(C)** theobromine. Black peaks represent samples collected at T0 (the beginning of the experiment) and blue peaks represent samples collected at the indicated hours after the experiment started. A map of the *N*-demethylation pathway is provided as a reference. The red X's represent the steps of the pathway expected to be blocked in the Δ*ndmA* mutant.

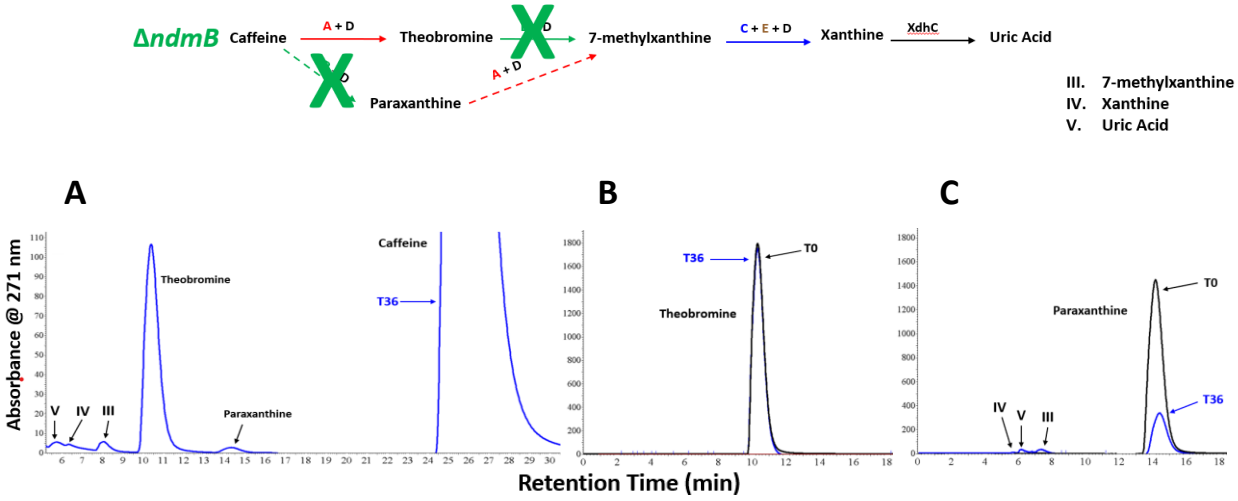



**Figure 2.17**: HPLC chromatograms showing the metabolites formed when the Δ*ndmB* mutant was grown on 5 mM **(A)** caffeine, **(B)** theobromine, or **(C)** paraxanthine. Black peaks represent samples collected at T0 (the beginning of the experiment) and blue peaks represent samples collected at the indicated hours after the experiment started. A map of the *N*-demethylation pathway is provided as a reference. The green X's represent the steps of the pathway expected to be blocked in the Δ*ndmB* mutant.

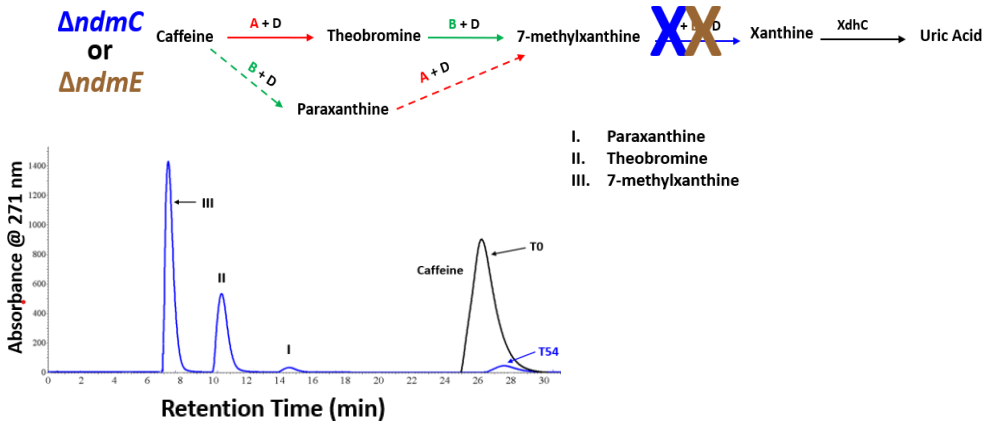



**Figure 2.18**: HPLC chromatogram showing the metabolites formed when the Δ*ndmC* or Δ*ndmE* mutant was grown on 5 mM caffeine. The black peak represents the sample collected at T0 (the beginning of the experiment) and the blue peaks represent the sample collected 54 hours after starting the experiment. A map of the *N*-demethylation pathway is provided as a reference. The blue and brown X's represent the steps of the pathway expected to be blocked in the Δ*ndmC* and Δ*ndmE* mutants. Data for the Δ*ndmC* mutant is shown; results were the same for the Δ*ndmE* mutant.

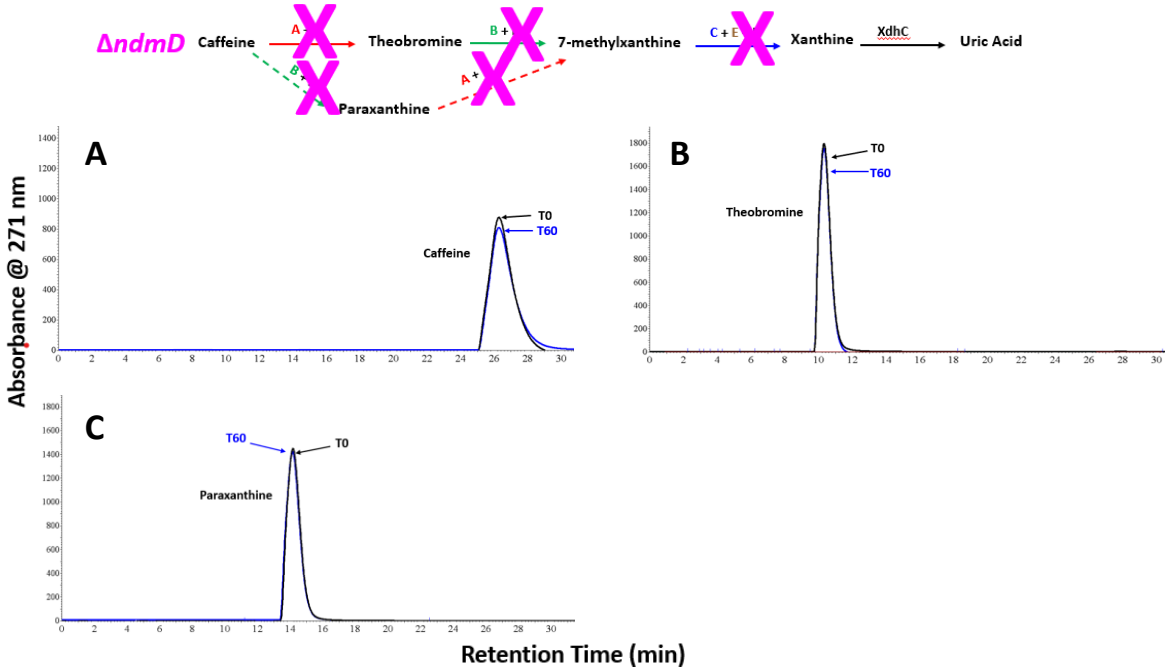



**Figure 2.19**: HPLC chromatograms showing the metabolites formed when the Δ*ndmD* mutant was grown on 5 mM **(A)** caffeine, **(B)** theobromine, or **(C)** paraxanthine. Black peaks represent samples collected at T0 (the beginning of the experiment) and blue peaks represent samples collected 60 hours after the experiment started. A map of the *N*-demethylation pathway is provided as a reference. The pink X's represents the steps of the pathway expected to be affected in the Δ*ndmD* mutant.

## Discussion

### Caffeine and related methylxanthines are catabolized via the *N*-demethylation pathway in *Methylorubrum populi* PINKEL

The only known caffeine-degrading methylotrophic bacterium, *M. populi* PINKEL, was isolated from compost containing coffee grounds and can grow not only on caffeine, but also on theobromine, paraxanthine, 7-methylxanthine, xanthine, and uric acid as sole carbon, nitrogen, and energy sources. Of all tested methylxanthines, *M. populi* PINKEL had the fastest growth rate on caffeine followed by 7-methylxanthine, theobromine and paraxanthine (Figure 2.13). Similar to the majority of caffeine-degrading bacteria, PINKEL catabolizes methylxanthines via the *N*-demethylation pathway as shown in Figure 1.13. This is further supported by the fact that the same pathway intermediates were seen when PINKEL metabolized theobromine, paraxanthine, and 7-methylxanthine (Figure 2.15). PINKEL catabolizes caffeine to xanthine via three successive *N*-demethylation steps, with theobromine as the major metabolite and paraxanthine as the minor metabolite of the first *N*-demethylation step. The 7-methylxanthine → xanthine → uric acid → purine catabolism route is followed afterwards.

### *M. populi* PINKEL is more efficient at catabolizing caffeine than *P. putida* CBB5

Although both *M. populi* PINKEL and *P*. *putida* CBB5 follow the same pathway and use similar enzymes to metabolize methylxanthines, *M. populi* PINKEL can grow and consume caffeine faster when grown under similar conditions (Figure 2.11). To increase the rate of methylxanthine consumption by *P. putida* CBB5, the medium

needed to be supplemented with soytone as an additional nitrogen source [Yu et. Al., 2009]. However, there is no need to provide *M. populi* PINKEL with additional nitrogen or carbon sources to improve its growth or caffeine consumption rate. One possible explanation for the observed difference in growth rate and final cell density could be due to differences in the activity of the *N*-demethylases in the two bacterial strains. Although the activity of the *P. putida* CBB5 *N*-demethylases has already been reported [Summers et al., 2011], detailed studies of the enzymes from *M. populi* PINKEL still need to be completed to make a direct comparison.

An alternative explanation may have to do with the formaldehyde produced during each demethylation reaction. Formaldehyde is a very reactive aldehyde that can cause protein and DNA damage and has been shown to have bacteriostatic activity at sublethal concentrations and bactericidal outcomes at higher concentrations [Neely, 1993; Manchee et al., 1994; Sagripanti and Bonifacino, 1996]. Certain bacterial strains have formaldehyde and formate dehydrogenases that help dissimilate formaldehyde by oxidizing it to carbon dioxide as a mechanism to deal with the formaldehyde produced as part of metabolism. However, some *Pseudomonas* strains, such as *P. putida* KT2440, can only tolerate up to 1.5 mM formaldehyde, and their growth rates decrease by as much as 40% when grown in the presence of 0.5 mM formaldehyde [Roca et al., 2008]. On the other hand, methylotrophs can tolerate higher concentrations of formaldehyde as they have both dissimilation mechanisms and can assimilate formaldehyde via the serine cycle. In the case of *M. populi* PINKEL, the cells can grow in the presence of up to 8 mM formaldehyde without affecting the growth rate (Figure 2.12). The ability of PINKEL to assimilate formaldehyde may contribute to its higher

tolerance to formaldehyde and in turn its faster growth and caffeine consumption rate in comparison to *P. putida* CBB5.

It has been discussed in the literature that using microbes as biocatalysts for the production of valuable methylxanthines, such as paraxanthine and 7-methylxanthine, may be a more efficient alternative to current synthetic methods [Mock et al., 2022]. This is primarily because the chemical synthesis of methylxanthines requires multiple steps, uses highly toxic and volatile reagents, as well as high temperatures and pressures as part of the process [Gulevskaya and Pozharskii, 1991; Muller and Sandoval-Ramirez, 1995; Zavialov et al., 2004]. These extensive and laborious processes result in high prices for methylxanthines that do not naturally exist in high concentrations in the environment as shown in Figure 2.20.

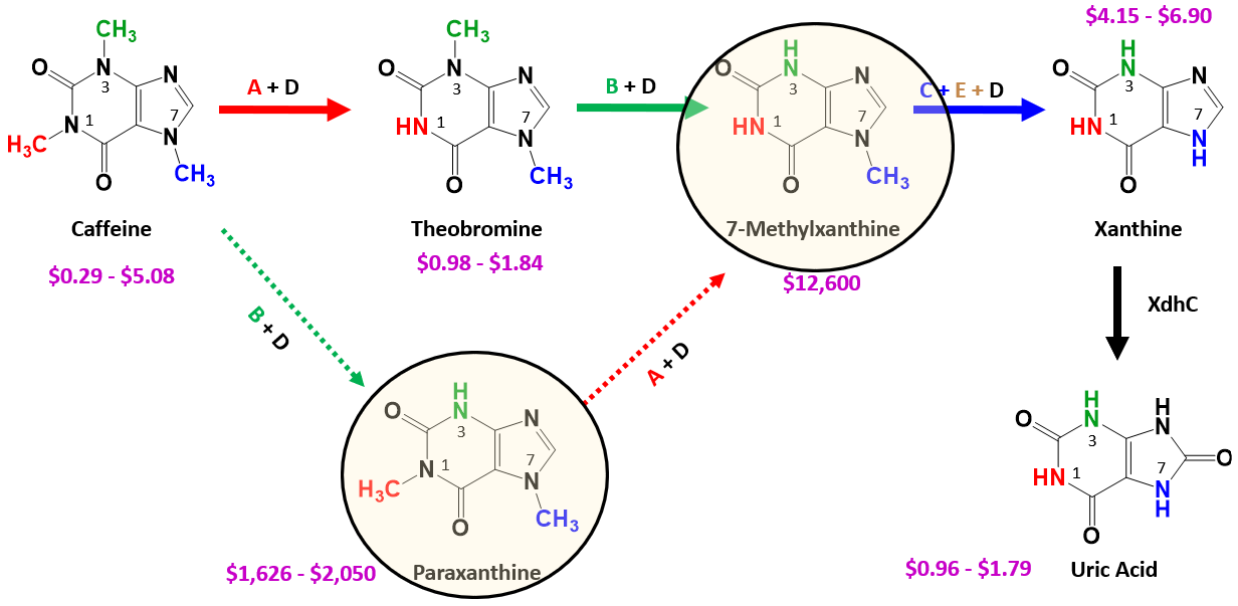



**Figure 2.20**: Price per gram of each methylxanthine shown is based on the Sigma-Aldrich catalog as of June 14, 2023.

Mock and Summers [2023] recently reported a mixed culture of *Escherichia coli* cells engineered to convert caffeine to 7-methylxanthine. One of the *E. coli* strains used in the mixture was specialized at converting caffeine to theobromine, while the second strain was specialized at converting theobromine to 7-methylxanthine. Although the conversion of 1 mM caffeine to 7-methylxanthine by their mixed culture was successful, there was a decrease in resting cell activity with increasing caffeine concentrations. This low activity limits the use of this biotransformation system for the production of methylxanthines at the gram-scale. Additionally, other unidentified metabolites were detected during the biotransformation. PINKEL may represent an alternative caffeine-degrading microbe that can be engineered to develop an optimized process for the production of methylxanthines without the formation of undesired metabolites.

## NdmA catalyzes the conversion of caffeine to paraxanthine

Enzymatic assays using purified NdmA from *Pseudomonas putida* CBB5 showed that NdmA specifically removes methyl groups at the $N_1$ position of methylxanthines and thus is responsible for the conversion of caffeine (1,3,7-trimethylxanthine) to theobromine (3,7-dimethylxanthine) and paraxanthine (1,7-dimethylxanthine) to 7-methylxanthine [Summers et al., 2015]. On the other hand, NdmB from *P. putida* CBB5 was reported to specifically remove $N_3$ methyl groups, thereby converting caffeine (1,3,7-trimethylxanthine) to paraxanthine (1,7-dimethylxanthine) and theobromine (3,7-dimethylxanthine) to 7-methylxanthine. In contrast, however, the results obtained in the present study indicate that NdmA from *M. populi* PINKEL is responsible for the conversion of caffeine to a mixture of theobromine and paraxanthine. This conclusion is supported by the observation that the Δ*ndmA* mutant did not convert caffeine to

paraxanthine, but rather the Δ*ndmB* mutant was capable of catalyzing that conversion (Figure 2.16A and 2.17A). In light of these results, the first step of the minor route of the *N*-demethylation pathway changes in terms of the enzyme that catalyzes the reaction as shown in Figure 2.21.

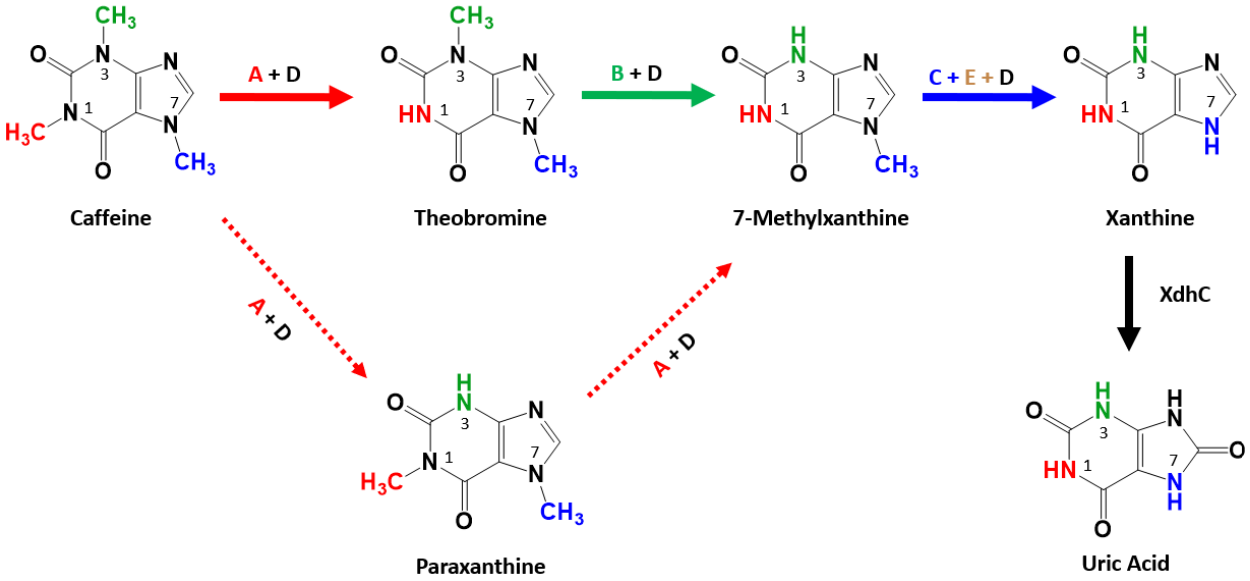



**Figure 2.21**: *N*-demethylation pathway with corrected first step of the minor route.

## References


Anaya AL, Cruz-Ortega R, & Waller GR. (2006). Metabolism and ecology of purine alkaloids. *Frontiers in Bioscience* 11: 2354-2370.

Anthony C. (1982). The biochemistry of methylotrophs. Academic Press, London, United Kingdom.

Bruns RF, & Fergus JH. (1989). Solubilities of adenosine antagonists determined by radioreceptor assay. *Journal of Pharmacy and Pharmacology* 41: 590-594.

Chubiz LM, Purswani J, Carroll SM, & Marx CJ. (2013). A novel pair of inducible expression vectors for use in *Methylobacterium extorquens*. *BMC Research Notes* 6: 183. https://doi.org/10.1186/1756-0500-6-183.

Daly JW. (2007). Caffeine analogs: biomedical impact. *Cellular and Molecular Life Sciences* 64: 2153-2169.

Dash SS & Gummadi SN. (2006). Catabolic pathways and biotechnological applications of microbial caffeine degradation. *Biotechnology Letters* 28: 1993-2002.

Delaney NF, Kaczmarek ME, Ward LM, Swanson PK, Lee MC, & Marx CJ. (2013). Development of an optimized medium, strain and high-throughput culturing methods for *Methylobacterium extorquens*. *PLoS One* 8(4): e62957. doi: 10.1371/journal.pone.0062957. PMID: 23646164; PMCID: PMC3639900.

Ditty JL, & Parales RE. (2015). Protocols for the measurement of bacterial chemotaxis to hydrocarbons. In: McGenity T, Timmis KN, & Nogales B. (eds) Hydrocarbon and Lipid Microbiology Protocols. Springer Protocols Handbooks. Springer, Berlin, Heidelberg. https://doi.org/10.1007/8623_2015_112.

Ferraro DJ, Gakhar L, & Ramaswamy S. (2005). Rieske business: structure-function of Rieske non-heme oxygenases. *Biochemical and Biophysical Research Communications* 338(1): 175-90. doi: 10.1016/j.bbrc.2005.08.222. Epub 2005 Sep 8. PMID: 16168954.

Gokulakrishnan S, Chandraraj K & Gummadi SN. (2005). Microbial and enzymatic methods for the removal of caffeine. *Enzyme and Microbial Technology* 37: 225-232.

Gulevskaya A, & Pozharskii A. (1991). Synthesis of *N*-substituted xanthines. *Chemistry of Heterocyclic Compounds* 27(1): 1–23.

Gummadi SN, Bhavya B, & Ashock N. (2012). Physiology, biochemistry and possible applications of microbial caffeine degradation. *Applied Microbiology Biotechnology* 93: 545–554.

Manchee RJ, Broster MG, Stagg AJ, & Hibbs SE. (1994). Formaldehyde solution effectively inactivates spores of *Bacillus anthracis* on the Scottish island of Gruinard. *Applied Environmental Microbiology* 60: 4167–4171.

Marx CJ, & Lidstrom ME. (2002). Broad-host-range *cre-lox* system for antibiotic marker recycling in Gram-negative bacteria. *BioTechniques* 33(5): 1062-1067.

Mock MB, Mills SB, Cyrus A, Campo H, Dreischarf T, Strock S, & Summers, R.M. (2022). Biocatalytic production and purification of the high-value biochemical paraxanthine. *Biotechnology and Bioprocess Engineering* 27: 640–651. https://doi.org/10.1007/s12257-021-0301-0.

Mock MB & Summers RM. (2023). Mixed culture biocatalytic production of the high-value biochemical 7-methylxanthine. *Journal of Biological Engineering* 17: 2. https://doi.org/10.1186/s13036-022-00316-6.

Muller CE, & Sandoval-Ramirez J. (1995). A new versatile synthesis of xanthines with variable substituents in the 1-, 3-, 7-and 8-positions. *Synthesis* 10:1295–9.

Nathanson JA. (1984). Caffeine and related methylxanthines: Possible naturally occurring pesticides. *Science* 226 (4671): 184-187.

Neely WB. (1963). Action of formaldehyde on microorganisms III: bactericidal action of sublethal concentrations of formaldehyde on *Aerobacter aerogenes*. *Journal of Bacteriology* 86: 445–448.

Parales RE, Sharma G, Zhang X, Subuyuj GA, Langner JT, Wright ME, Ditty JL, & Dawson SC. (2019). Draft genome sequence of the caffeine-degrading methylotroph *Methylorubrum populi* Pinkel. *Microbiology Resource Announcements* 8: e01300-19.

Roca A, Rodriguez-Herva JJ, Duque E & Ramos JL. (2008). Physiological responses of *Pseudomonas putida* to formaldehyde during detoxification. *Microbial Biotechnology* 1 (2): 158-169.

Rogers PJ, & Dernoncourt C. (1998). Regular caffeine consumption: a balance of adverse and beneficial effects for mood and psychomotor performance. *Pharmacology Biochemistry and Behavior* 59: 1039-45.

Roszczenko-Jasińska P, Vu HN, Subuyuj GA, Crisostomo RV, Cai J, Lien NF, Clippard EJ, Ayala E, Ngo RT, Yarza F, Wingett JP, Raghuraman C, Hoeber CA, Martinez-Gomez NC, & Skovran E. (2020). Gene products and processes contributing to lanthanide homeostasis and methanol metabolism in *Methylorubrum extorquens* AM1. *Science Reports* 10: 12663. https://doi.org/10.1038/s41598-020-69401-4.

Sagripanti JL, & Bonifacino A. (1996). Comparative sporicidal effect of liquid chemical agents. *Applied Environmental Microbiology* 62: 545–551.

Sambrook J, Fritsch EF, & Maniatis T. (1989). Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sankar J, Lodha R & Kabra SG. (2008). Doxofylline: The next generation methylxanthine. *Indian Journal of Pediatrics* 68 (3): 251-254.

Simon R, Priefer U & Pühler A. (1983). A broad host range mobilization system for *in vivo* genetic engineering: Transposon mutagenesis in Gram negative bacteria. *Nature Biotechnology* 1: 784–791. https://doi.org/10.1038/nbt1183-784

Skovran E, Palmer AD, Rountree AM, Good NM, & Lidstrom ME. (2011). XoxF is required for expression of methanol dehydrogenase in *Methylobacterium extorquens* AM1. *Journal of Bacteriology* 193: 6032–6038.

Stanier RY, Palleroni NJ, & Doudoroff M. (1966). The aerobic pseudomonads: a taxonomic study. *Journal of General Microbiology* 43: 159 –271. https://doi.org/10.1099/00221287-43-2-159.

Summers RM, Louie TM, Yu CL & Subramanian M. (2011). Characterization of a broad specificity non-haem iron N-demethylase from *Pseudomonas putida* CBB5 capable of utilizing several purine alkaloids as a sole carbon and nitrogen source. *Microbiology* 157: 583-592.

Summers RM, Louie TM, Yu CL, Gakhar L, Louie KC & Subramanian M. (2012). Novel, highly specific *N*-demethylases enable bacteria to live on caffeine and related purine alkaloids. *Journal of Bacteriology* 194: 2041-2049.

Summers RM, Seffernick JL, Quandt EM, Yu CL, Barrick JE & Subramanian M. (2013). Caffeine junkie: an unprecedented glutathione *S*-transferase-dependent oxygenase required for caffeine degradation by *Pseudomonas putida* CBB5. *Journal of Bacteriology* 195: 3933-3939.

Summers R, Gopishetty S, Mohanty S, & Subramanian M. (2014). New genetic insights to consider coffee waste as feedstock for fuel, feed, and chemicals. *Central Journal of Chemistry* 12: 1271-1279.

Summers RM, Mohanty SK, Gopishetty S & Subramanian M. (2015). Genetic characterization of caffeine degradation by bacteria and its potential applications. *Microbial Biotechnology* 8: 369-378.

Vega FE, Emche S, Shao J, Simpkins A, Summers RM, Mock MB, Ebert D, Infante F, Aoki S & Maul JE. (2021). Cultivation and genome sequencing of bacteria isolated from the coffee berry borer (*Hypothenemus hampei*), with emphasis on the role of caffeine degradation. *Frontiers in Microbiology* 12: 644768. doi: 10.3389/fmicb.2021.644768.

Vu HN, Subuyuj GA, Crisostomo RV, & Skovran E. (2021). Transposon mutagenesis for methylotrophic bacteria using *Methylorubrum extorquens* AM1 as a model system. *Methods in Enzymology* 650: 159-184. doi: 10.1016/bs.mie.2021.01.015. Epub 2021 Mar 3. PMID: 33867020.

Yu CL, Louie TM, Summers R, Kale Y, Gopishetty S & Subramanian M. (2009). Two distinct pathways for metabolism of theophylline and caffeine are coexpressed in *Pseudomonas putida* CBB5. *Journal of Bacteriology* 191 (14): 4624-4632.

Zavialov IA, Dahanukar VH, Nguyen H, Orr C & Andrews DR. (2004). New and practical method for synthesis of 1-and 1,3-substituted xanthines. *Organic Letters* 6 (13): 2237-2240.

# Chapter 3

## Transcriptional regulation of the *N*-demethylation pathway in *Methylorubrum populi* PINKEL

## Introduction

The worldwide consumption of caffeine and related methylxanthines via various food products and pharmaceuticals has made such compounds ubiquitous in the environment [Li et al., 2020]. Hence, it comes to no surprise that caffeine-degrading microorganisms, such as yeast, fungi, and bacteria, have been isolated from the environment. Currently, there is a great interest in engineering caffeine-degrading microbes for the bio-decaffeination of food products, production of medically relevant methylxanthines, and bioremediation of contaminated soil and water systems [Summers et al., 2015]. Therefore, a complete understanding, not only of the bacterial enzymes necessary for caffeine biodegradation, but how those genes are regulated at the transcriptional level and how bacteria sense methylxanthines in the environment may contribute to these potential biotechnological applications.

Up until the 1970s, caffeine metabolism in microbes had not received much attention since caffeine was believed to be toxic to microbes [Gopishetty et al., 2011]. Since then, approximately 71 bacterial strains capable of degrading caffeine have been reported, most of which belong to the genus *Pseudomonas*, primarily *Pseudomonas putida* [Vega et al., 2021]. Recently, the only known methylotrophic bacterium, *Methylorubrum populi* strain PINKEL, capable of using caffeine as its sole carbon, nitrogen, and energy source was isolated via a caffeine enrichment strategy [Parales et al., 2019]. The ability of PINKEL to use the formaldehyde produced during caffeine catabolism for growth provides a new caffeine-degrading microbe with a unique metabolism that may serve as a tool to investigate the transcriptional regulation of genes encoding enzymes that participate in the degradation of methylxanthines.

Early studies using these caffeine-degrading bacteria revealed two caffeine catabolic pathways: the *N*-demethylation pathway and the C-8 oxidation pathway [Summers et al., 2015]. The *N*-demethylation pathway appears to be the most common pathway since it has been observed in more than 80% of the isolates in which caffeine degradation has been studied (Table 1.5). In *M. populi* PINKEL, the *N*-demethylation pathway involves three demethylation steps to convert caffeine (1,3,7-trimethylxanthine) to xanthine and requires five proteins: NdmA, NdmB, NdmC, NdmD and NdmE (Figure 3.1A). The first step is catalyzed by NdmA, which removes the methyl group at either the $N_1$ position to convert caffeine (1,3,7-trimethylxanthine) to theobromine (3,7-dimethylxanthine), or at the $N_3$ position to convert caffeine (1,3,7-trimethylxanthine) to paraxanthine (1,7-dimethylxanthine). However, theobromine appears to be the major metabolite of the first step of the pathway possibly because the $N_1$ methyl group seems to fit better at the catalytic site in comparison to the $N_3$ methyl group when caffeine is the initial substrate [Kim et al., 2019]. Both of these dimethylxanthines are converted to 7-methylxanthine in the second step of the pathway. The $N_1$ methyl group from paraxanthine is removed by NdmA, whereas the $N_3$ methyl group from theobromine is removed by NdmB. Both NdmA and NdmB require the aid of NdmD to acquire and transfer electrons from NADH to power the demethylation reactions, which generates formaldehyde as a byproduct. The last methyl group is removed by the NdmCDE protein complex to convert 7-methylxanthine to xanthine, which then gets oxidized at the C-8 position to enter the purine catabolism pathway.

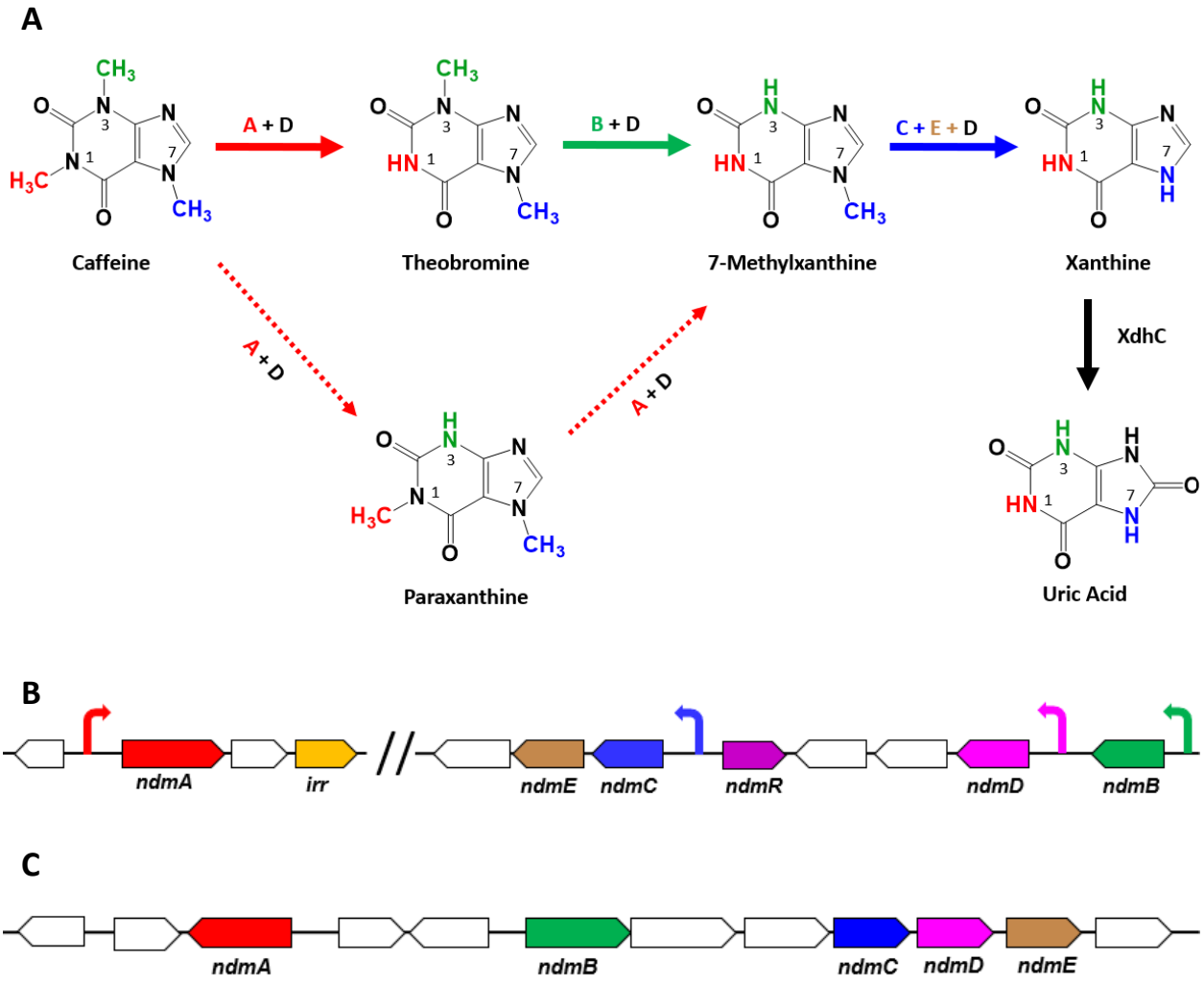



**Figure 3.1**: (A) *N*-demethylation pathway in *Methylorubrum populi* PINKEL. Solid arrows represent the major pathway, and dashed arrows represent the minor pathway. Letters above each arrow indicate the *N*-demethylase proteins that catalyze each step. (B) Genetic organization of *ndm* genes in *M. populi* PINKEL. The colored arrows upstream of the *ndmA, ndmB, ndmC, and ndmD* genes represent regions predicted to contain a promoter region. The labeled *ndmR* gene is predicted to encode a LysR-type transcription regulator and the *irr* gene is predicted to encode an iron responsive regulator. (C) Genetic organization of *ndm* genes in *Pseudomonas putida* CBB5.

The *N*-demethylase proteins have been characterized as Rieske [2Fe-2S] monooxygenases with a non-heme ferrous iron at the catalytic site [Summers et al., 2015]. This system consists of two components, a reductase, and a monooxygenase,

that work together to transfer the electrons acquired from NADH to the catalytic site of the monooxygenase to power the removal of methyl groups from the corresponding methylxanthine in the form of formaldehyde (Figure 3.2). In *P. putida* CBB5, NdmA, NdmB, and NdmC have been identified as monooxygenase components, whereas NdmD has been identified as iron-sulfur flavoprotein reductase [Summers et al., 2012; Summers et al., 2013]. NdmE has been shown to be required for structural support when the NdmCDE complex is formed for the removal of the $N_7$ methyl group of methylxanthines [Summers et al., 2014].

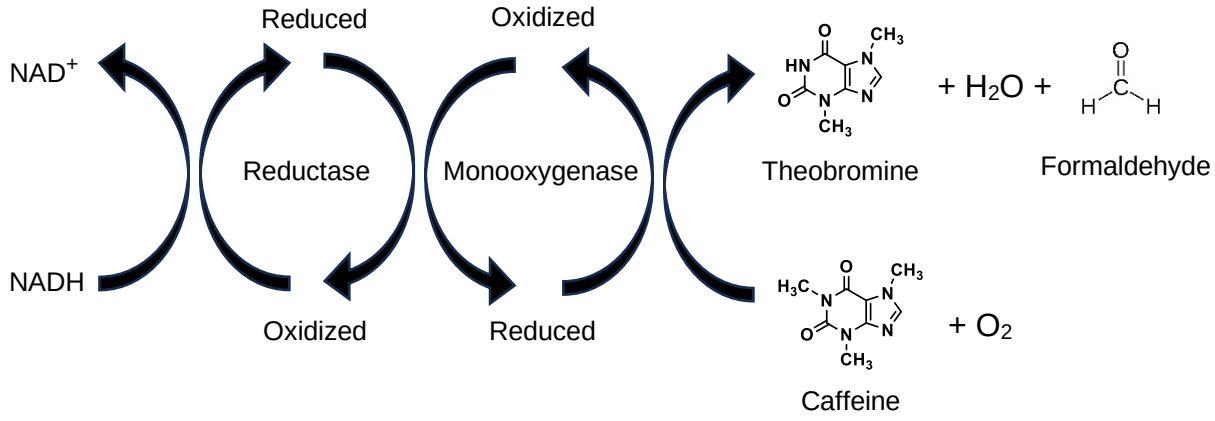



**Figure 3.2**: Electron flow in the demethylation of methylxanthines by the two component Rieske [2Fe-2S] monooxygenase. The $N_1$ demethylation of caffeine to produce theobromine is shown as an example. Formaldehyde is a byproduct of each demethylation step.

Even though the caffeine catabolic enzymes have been well characterized, there is not much information about how those genes are regulated at the transcriptional level [Summers et al., 2015]. Initial enzyme assay experiments using *Pseudomonas* sp. cell lysates obtained from induced versus uninduced cells hinted towards the inducible

nature of the *N*-demethylation pathway [Dash and Gummadi, 2008]. Growth experiments using *M. populi* PINKELnc also supports this finding since there is a shorter growth lag observed for cells pre-grown in caffeine in comparison to cells pre-grown in the absence of caffeine (Figure 3.3). In contrast to *P. putida* CBB5, which has a single gene cluster encoding all of the genes necessary to convert caffeine to xanthine, the *ndm* genes in PINKEL are in two separate clusters (Figure 3.1B and C). Additionally, NdmR, a putative LysR-type transcriptional regulator encoded by peg_1778, is located within the locus where the *ndm* genes are found and is a good candidate for the transcriptional regulation of the *N*-demethylation pathway.

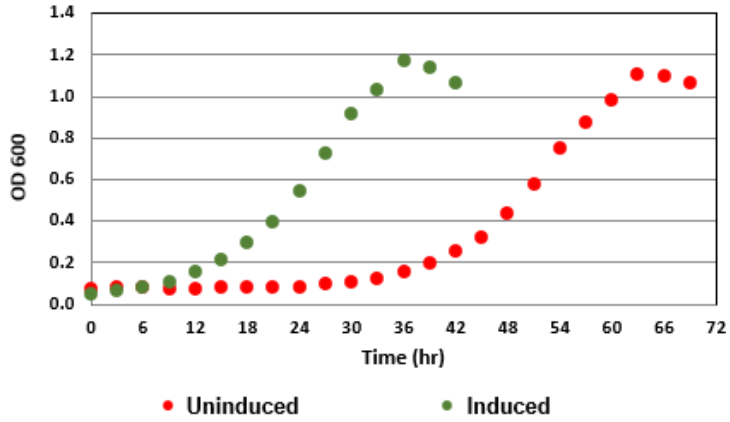



**Figure 3.3**: Growth of *Methylorubrum populi* PINKELnc on 10 mM caffeine. Cells were either pre-grown in the presence (induced) or absence (uninduced) of caffeine.

The LysR-type transcriptional regulators (LTTRs) comprise what is considered the most abundant and diverse family of transcriptional regulators among bacteria (Table 3.1) [Henikoff et al., 1988; Maddocks and Oyston, 2008; Koentjoro and Ogawa, 2018], with functional orthologues found in some archaea [Perez-Rueda and Collado-Vides, 2001; Sun and Klein, 2004] and eukaryotic organisms [Minoda et al., 2010].

**Table 3.1**: Examples of LysR-type transcriptional regulators

| LTTR | Regulation | Target gene function | Co-factor | Origin | Subdivision* | Reference |
|---|---|---|---|---|---|---|
| AlsR | Activator (local) | Acetoin synthesis | Acetate | *Bacillus subtilis* | Gm + | Renna *et al.* (1993) |
| AmpR | Activator (local) | β-Lactamase | Unknown | *Rhodobacter capsulatus* | α | Bartowsky & Normark (1993) |
| | | | | *Enterobacter cloacae* | γ | |
| | | | | *Citrobacter freundii* | γ | |
| ArgP | Activator (local) | Arginine transport | Arginine | *Escherichia coli* | γ | Nandineni & Gowrishankar (2004) |
| BenM | Activator (global) | Aromatic compound degradation | *cis,cis*-Muconate or benzoate | *Acinetobacter* spp. | γ | Collier *et al.* (1998) |
| BlaA | Activator (local) | β-Lactamase | Unknown | *Streptomyces* spp. | Gm + | Raskin *et al.* (2003) |
| CatM | Activator (global) | Catechol catabolism | *cis,cis*-Muconate | *Acinetobacter calcoaceticus* | γ | Ezezika *et al.* (2006) |
| CatR | Repressor (local) | Catechol catabolism | *cis,cis*-Muconate | *Pseudomonas putida* | γ | Chugani *et al.* (1998) |
| CbbR | Activator (global) | Carbon dioxide fixing | | *Xanthobacter flavus* | α | van Keulen *et al.* (2003) |
| CfxR | Activator (global) | Carbon dioxide fixing | Auxotrophic growth conditions | *Alcaligenes eutrophus* | β | Windhövel & Bowien (1991) |
| ChiR | Activator (local) | Chitin binding/chitinase | Unknown | *Serratia marcescens* | γ | Suzuki *et al.* (2001) |
| CidR | Activator (local) | *cid* operon (murein hydrolase regulation) | Acetic acid | *Staphylococcus* spp. | Gm + | Yang *et al.* (2005) |
| | | | | *Bacillus anthracis* | Gm + | Ahn *et al.* (2006) |
| ClcR | Activator (local) | Chlorocatechol catabolism | *cis,cis*-Muconate | *Pseudomonas putida* | γ | Coco *et al.* (1993) |
| CrgA | Activator/Repressor (global) | Pili/capsule synthesis | α-Methylene-γ-butyrolactone | *Neisseria meningitidis* | γ | Deghmane *et al.* (2000) |
| CynR | Activator (local) | Cyanate detoxification | Cyanate | *Escherichia coli* | γ | Sung & Fuchs (1992) |
| CysB | Activator (global) | Cysteine biosynthesis | *N*-Acetylserine | *Salmonella enterica* serovar Typhimurium | γ | van der Ploeg *et al.* (1997) |
| | | | | *Escherichia coli* | γ | |
| CysL | Activator (global) | Sulphite reductase | Sulphur source dependent | *Bacillus subtilis* | Gm + | Guillouard *et al.* (2002) |
| GltC | Activator (local) | Glutamate synthase | | *Bacillus subtilis* | Gm + | Picossi *et al.* (2007) |
| HupR | Activator (global) | Haem uptake | Unknown | *Vibrio vulnificus* | γ | Litwin & Quackenbush (2001) |
| HvrB | Activator (global) | *S*-Adenosyl-L-homocysteine hydrolase expression | Light sensitivity | *Rhodobacter capsulatus* | α | Buggy *et al.* (1994) |
| IlvR | Activator (local) | Isoleucine/valine biosynthesis | Unknown | *Caulobacter crescentus* | α | Malakooti & Ely (1994) |
| IlvY | Activator (local) | Isoleucine/valine biosynthesis | α-Acetolactate or α-acetohydroxybutyrate | *Escherichia coli* | γ | Wek & Hatfield (1988) |
| IrgB | Activator (local) | Iron-regulated virulence factor | Unknown | *Vibrio cholerae* | γ | Goldberg *et al.* (1991) |
| LeuO | Activator/Repressor (global) | Bacterial stringent response | Unknown | *Salmonella enterica* serovar Typhimurium | γ | Hernández-Lucas *et al.* (2008) |
| LrhA | Activator (global) | Flagella, motility and chemotaxis | Unknown | *Escherichia coli* | γ | Lehnen *et al.* (2002) |
| LysR | Activator (local) | Lysine biosynthesis | Diaminopimelate | *Escherichia coli* | γ | Stragier *et al.* (1983) |
| MdcR | Activator (local) | Malonate catabolism | Malonate | *Klebsiella pneumoniae* | γ | Peng *et al.* (1999) |
| MetR | Activator (global) | Methionine and cysteine transport/biosynthesis | Homocysteine | *Streptococcus* spp. | Gm + | Kovaleva & Gelfand (2007) |
| MleR | Activator (local) | Malolactic enzyme | Unknown | *Lactococcus lactis* | Gm + | Renault *et al.* (1989) |
| MtaR | Activator (global) | Methionine transport | Unknown | Group B streptococci | Gm + | Shelver *et al.* (2003) |
| MvfR | Activator (global) | Pathogenicity regulator | 4-Hydroxy-2-heptylquinone | *Pseudomonas aeruginosa* | γ | Cao *et al.* (2001) |
| NagR | Activator (local) | Naphthalene catabolism | Salicylate | *Ralstonia eutropha* | α | Jones *et al.* (2003) |
| NahR | Activator (local) | Naphthalene/salicylate catabolism | Salicylate | NAH7 plasmid of *Pseudomonas putida* | NA | Park *et al.* (2002) |
| NhaR | Activator (local) | $Na^+/H^+$ antiporter | $Na^+$ | *Escherichia coli* | γ | Dover & Padan (2001) |
| NocR | Activator (local) | Nopaline catabolism | Octopine | Ti plasmids of *Agrobacterium tumefaciens* | NA | von Lintig *et al.* (1994) |
| NodD | Activator (global) | Nitrogen fixation/symbiosis | Flavonoids | *Rhizobium* spp. | β | Schlaman *et al.* (1992b) |
| | | | | *Bradyrhizobium* spp. | β | |
| | | | | *Azorhizobium* spp. | β | |
| OccR | Activator (local) | Octopine catabolism | Octopine | Ti plasmids of *Agrobacterium tumefaciens* | NA | Habeeb *et al.* (1991) |
| OxyR | Activator (global) | Oxidative stress response ($H_2O_2$) | Redox changes | *Escherichia coli* | γ | Farr & Kogoma (1991) |
| | | | | *Salmonella enterica* serovar Typhimurium | γ | |
| PhcA | Activator (global) | Virulence regulator | Unknown | *Pseudomonas solanacearum* | γ | Brumbley *et al.* (1993) |
| QseA | Activator (global) | Quorum sensing | Unknown | Enterpathogenic/enterohaemorrhagic *Escherichia coli* | γ | Sperandio *et al.* (2002) |
| RbcR | Activator (local) | Carbon dioxide fixing | Unknown | *Chromatium vinosum*, *Thiobacillus ferrooxidans* | γβ | Viale *et al.* (1991) |
| RovM | Repressor (global) | Invasion/motility, virulence | Unknown | *Yersinia pestis* | γ | Heroven *et al.* (2007) |
| SpvR | Activator (local) | Virulence factor synthesis | Unknown | *Salmonella* spp. virulence plasmids | NA | Sheehan & Dorman (1998) |
| SyrM | Activator (global) | Exopolysaccharide synthesis | Unknown | *Rhizobium meliloti* | β | Barnett *et al.* (1998) |
| TcbR | Activator (local) | Chlorocatechol metabolism | 2-Chloromuconic acid | *Pseudomonas* spp. plasmid J51 | NA | van der Meer *et al.* (1991) |
| ToxR | Activator (local) | Quorum sensing | Toxoflavin | *Burkholderia glumae* | β | Kim *et al.* (2004) |
| YofA | Activator (local) | Cell division | Unknown | *Bacillus subtilis* | Gm+ | Lu *et al.* (2007) |
| YtxR | Activator (local/global) | ADP-ribosyl-transferase toxin | Unknown | *Yersinia entercolitica* | γ | Axler-Diperte *et al.* (2006) |

*Subdivisions of Proteobacteria indicated by α, β, γ; Gm +, Gram positive; NA, not applicable

Source: Maddocks and Oyston, 2008

Hence, it comes to no surprise that LTTRs have evolved to regulate the expression of genes whose products participate in diverse functions, including degradation of aromatic compounds, biosynthesis of amino acids, cell division, quorum sensing, virulence factors, motility, antibiotic resistance, nitrogen fixation, and oxidative stress responses [Kovacikova & Skorupski, 1999; Cao et al., 2001; Deghmane et al., 2002; Kim et al., 2004; Byrne et al., 2007; Sperandio et al., 2007]. Members of this family are typically 300 to 330 residues in length, autoregulate themselves, can act as transcriptional activators or repressors, and bind to their target DNA-binding sites as homotetramers [Maddocks and Oyston, 2008; Koentjoro and Ogawa, 2018; Baugh et al., 2023].

Despite the great diversity of functions that LTTRs regulate, there are two conserved domains found in all members of this family: a DNA-binding domain and a ligand (or effector) binding domain connected to each other by a linker helix (Figure 3. 4). The highly conserved DNA-binding domain is located at the N-terminus and contains a winged helix-turn-helix (wHTH) motif that recognizes and binds to the recognition binding site (RBS) and activation binding site (ABS) near the promoter regions of the regulons they control [MacLean et al., 2011]. Sequence comparison of various promoter regions for LTTRs has revealed T-$N_{11}$-A as the consensus sequence of the RBS (also referred to as the LTTR box) [Koentjoro and Ogawa, 2018]. The less conserved ligand binding domain is located at the C-terminus and consists of two subdomains (RD1 and RD2) that create a cleft for the binding of the corresponding inducer or effector [Baugh et al., 2023].

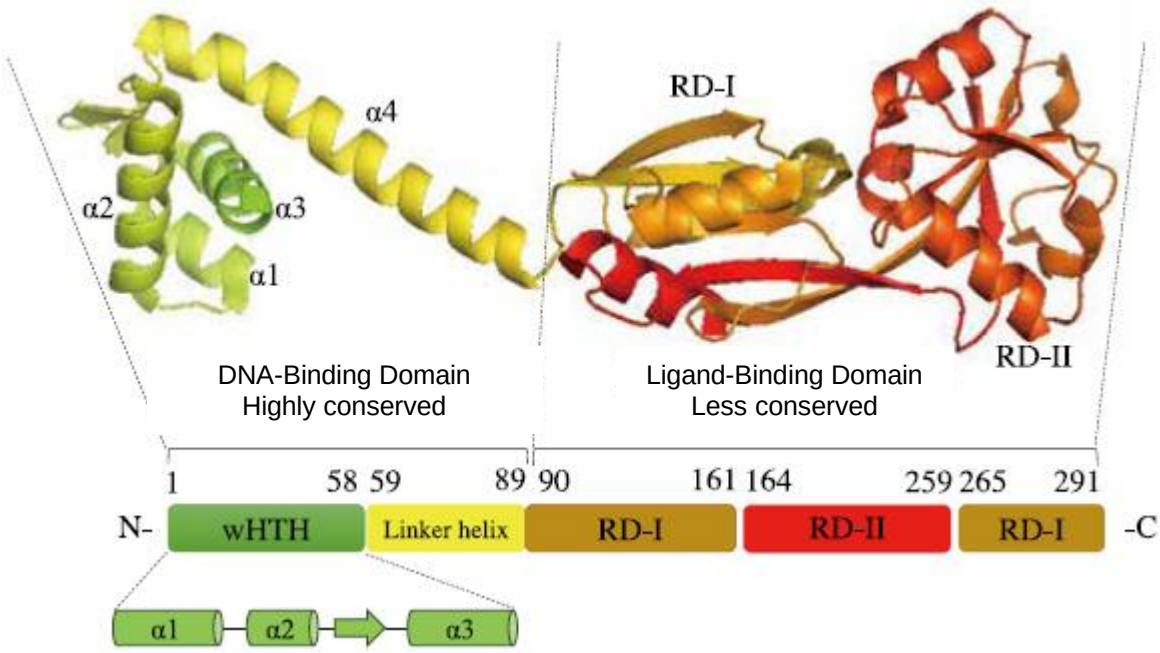



**Figure 3.4**: General structure of LysR-type transcriptional regulators. The N-terminus contains a winged helix-turn-helix (wHTH) DNA-binding domain and a C-terminus contains a Ligand/Effector-binding domain. Both of these domains are connected by a linker helix.

Source: Koentjoro and Ogawa, 2018.

Even though there is no unifying mechanism by which LTTRs control expression of their regulons, they are typically bound to the RBS and ABS binding sites regardless of the presence or absence of their corresponding inducer/effector as shown in Figure 3.5 [Baugh et al., 2023]. In the absence of the inducer, the DNA is bent in such a way that the RNA-polymerase cannot have a stable contact with the promoter, thus preventing transcription. On the other hand, binding of the inducer to the corresponding LTTR causes a conformational change that alters the angle in which the DNA is bent. This conformational change results in an interaction between the LTTR and RNA-polymerase that stabilizes the polymerase on the promoter to turn on expression of the regulated genes.

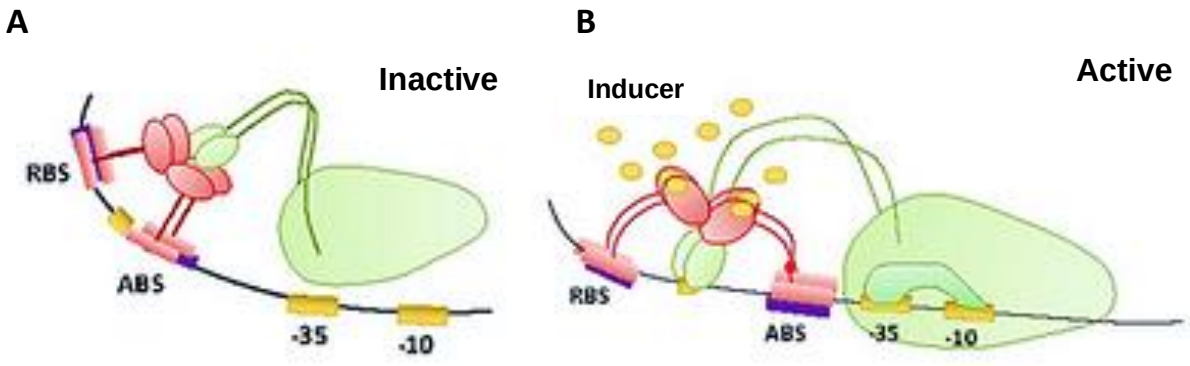



**Figure 3.5**: Example of transcriptional regulatory mechanism by which LysR-type transcriptional regulators control expression of their corresponding regulons. (A) In the absence of an inducer the RNA-polymerase cannot effectively bind to its promoter region. (B) When an inducer binds, the angle of DNA bending changes and stabilizes the LysR-RNA polymerase interactions to turn on expression.

Source: Fernandez-Lopez et al., 2015.

The aim of this study is to identify gene(s) encoding protein(s) involved in the transcriptional regulatory mechanism of the *N*-demethylation pathway in *M. populi* PINKEL, to identify the metabolite(s) acting as the inducer(s), and to narrow down the fragment of DNA within the identified promoter regions relevant for the transcriptional regulation. Transcriptional promoter fusion assays using wild-type and pathway mutant strains in combination with bioinformatic tools will be used to investigate the transcriptional regulation in *M. populi* PINKEL.

## Materials and Methods

### Chemical reagents

All the xanthine alkaloids used in this study, including caffeine (1,3,7-trimethylxanthine), theobromine (3,7-dimethylxanthine), paraxanthine (1,7-dimethylxanthine), 7-methylxanthine, xanthine, and uric acid were purchased from Sigma-Aldrich (St. Louis, MO). Unless otherwise stated, all reagents used to make growth media, buffers, agarose gels, and HPLC-grade chemicals were bought from Fisher Scientific (Pittsburg, PA). Restriction endonucleases, Antarctic Phosphatase, T4 DNA ligase, NEB builder HIFI DNA assembly cloning kit, and the Quick-load purple DNA ladder (50 bp and 1 kb) were obtained from New England Biolabs (NEB, Ipswitch, MA). The GoTaq Green (2x) Master Mix was purchased from Promega (Madison, WI). The custom-made DNA oligonucleotides used in this study were bought and analyzed using the Integrated DNA Technologies (IDT) platform (Coralville, IA).

### Methylxanthine stock solutions

It has been well documented that some of the methylxanthines, such as theobromine and xanthine, have poor solubility in aqueous solutions [Bruns and Fergus, 1989]. To overcome this problem and increase the solubility of methylxanthines in aqueous solutions, water was heated to 100°C before adding the corresponding methylxanthine. This method allowed us to make the final concentrations needed for theobromine, paraxanthine, and 7-methylxanthine. To be consistent, the same method was used to dissolve caffeine even though dissolving caffeine in aqueous solutions is not as difficult. Since xanthine does not dissolve, even after boiling the water, growth of cultures on xanthine was tested by providing solid crystals of xanthine as the sole

carbon source on solid agar media. Because theobromine precipitates out of solution within 3 days, all methylxanthine stock solutions were freshly made the day they were used. To ensure methylxanthines did not degrade as a result of the water boiling procedure, samples from the stock solutions were analyzed via High Performance Liquid Chromatography (HPLC). No additional metabolites were detected as a result of heating.

## Bacterial strains, primers, and plasmids

The bacterial strains, plasmids, and primers used in this study are listed in Tables 3.2, 3.3, and 3.4, respectively. *Escherichia coli* DH5α was used for cloning and plasmid propagation. *Escherichia coli* S17-1 was used to mate the engineered plasmids into *Methylorubrum populi* PINKEL. Prior to performing any experiments, *M. populi* PINKEL was swarmed several times in minimal salts base (MSB, see below) soft agar medium containing 0.25 mM caffeine as a carbon source, 0.1X Hutner's mineral base (trace elements solution) and 0.3% noble agar (Figure 3.6) [Hardwood et al., 1994; Ditty and Parales, 2015]. This strategy allowed the enrichment of a highly motile population of PINKEL cells (PINKELsw for swarmed PINKEL). One difficulty with working with certain pink-pigmented methylotrophic bacteria is their tendency to form clumps under certain growth conditions. This makes measuring the growth of the culture difficult. Additionally, performing chemotaxis assays would not be possible since studying the swimming behavior of bacteria requires free swimming cells. Other researchers have found that deleting three genes suggested to be involved in cellulose synthesis reduced or eliminated the clumping phenotype in *Methylobacterium extorquens* AM1 and PA1

[Delaney et al., 2013]. To determine if the same non-clumping phenotype was replicable in PINKEL, we deleted the gene cluster containing the three homologous genes that encode the cellulose synthase genes: *celA* (encoded by peg_1742), *celB* (encoded by peg_1743) and *celC* (encoded by peg_1744). Deletion of this gene cluster was done in cells obtained from the highly motile PINKEL population (PINKELsw). A reduction in the amount of clumping was observed and a freezer stock of the resulting strain was saved and used as the wild-type strain (PINKELnc for non-clumping PINKEL) for all the experiments performed in this study. All mutants generated in this study were derivatives of this strain as well.

**Table 3.2:** Bacterial strains used in this study

| Strain | Description* | Reference or source |
|---|---|---|
| *E. coli* strains | | |
| DH5α | Cloning host | Life Technologies, Gaithersburg, MD |
| NEB 5α | Cloning host | New England Biolabs, Ipswitch, MA |
| S17-1 | Helper strain for conjugation | Simon et al., 1983 |
| *M. populi* strains | | |
| PINKEL | Wild-type caffeine degrading strain | This Study |
| PINKELsw | PINKEL strain swarmed for a highly motile cell population | This Study |
| PINKELnc | PINKELsw strain with Δ*celA* gene cluster deletion (peg_1742 - peg_1745) | This Study |
| GS101 | PINKELnc Δ*ndmR* (peg_1778) | This Study |
| GS102 | PINKELnc Δ*irr* (peg_1792) | This Study |
| GS165 | PINKELnc Δ*ndmD* (peg_1781) | This Study |
| GS150 | PINKELnc with pAP5 plasmid ($Tc^R$) | This Study |
| GS151 | PINKELnc with pGS15 plasmid ($Tc^R$) | This Study |
| GS152 | PINKELnc with pGS16 plasmid ($Tc^R$) | This Study |
| GS153 | PINKELnc with pGS17 plasmid ($Tc^R$) | This Study |
| GS155 | PINKELnc with pGS18 plasmid ($Tc^R$) | This Study |
| GS170 | PINKELnc with pGS19 plasmid ($Tc^R$) | This Study |
| GS171 | PINKELnc with pGS20 plasmid ($Tc^R$) | This Study |

**Table 3.2:** Continued

| | | |
|---|---|---|
| GS172 | PINKELnc with pGS21 plasmid ($Tc^R$) | This Study |
| GS173 | PINKELnc with pGS22 plasmid ($Tc^R$) | This Study |
| GS174 | PINKELnc with pGS23 plasmid ($Tc^R$) | This Study |
| GS175 | PINKELnc with pGS24 plasmid ($Tc^R$) | This Study |
| GS176 | PINKELnc with pGS25 plasmid ($Tc^R$) | This Study |
| GS177 | PINKELnc with pGS26 plasmid ($Tc^R$) | This Study |
| GS178 | PINKELnc with pGS27 plasmid ($Tc^R$) | This Study |
| GS180 | PINKELnc with pGS28 plasmid ($Tc^R$) | This Study |
| GS181 | PINKELnc with pGS29 plasmid ($Tc^R$) | This Study |
| GS182 | PINKELnc with pGS30 plasmid ($Tc^R$) | This Study |
| GS183 | PINKELnc with pGS31 plasmid ($Tc^R$) | This Study |
| GS184 | PINKELnc with pGS32 plasmid ($Tc^R$) | This Study |
| GS185 | PINKELnc with pGS33 plasmid ($Tc^R$) | This Study |
| GS111 | PINKELnc Δ*ndmR* with pAP5 plasmid ($Tc^R$) | This Study |
| GS112 | PINKELnc Δ*ndmR* with pGS15 plasmid ($Tc^R$) | This Study |
| GS113 | PINKELnc Δ*ndmR* with pGS16 plasmid ($Tc^R$) | This Study |
| GS143 | PINKELnc Δ*ndmR* with pGS17 plasmid ($Tc^R$) | This Study |
| GS107 | PINKELnc Δ*irr* with pAP5 plasmid ($Tc^R$) | This Study |
| GS108 | PINKELnc Δ*irr* with pGS15 plasmid ($Tc^R$) | This Study |
| GS109 | PINKELnc Δ*irr* with pGS16 plasmid ($Tc^R$) | This Study |
| GS145 | PINKELnc Δ*irr* with pGS17 plasmid ($Tc^R$) | This Study |
| GS248 | PINKELnc Δ*ndmR* complemented with pLC290 ($Kan^R$) | This Study |
| GS249 | PINKELnc Δ*ndmR* complemented with pGS14 ($Kan^R$) | This Study |
| GS250 | PINKELnc Δ*ndmR* complemented with pGS14 containing pAP5 ($Kan^R$, $Tc^R$) | This Study |
| GS251 | PINKELnc Δ*ndmR* complemented with pGS14 containing pGS15 ($Kan^R$, $Tc^R$) | This Study |
| GS252 | PINKELnc Δ*ndmR* complemented with pGS14 containing pGS16 ($Kan^R$, $Tc^R$) | This Study |
| GS253 | PINKELnc Δ*ndmR* complemented with pGS14 containing pGS17 ($Kan^R$, $Tc^R$) | This Study |
| GS254 | PINKELnc Δ*ndmD* with pGS15 plasmid ($Tc^R$) | This Study |
| GS255 | PINKELnc Δ*ndmD* with pGS16 plasmid ($Tc^R$) | This Study |
| GS256 | PINKELnc Δ*ndmD* with pGS17 plasmid ($Tc^R$) | This Study |

All *M. populi* PINKEL strains used in this study are isogenic derivatives of the swarmed *celA* deletion mutant (PINKELnc).

* $Kan^R$, Kanamycin resistant; $Tc^R$, tetracycline resistant

**Table 3.3:** Plasmids strains used in this study

| Plasmid | Description* | Reference or source |
|---|---|---|
| pHV2 | Modified pCM184 suicide vector expressing *sacB* ($Kan^R$, $Tc^R$, $Amp^R$) | Roszczenko-Jasińska et al., 2020 |
| pLC290 | Expression vector using a cumate inducible promoter ($Kan^R$) | Chubiz et al., 2013 |
| pCM157 | Modified pCM62 vector expressing *cre* ($Tc^R$). Used to delete kanamycin insertion cassette. | Marx and Lidstrom, 2002 |
| pAP5 | Promoterless *yfp* (venus) fusion vector created using pCM62 ($Tc^R$) as a backbone | Skovran et al., 2011 |
| pGS6 | *celABCD* gene cluster (peg_1742 - peg_1745)<br>767 bp upstream PCR fragment cloned into MfeI - NdeI sites and 964 bp downstream PCR fragment cloned into AgeI - SacI sites of pHV2 ($Kan^R$, $Tc^R$, $Amp^R$) | This Study |
| pGS10 | *ndmD* gene (peg_1781)<br>523 bp upstream PCR fragment cloned into MfeI - NcoI sites and 1008 bp downstream PCR fragment cloned into AgeI - SacI sites of pHV2 ($Kan^R$, $Tc^R$, $Amp^R$) | This Study |
| pGS12 | *ndmR* gene (peg_1778)<br>1026 bp upstream PCR fragment cloned into MfeI - NcoI sites and 460 bp downstream PCR fragment cloned into AgeI - SacI sites of pHV2 ($Kan^R$, $Tc^R$, $Amp^R$) | This Study |
| pGS13 | *irr* gene(peg_1792)<br>951 bp upstream PCR fragment cloned into MfeI - NdeI sites and 538 bp downstream PCR fragment cloned into AgeI - SacI sites of pHV2 ($Kan^R$, $Tc^R$, $Amp^R$) | This Study |
| pGS14 | *ndmR* gene (peg_1778) cloned into pLC290 ($Kan^R$) | This Study |
| pGS15 | WT ⱷ*ndmA* - 678 bp promoter region (peg_1790)<br>678 bp PCR fragment upstream of the *ndmA* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS16 | WT ⱷ*ndmB* - 257 bp promoter region (peg_1782)<br>257 bp PCR fragment upstream of the *ndmB* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS17 | WT ⱷ*ndmC* - 1844 bp promoter region (peg_1777)<br>1844 bp PCR fragment upstream of the *ndmC* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS18 | ⱷ*ndmD* 1332 bp promoter region [WT ⱷndm*D*] (peg_1781)<br>1332 bp PCR fragment upstream of the *ndmD* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS19 | ⱷ*ndmA* - 595 bp promoter region (peg_1790)<br>595 bp PCR fragment upstream of the *ndmA* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS20 | ⱷ*ndm*A - 545 bp promoter region (peg_1790)<br>545 bp PCR fragment upstream of the *ndm*A gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |

**Table 3.3:** Continued

| | | |
|---|---|---|
| pGS21 | ↱*ndm*A - 493 bp promoter region (peg_1790)<br>493 bp PCR fragment upstream of the *ndmA* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS22 | ↱*ndm*A - 460 bp promoter region (peg_1790)<br>460 bp PCR fragment upstream of the *ndmA* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS23 | ↱*ndm*A - 400 bp promoter region (peg_1790)<br>400 bp PCR fragment upstream of the *ndmA* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS24 | ↱*ndm*A - 354 bp promoter region (peg_1790)<br>354 bp PCR fragment upstream of the *ndmA* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS25 | ↱*ndm*A - 314 bp promoter region (peg_1790)<br>314 bp PCR fragment upstream of the *ndmA* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS26 | ↱*ndm*A - 244 bp promoter region (peg_1790)<br>244 bp PCR fragment upstream of the *ndmA* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS27 | ↱*ndm*A - 203 bp promoter region (peg_1790)<br>203 bp PCR fragment upstream of the *ndmA* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS28 | ↱*ndm*A - 111 bp promoter region (peg_1790)<br>111 bp PCR fragment upstream of the *ndmA* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS29 | ↱*ndm*A - 54 bp promoter region (peg_1790)<br>54 bp PCR fragment upstream of the *ndmA* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS30 | ↱*ndm*B - 206 bp promoter region (peg_1782)<br>206 bp PCR fragment upstream of the *ndmB* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS31 | ↱*ndm*B - 147 bp promoter region (peg_1782)<br>147 bp PCR fragment upstream of the *ndmB* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS32 | ↱*ndm*B - 117 bp promoter region (peg_1782)<br>117 bp PCR fragment upstream of the *ndmB* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |
| pGS33 | ↱*ndm*B - 49 bp promoter region (peg_1782)<br>49 bp PCR fragment upstream of the *ndmB* gene cloned into AclI - NcoI sites of pAP5 ($Tc^R$) | This Study |

* $Kan^R$, Kanamycin resistant; $Tc^R$, tetracycline resistant; $Amp^R$, ampicillin resistant

**Table 3.4:** Primers used in this study

| Primer name | *Sequence (5' - 3') |
|---|---|
| 1742_MfeI_UF | atatcaattgtcctactcgaccaagttcaacgc |
| 1742_NdeI_UR | tatacatatgaggatgtagcggagcaccacg |
| 1745_AgeI_DF | atataccggtctacgcctcgcagaagcagacc |
| 1745_SacI_DR | atatgagctcttcgagcacgtcgatgaacagca |
| 1781_MfeI_UF | atatcaattgtcaggaccatcggagggacttc |
| 1781_NcoI_UR | atatccatgggccatcctcagggactagct |
| 1781_AgeI_DF | atataccggtggcggagaggaaggccaataaga |
| 1781_SacI_DR | atatgagctccgcgctcgatcagaagagtgacc |
| 1778_MfeI_UF | tatacaattggtcgtagccccagccgtataaa |
| 1778_NcoI_UR | tataccatggctgcagatctgtctcgtcttc |
| 1778_AgeI_DF | ttaaaccggtacgcattgcgaagctggtcgata |
| 1778_SacI_DR | atatgagctctggtgtgtcccgtcggaggta |
| 1792_MfeI_UF | tatacaattgatgatggaagccgccgagtagag |
| 1792_NdeI_UR | tatacatatggagagatggcgatgctcctttc |
| 1792_AgeI_DF | tataaccggtaccagtcggcgatcaatggagtg |
| 1792_SacI_DR | atatgagctctttcacccttgcggacgagtat |
| 1778C_KpnI_G290fw | ctggtctgtttgtggtacctgcgaaacaatacgcctgagaga |
| 1778C_MluI_G290rv | gctcgcatgcactagtacgcgtccacgacctccgtcttcgtc |
| p_1790_AclI_fw | atataacgttcgttcgacgtctctcgatcttcg |
| p_1790_NcoI_rv | atatccatgggttgatcagtggacgcggttcag |
| p_1782_AclI_fw | atataacgttcttcgtgctcgcttggtcagc |
| p_1782_NcoI_rv | atatccatgggttgatcagtggacgcggttcag |
| p_1781_AclI_fw | atataacgttcaacatcatcgagagcgcctgag |
| p_1781_NcoI_rv | atatccatgggttgatcagtggacgcggttcag |
| p_1777_AclI_fw | atataacgttgcacccatctctcaggcgtattg |
| p_1777_NcoI_rv | atatccatgggttgatcagtggacgcggttcag |
| p_1790_AclI_595_fw | atataacgtttcatgcgaagggattggaagctc |
| p_1790_NcoI_rv | atatccatgggttgatcagtggacgcggttcag |
| p_1790_AclI_545_fw | atataacgttccataggagccggacgtgcat |
| p_1790_AclI_493_fw | atagaacgttactcacaattccgctgccgatg |
| p_1790_AclI_460_fw | atataacgttctgcggtgccggaaactcc |
| p_1790_AclI_400_fw | atataacgttgaagacgatcggtccgacacg |
| p_1790_AclI_354_fw | atataacgttgcgcgggcgtttgcctagta |
| p_1790_AclI_314_fw | atataacgttgcgtatccatgaccttccattcg |
| p_1790_AclI_244_fw | atataacgtttcataccgcagcgcacatgg |
| p_1790_AclI_203_fw | atataacgttgaggcgtcggctgtccatcag |
| p_1790_AclI_111_fw | atagaacgttatcaagaattgcggcgctgatc |
| p_1790_AclI_54_fw | atataacgttcggtaggaaacagggagtgagcc |

**Table 3.4:** Continued

| | |
|---|---|
| p_1782_AclI_206_fw | atagaacgttacaggcccatctcagagcgaaaa |
| p_1782_NcoI_rv | atagccatggacgagcgaacgaggattgagcat |
| p_1782_AclI_147_fw | atataacgttttgatgccgagcgtccaaaat |
| p_1782_AclI_117_fw | atataacgttcatctcttgcagacgcctaccc |
| p_1782_AclI_49_fw | atataacgttgaacggaagacgccacatgaacc |

* Underlined sequences indicate restriction sites. The PCR product generated using primers with no underlined sequences were cloned using Gibson cloning.

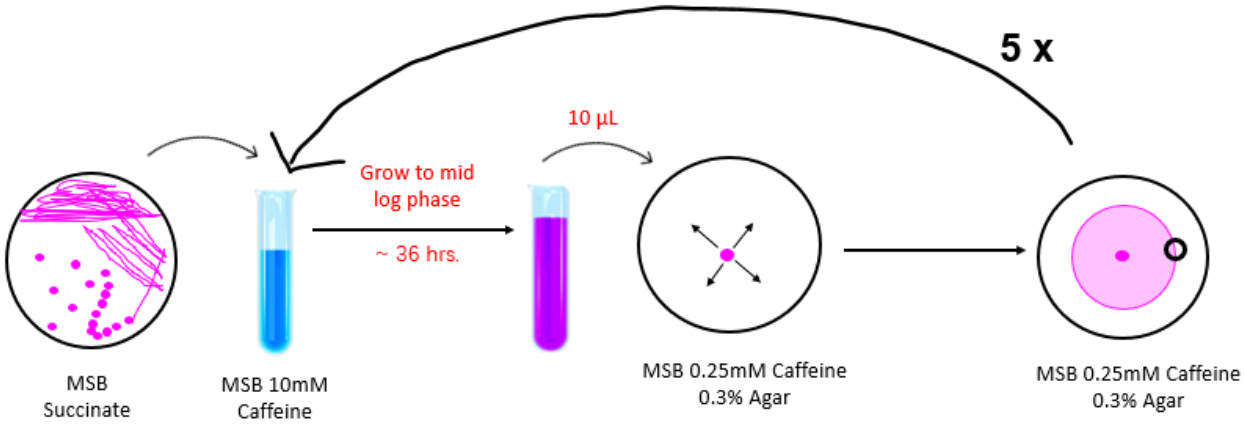



**Figure 3.6**: Growth in soft agar swim plates to enrich for a population of highly motile *M. populi* PINKEL cells. A seed culture of PINKEL was grown in 10 mL of MSB minimal medium containing 10 mM caffeine. Once the culture reached late exponential phase ($OD_{600}$ ~ 0.9) 10 µL was pipetted into the center of an MSB soft agar (0.3% Noble agar) Petri plate containing a low concentration of caffeine (0.25 mM). A chemotactic response is visualized as a sharp pink ring of growth that formed as cells multiplied in number and swam away from the point of inoculation following the gradient that was generated as the cells metabolized the caffeine. Once the cells moved 20 mm away from the point of inoculation (in ~ 10 days), the wide end of a sterile P1000 pipette tip was used to obtain a plug of agar containing cells from the edge of outer ring. The plug was immediately inoculated into 10 mL of MSB containing 10 mM caffeine. This enrichment process was repeated five times after which the populations swam 20 mm in 2 days rather than 10 days.

## Culture media and growth conditions

*Escherichia coli* strains were grown aerobically in Lysogeny broth (LB) or on LB agar plates at 37°C (5 g Bacto tryptone, 2.5 g yeast extract, 2.5 g NaCl with or without 15 g Bacto Agar per 1 L distilled water). *Methylorubrum populi* PINKEL was grown aerobically at 30°C in a phosphate buffered minimal medium (MSB) [Stanier et al., 1966] with the following modifications: 1. The Hutner's base (trace elements solution) stock solution was filter sterilized and added to the medium after autoclaving, and 2. The ammonium sulfate solution was omitted when testing whether PINKEL could use methylxanthines as sole nitrogen sources. MSB was supplemented with 10 mM succinate, 124 mM methanol, 5- or 10-mM caffeine, 5 mM theobromine, 5 mM paraxanthine, 5 mM 7-methylxanthine or a combination of 5 mM succinate and 5 mM methylxanthine or increasing concentrations of formaldehyde as indicated in the text or figure legends. To select and maintain plasmids in *M. populi* or *E. coli* strains, the following concentrations of antibiotics were added when appropriate: 10 µg/ml tetracycline, 50 µg/mL kanamycin, or 100 µg/ml ampicillin. For complementation studies, cumate at a final concentration of 30 µg/mL was added to the medium to induce expression of the gene cloned downstream of a cumate-inducible promoter in the complementation plasmid pLC290 (described in more detail below). During the procedures to generate kanamycin insertion mutants, 10% sucrose was added to counter-select against cells containing a single crossover (procedure described in more detail below) [Chubiz et al., 2013]. Liquid cultures were incubated at 200 rpm in a shaking incubator.

## Cloning and DNA manipulations

All molecular biology kits mentioned below were purchased from Thermo Fisher Scientific (Chicago IL). Genomic DNA from *M. populi* PINKEL was isolated using the PureLink Genomic DNA Mini Kit, and plasmids were purified using a GeneJet plasmid Miniprep Kit. Manipulations of DNA fragments, plasmids, and transformation of *E. coli* strains were carried out by standard methods [Sambrook et al., 1989]. All PCR amplifications were performed using Pfu polymerase in Pfu reaction buffer (200 mM Tris-Cl (pH 8.8), 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 1% Triton X-100, 1 mg/ml bovine serum albumin, 20 mM $Mg_2SO_4$) under the following conditions: 95°C denaturation, 65-60°C annealing (decreasing by 1°C for the first 5 cycles), 72°C elongation, with an elongation time of 1 min/kb of PCR product. All PCRs were set-up for a total of 30 cycles with the initial denaturation and final extension periods each being 10 minutes. Since the genome of PINKEL has a high G+C content, a final concentration of 3% DMSO was added to each PCR as described for other pink-pigmented methylotrophic bacteria [Vu et al., 2021]. PCR products were purified using a GeneJet Gel Extraction Kit and amplicons digested with restriction endonucleases were purified using a GeneJet PCR Purification Kit. Restriction endonucleases used to digest each PCR product are specified in the primer list (Table 2). Some amplicons were cloned into their corresponding plasmid following the recommendations stated in the NEB builder HIFI DNA assembly cloning kit manual. Gene manipulations were corroborated via standard colony PCR procedures using the Promega GoTaq Green (2x) Master Mix. The sequences of cloned PCR products were verified by fluorescent automated DNA

sequencing using the University of California Davis DNA Sequencing Facility with an Applied Biosystems 3730 automated sequencer.

## Construction of promoter fusion strains

In this study, we used the promoterless plasmid pAP5 (Figure 3.7A) as the backbone to clone promoter regions of interest [Skovran et al., 2011]. This plasmid has a tetracycline resistance gene cassette as a selection marker, a yellow fluorescent protein gene (*yfp*) as the reporter and a multiple cloning site upstream of the *yfp* gene. To identify the promoter regions involved in expression of the *ndm* genes, to identify transcriptional regulator(s) involved in the regulatory system of the *N*-demethylation pathway, and to identify the inducer(s), four predicted *ndm* promoter regions located upstream of the *ndmA, ndmB, ndmC,* and *ndmD* genes were amplified (Figure 3.8) and cloned into pAP5. The primers used to amplify each *ndm* promoter region are provided in Table 3.4 and the amplified sequences are provided in Appendix A. The resulting amplicons are referred to as the wild-type *ndm* promoter regions in this study. The AclI and NcoI restriction sites were used to digest and clone the *ndm* promoter regions into pAP5. Generation of the correct pAP5-*ndm* promoter constructs was verified via DNA sequencing analysis. To narrow down the nucleotides within the wild-type *ndm* promoter regions that may be relevant for transcriptional regulation, decreasing lengths of wild-type promoter regions were amplified and cloned as described above. All the different pAP5-*ndm* promoter constructs were mated into the corresponding PINKEL strains following the methods shown in Figure 3.9.

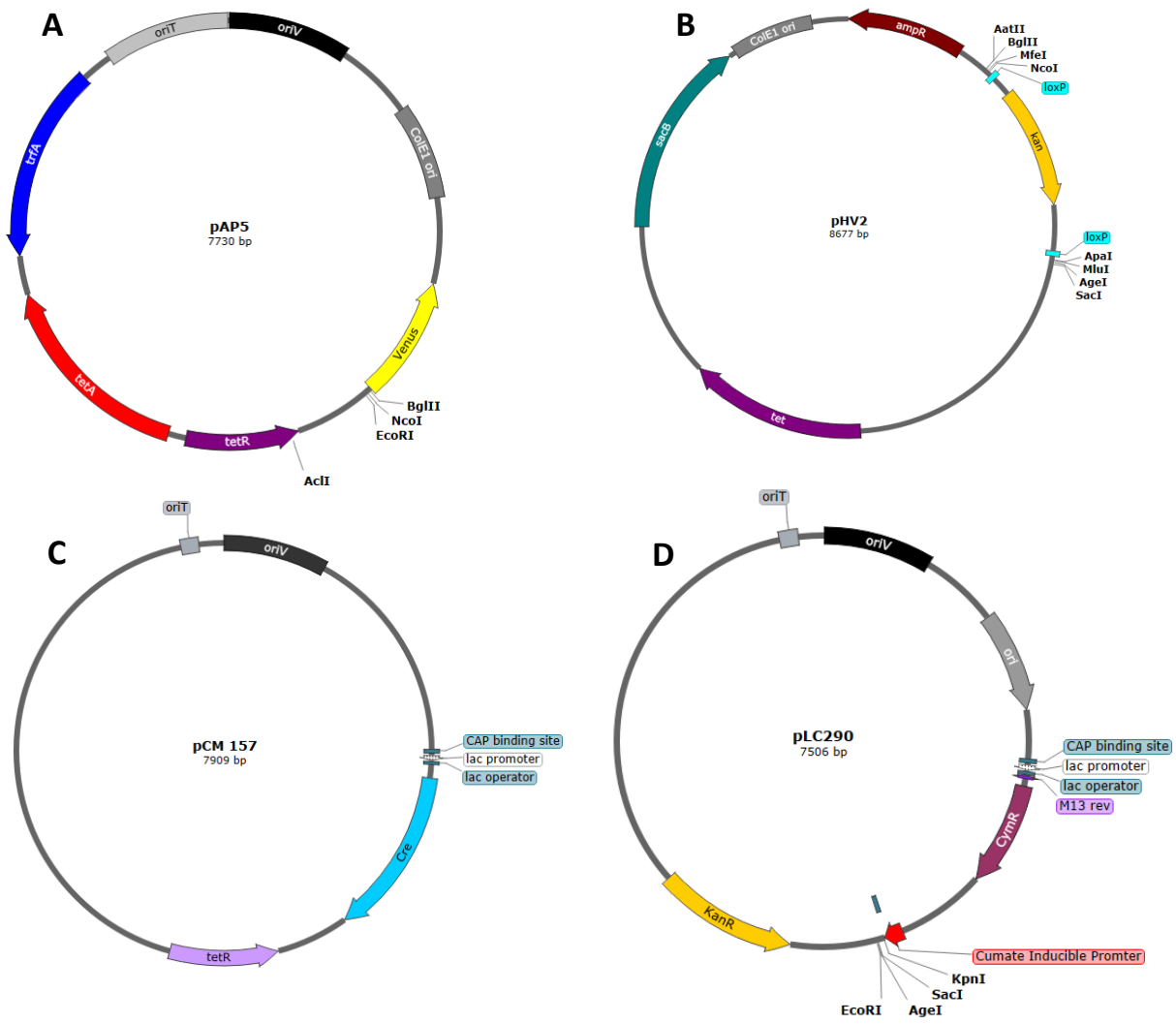



**Figure 3.7**: Vector maps of plasmids used in this study. (A) pAP5, (B) pHV2, (C) pCM157, and (D) pLC290

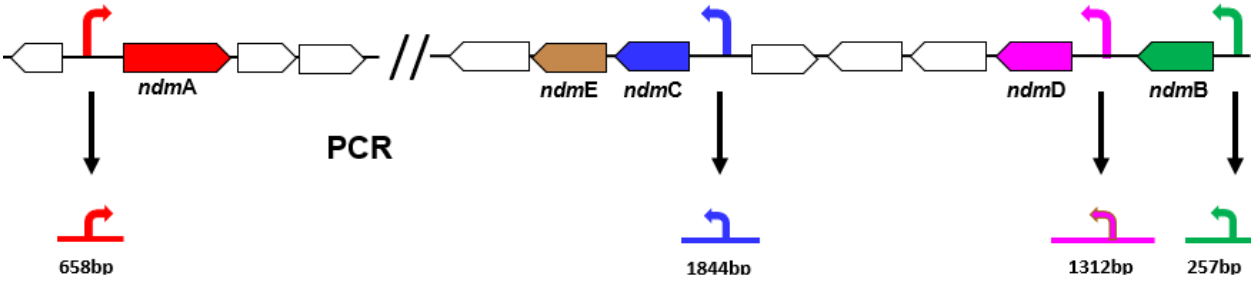



**Figure 3.8**: The four predicted *ndm* promoter regions that were amplified and cloned into pAP5 for transcriptional promoter fusions assays.

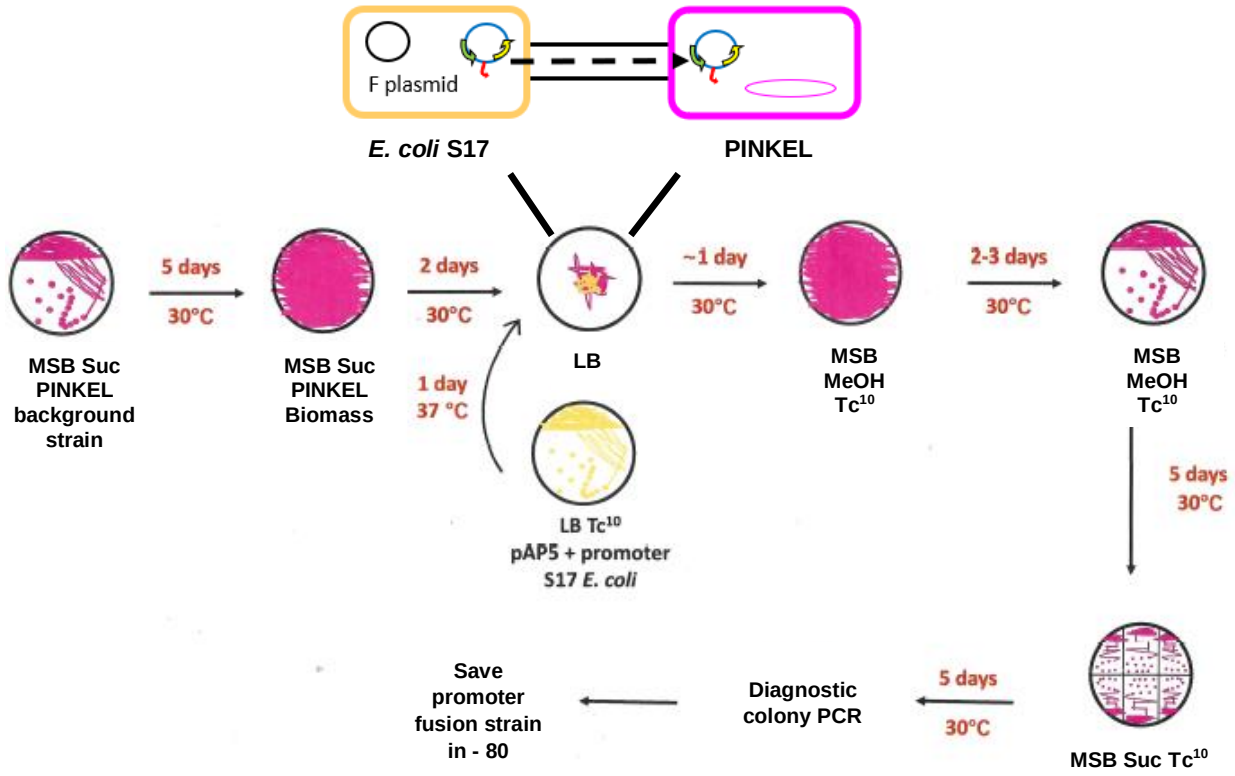



**Figure 3.9**: Method for mating pAP5-*ndm* promoter fusion plasmids into the desired *M. populi* PINKEL background strain. A few swabs of *M. populi* cells from a biomass plate were inoculated together with one colony of *E. coli* S17-1 containing the pAP5-*ndm* promoter plasmid on an LB agar plate. The cells were allowed to mate for 1 day (~24 h) at 30°C. An MSB agar plate containing methanol (1% MeOH) and 10 ug/ml tetracycline ($Tc^{10}$) was used to select for *M. populi* cells that received the promoter fusion plasmid. Isolated colonies were passed onto a fresh MSB MeOH $Tc^{10}$ plate and then onto an MSB succinate (Suc) $Tc^{10}$ plate until a pure culture was obtained. Diagnostic colony PCR was performed to verify that *M. populi* cells had the promoter fusion plasmid before saving a freezer stock.

## Construction of mutant strains

The method used to generate all markerless gene deletion mutants listed in Table 1 is divided into two stages (Figure 3.10): 1. Generation of a kanamycin insertion mutant via an allelic exchange mechanism and 2. Deletion of the kanamycin resistance cassette introduced during the first stage. The allelic exchange suicide plasmid pHV2 (Figure 3.7B) [Roszczenko-Jasinska et al., 2020] was used during the first stage to generate the kanamycin insertion mutants (Figure 3.10A). This vector has a *sacB* gene to allow for counter-selection against single crossovers and contains a kanamycin

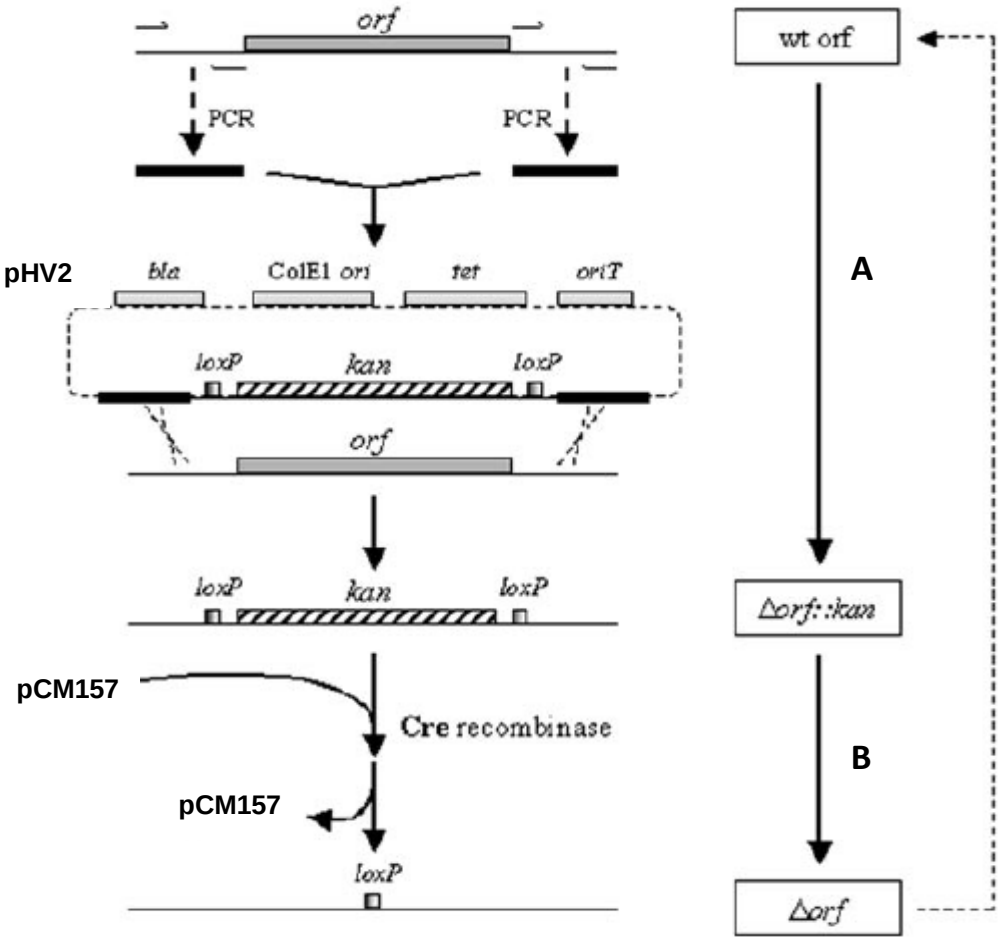



**Figure 3.10**: Strategy to generate markerless gene deletions in *M. populi*. (A) DNA fragments upstream and downstream of the target gene are amplified by PCR and cloned into pHV2. Allelic exchange leads to a kanamycin insertion mutant (B), which can then be rendered unmarked through the introduction of the *cre* expression plasmid pCM157. The process can then be repeated for a second target gene to generate a strain bearing multiple genetic manipulations.

Source: adapted from Marx and Lidstrom, 2002.

resistance cassette flanked by two *loxP* sites, which are flanked by two multiple cloning sites. The upstream and downstream regions of each target gene were amplified via PCR, cloned into pHV2, and the resulting plasmid was used to transform *E. coli* S17-1. The steps necessary to mate the engineered plasmid into PINKEL and select for cells that underwent allelic exchange that resulted in a kanamycin insertion mutant (Figure 3.11) are similar to those previously described [Marx and Lidstrom, 2002; Vu et al., 2021] with the following modifications: 1. A final concentration of 10% sucrose was used

to select against cells containing single crossovers and 2. Methanol was used as the sole carbon source to select for PINKEL cells after the biparental mating step instead of selecting rifamycin resistance. To remove the kanamycin resistance cassette and generate makerless deletion mutants during the second stage (Figure 3.10B), the tetracycline resistance plasmid pCM157 (Figure 3.7C) was used [Marx and Lidstrom, 2002]. This vector encodes the Cre recombinase, which is necessary to catalyze the recombination of the *loxP* sites flanking the kanamycin resistance cassette. This recombination leads to the unmarked removal of the kanamycin resistance cassette. The steps necessary to select for gene deletion strains are summarized in Figure 3.12.

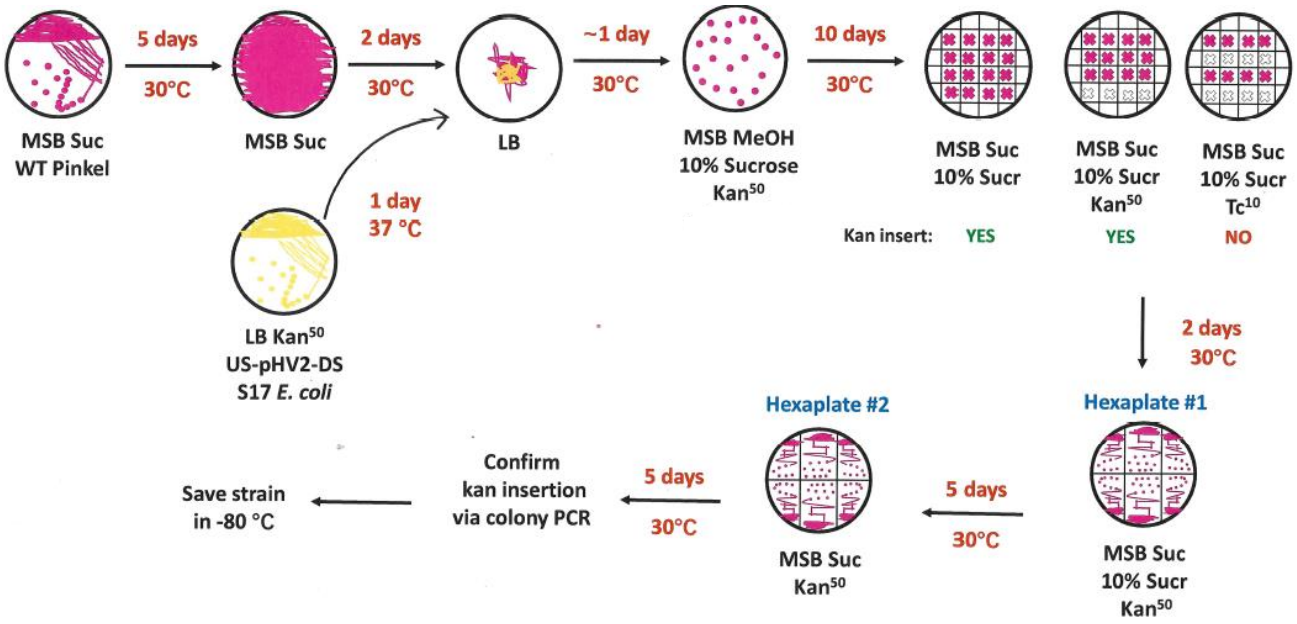



**Figure 3.11**: Method for mating pHV2 plasmids into *M. populi* to select for cells that underwent allelic exchange leading to kanamycin insertion mutants. A few swabs of *M. populi* cells from a biomass plate were mated with one colony of *E. coli* S17-1 [containing an upstream-pHV2-downstream (US-pHV2-DS) plasmid engineered to knock out a target gene] on an LB agar plate. The cells were allowed to mate for 1 day (~24 h) at 30°C. An MSB agar plate containing methanol (1% MeOH), 10% sucrose, and 50 µg/ml kanamycin ($kan^{50}$) was used to select for *M. populi* cells with a kanamycin insertion mutation. Candidate colonies were patched onto MSB 10 mM succinate 10% sucrose, MSB 10 mM succinate 10% sucrose 50 µg/ml kanamycin, and MSB 10 mM succinate 10% sucrose 10 µg/ml tetracycline plates to screen for double crossovers and then passed onto fresh MSB succinate 50 µg/ml plates until a pure culture was obtained. Diagnostic colony PCR was performed to verify that *M. populi* cells had a kanamycin cassette inserted within the target gene before saving a freezer stock.

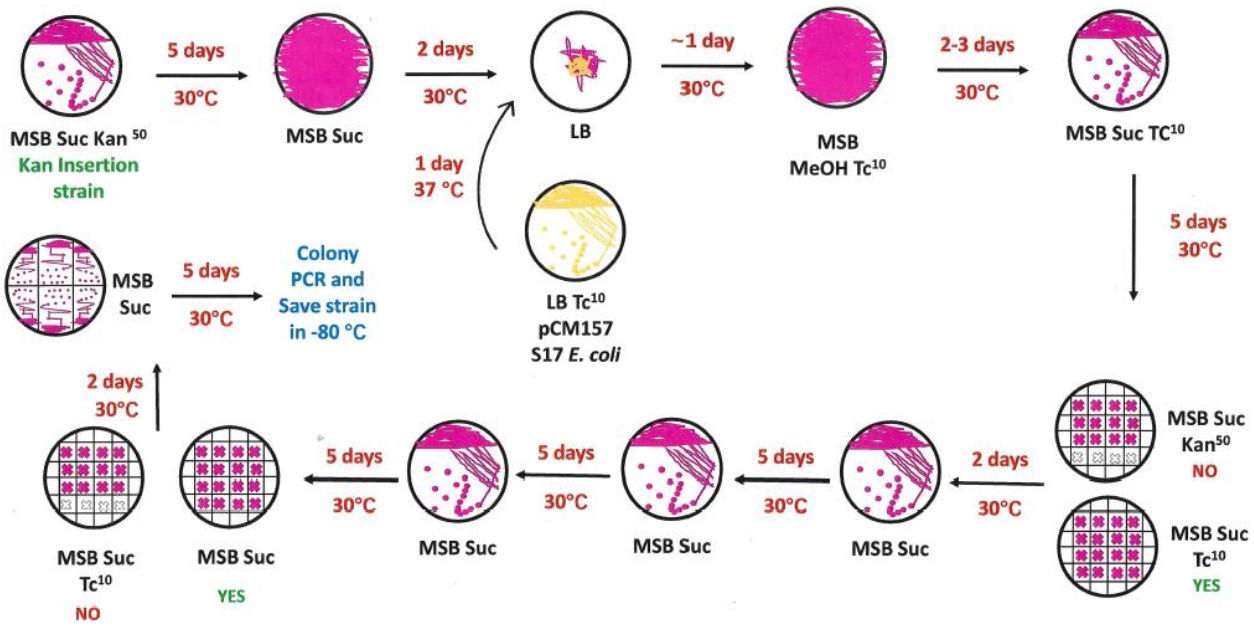



**Figure 3.12**: Method for resolving the insertion of the kanamycin resistance cassette using pCM157 to achieve markerless deletions. Matings between the *M. populi* strains bearing a kanamycin resistance cassette insertion and *E. coli* S17-1 containing the pCM157 plasmid were done on LB agar plates. The cells were allowed to mate for 1 day at 30°C. An MSB agar plate containing methanol (1% MeOH) and 10 µg/ml tetracycline ($Tc^{10}$) was used to select for *M. populi* cells that received the pCM157 plasmid. This plasmid carries the gene encoding the Cre recombinase, which is necessary to catalyze the recombination of the *loxP* sites flanking the kanamycin resistance cassette. The recombination leads to the unmarked removal of the kanamycin resistance cassette. Colonies were patched onto MSB succinate plates containing either kanamycin or tetracycline to screen for colonies that had lost the kanamycin resistance cassette. Kanamycin-sensitive, tetracycline-resistant cultures were passed several times on MSB succinate plates with no antibiotic to cure the pCM157 plasmid. Curing of the plasmid was confirmed by patching colonies onto MSB succinate medium with and without tetracycline. Tetracycline-sensitive cultures were streaked onto fresh MSB succinate plates and diagnostic colony PCR was performed to confirm that markerless gene deletion was achieved before saving strains as freezer stocks.

## Construction of complementation strains

To complement mutants with a functional copy of a deleted gene, the vector pLC290 was used (Figure 3.7D) [Chubiz et al., 2013]. This plasmid carries a kanamycin resistance gene and has a multiple cloning site immediately downstream of a cumate-inducible promoter. The KpnI and MluI restriction sites were used to linearize the pLC290 plasmid and an NEB builder HIFI DNA assembly cloning kit was used to clone the PCR amplified genes used for complementation (complete sequences of these

genes are provided in Appendix A). All PCR products were sequenced before cloning to ensure there were no point mutations present. A protocol similar to that shown in Figure 3.9 was used to introduce complementation plasmids into *M. populi* mutants except that kanamycin was added to the medium instead of tetracycline to maintain the plasmid. Empty vector (pLC290 with no cloned gene) was mated into wild-type PINKEL to serve as a control.

## Determination of growth rates

Growth experiments were conducted at 30°C in borosilicate glass tubes (20- by 150-mm) using a minimum of three biological replicates. Tubes were placed in test tube racks at an approximate angle of 60° to ensure maximum aeration. The cultures were shaken at 30 ºC, 200 rpm using a shaking incubator. To obtain a seed culture, *M. populi* cells were grown to late exponential phase in 6 ml of MSB medium containing 10 mM succinate as the carbon source. Approximately 200 µl of the seed culture was transferred into 10 ml of MSB plus 5 mM of the test methylxanthine. Optical density at 600 nm ($OD_{600}$) was recorded every three hours using a Spectronic 20D spectrophotometer containing an adapter to read absorbance in 20 x 150-mm tubes. Measurements were taken until three consecutive measurements in the stationary phase were recorded.

## Transcriptional promoter fusion assays

Seed cultures of the engineered pAP5-*ndm* promoter fusion strains were grown as described in the growth rate determination section, with the exception that tetracycline was added to the MSB succinate medium to maintain the promoter fusion plasmids. The cells were subcultured into uninduced (succinate only) and induced (succinate plus methylxanthine) conditions (Figure 3.13). Once the cells reached an $OD_{600}$ between 0.9 to 0.93, a 200 µl aliquot was taken, and relative fluorescence units (RFU) were measured in an optical-bottom black 96-well plate using a Victor X model 3 plate reader (PerkinElmer, Waltham, MA). The excitation spectra were set to 485 nm, and the emission spectra was set to 535 nm. The emission aperture, CW-lamp energy, and exposure time were set to small, 400 volts, and 3 seconds, respectively. The $OD_{600}$ measurements obtained from the Victor X model 3 plate reader were used to normalize fluorescence to culture density ($RFU/OD_{600}$). Background expression was defined as the RFU/OD detected from the wild-type PINKEL strain containing the promoterless *yfp* fusion construct. At least three biological replicates were carried out for each tested strain and condition. A time lapse promoter fusion in which expression was measured every hour for a duration of 36 hours was performed using the wild-type strain containing the wild-type *ndmA*-pAP5 promoter region to determine the optimal experimental conditions as well as the optimal $OD_{600}$ at which to measure expression.

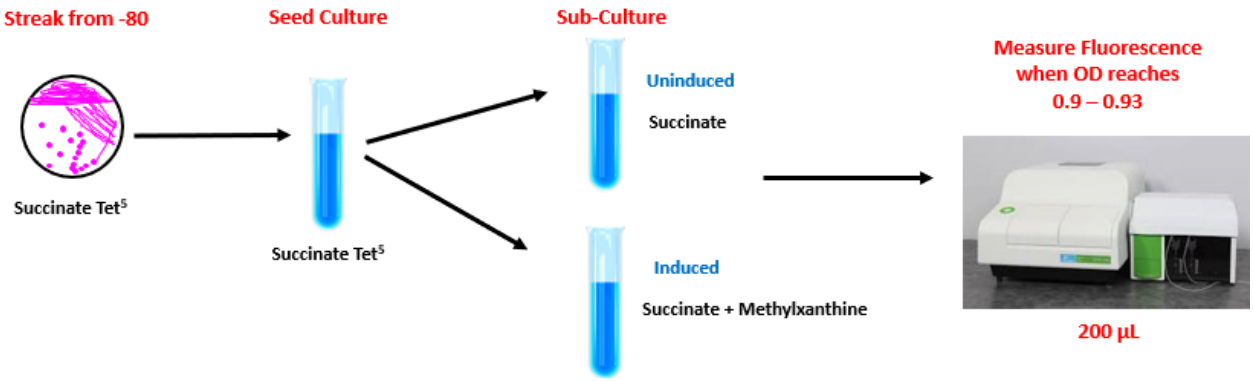



**Figure 3.13**: Transcriptional promoter fusion assay workflow. Seed cultures were used to inoculate uninduced (succinate + tetracycline) and induced (succinate + methylxanthine + tetracycline) cultures. Once cells reached an $OD_{600}$ between 0.9 to 0.93, a 200 µl aliquot was taken to measure fluorescence.

## Results

### Identification of *ndm* promoter regions

To study the transcriptional regulatory system that controls the expression of the *ndm* genes, the number and location of the *ndm* promoter regions were first determined. As shown in Figure 3.8, we hypothesized that promoters were located upstream of the *ndmA, ndmB, ndmC,* and *ndmD* genes. To test for the presence of a promoter, the four potential *ndm* promoter regions were PCR-amplified and cloned into the plasmid pAP5, upstream of a promoterless yellow fluorescent protein gene (*yfp)* and mated into the *M. populi* PINKELnc strain. The wild-type differential expression pattern was measured for each strain as an output of YFP via a transcriptional promoter fusion assay. When the cells were grown in the absence of methylxanthines (uninduced) no expression was detected from any of the promoter regions (Figure 3.14A). In contrast, when *M. populi* PINKELnc was grown in the presence of caffeine (induced) expression from the *ndmA, ndmB,* and *ndmC* promoter regions was detected (Figure 3.14B). The expression level from the *ndmD* promoter region was similar to the expression level obtained from the promoterless vector.

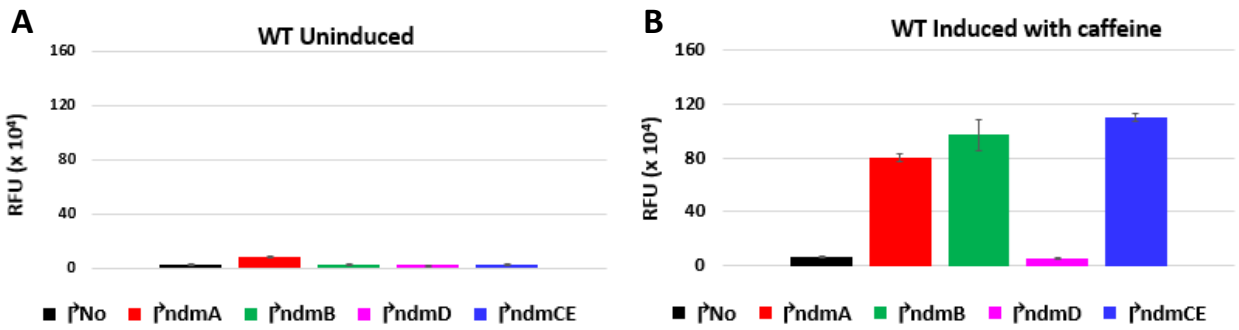
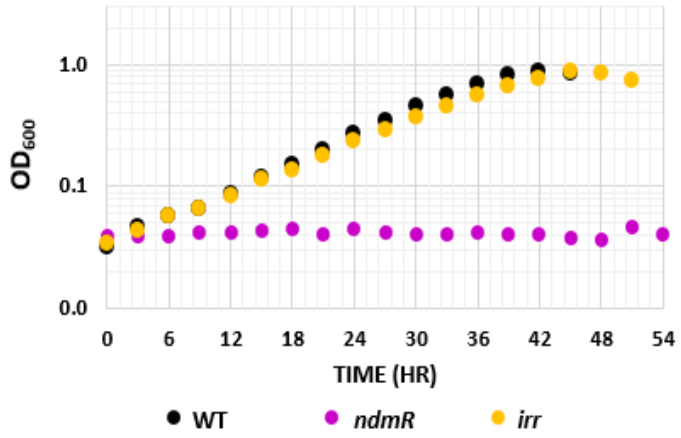



**Figure 3.14**: Transcriptional promoter fusion assay using the *M. populi* PINKELnc strain containing each of the four predicted *ndm* promoter regions. (A) Expression pattern when cells were grown in the absence of caffeine (uninduced). (B) Expression pattern when cells were grown in the presence of 5 mM caffeine (induced). A promoterless plasmid (black bars) was also mated in *M. populi* PINKELnc as a control for background expression. At least three biological replicates were done for each strain and test condition.

## Identification of transcriptional regulator of the *N*-demethylation pathway in *M. populi* PINKEL

Due to its localization in the vicinity of the *ndm* genes, one of the natural candidates to regulate the *N*-demethylation pathway in PINKEL is NdmR, a putative LysR-type transcriptional regulator encoded by peg_1778 (Figure 3.1B). Another candidate for the regulation of this pathway is a putative iron-responsive regulator named Irr (encoded by peg_1792), which is encoded in the same gene cluster as the *ndmA* gene (Figure 3.1B). To determine if either of these two proteins is involved in regulating the *ndm* genes, mutant strains lacking either *ndmR* or *irr* were constructed and their ability to grow on caffeine was tested. As shown in Figure 3.15, when *ndmR* was deleted, *M. populi* PINKEL was no longer able to grow on caffeine. However, deleting the *irr* gene did not affect the growth of *M. populi* PINKEL on caffeine.




| Strain | Growth Rate ($hr^{-1}$) | Doubling Time (hr) |
|---|---|---|
| WT | 0.093 | 7.4 |
| Δ *ndmR* | No growth | No growth |
| Δ *irr* | 0.083 | 8.3 |

**Figure 3.15**: Growth of wild-type PINKELnc (Black circles), and mutant strains Δ*ndmR* (purple circles) and Δ*irr* (orange circles) on 10 mM caffeine. The growth rate and doubling time for each strain is also provided.

To corroborate if either NdmR or Irr functions as a positive or negative transcriptional regulator of the *ndm* genes, the three identified *yfp-ndm* promoter fusion constructs (Figure 3.16A) were mated into Δ*ndmR* and Δ*irr* backgrounds and their differential expression patterns were measured and compared to that of the wild-type *M. populi* PINKELnc strain. There was no induction from any promoter in any of the strain backgrounds when cells were grown in the absence of caffeine (Figure 3.16B). In contrast to the wild-type strain, *M. populi* PINKELnc cells lacking the *ndmR* gene did not turn on expression from any of the three *ndm* promoter regions, even when grown in the presence of caffeine (Figure 3.16C). However, complementing the Δ*ndmR* strain with a functional copy of the *ndmR* gene restored the wild-type expression patter for all three promoter regions during inducing conditions. In contrast, deletion of the *irr* gene did not affect the expression patterns of any of the *ndm* promoters.

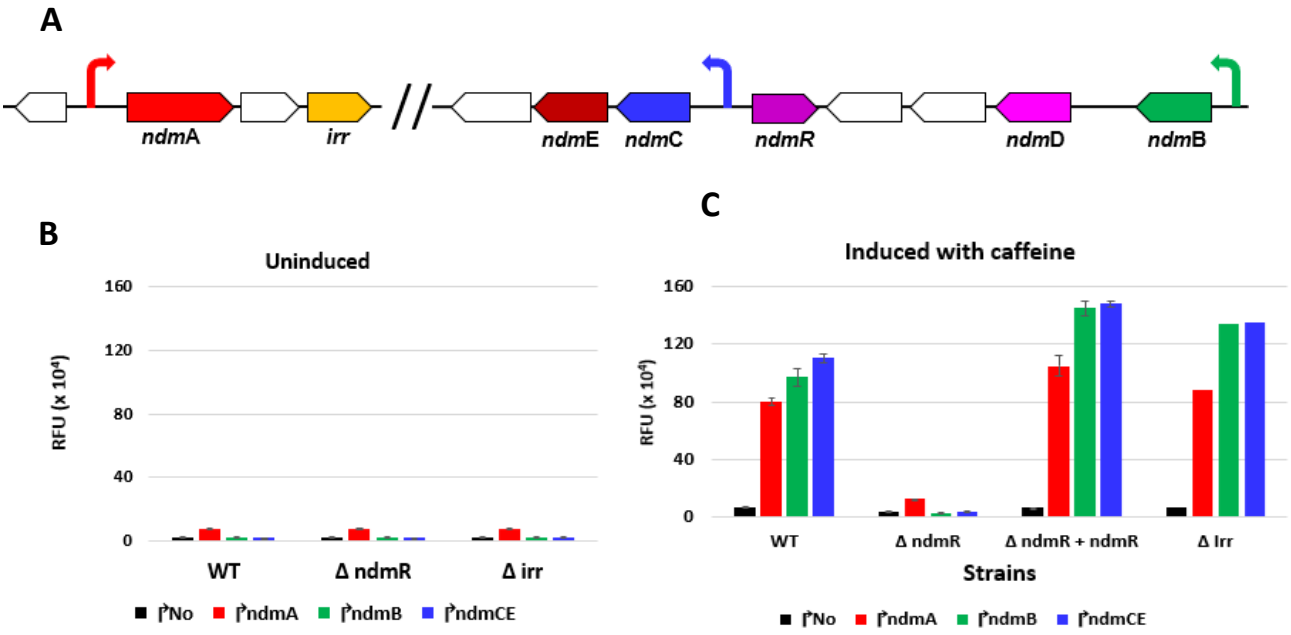



**Figure 3.16**: Transcriptional promoter fusion assay using the wild-type, Δ*ndmR*, Δ*irr*, and complemented Δ*ndmR* strains containing the three identified *ndm* promoter regions. (A) Reference map showing the three identified *ndm* promoter regions. (B) Expression pattern when cells were grown in the absence of caffeine (uninduced). (C) Expression pattern when cells were grown in the presence of 5 mM caffeine (induced). A promoterless plasmid (black bars) was used as a control for background expression.

## Identification of the inducers of the *N*-demethylation pathway

Since NdmD is the only reductase in *M. populi* PINKELnc that participates in the *N*-demethylation of methylxanthines, deletion of *ndmD* blocks every step of the *N*-demethylation pathway (Figure 2.19). Thus, methylxanthines that are provided to an Δ*ndmD* mutant strain cannot be converted to downstream metabolites. To determine the identity of the inducer, the *ndm* promoter fusion plasmids were mated into the Δ*ndmD* mutant and expression was measured in the presence or absence of each metabolite of the *N*-demethylation pathway (Figure 3.17). Caffeine, theobromine, paraxanthine and 7-methylxanthine upregulated expression from all three *ndm* promoter

regions, although at different levels. On the other hand, there was no expression from any of the *ndm* promoters in the absence of added methylxanthines or in the presence of xanthine.

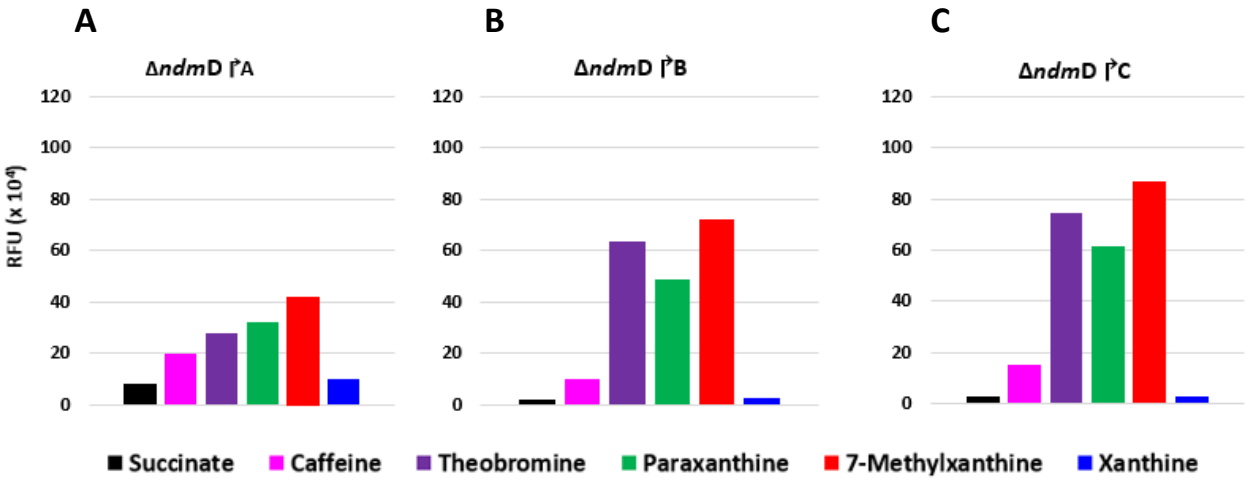



**Figure 3.17**: Transcriptional promoter fusion assay using the Δ*ndmD* strain as the host strain. Each methylxanthine (plus 5 mM succinate) produced during the *N*-demethylation pathway was individually tested as a potential inducer at a final concentration of 2 mM. Expression pattern for Δ*ndmD* cells containing the (A) *ndmA* promoter region, (B) *ndmB* promoter region, or (C) *ndmC* promoter region were tested. Growth with succinate alone (5 mM) served as the control for no expression.

## Analysis of *ndm* promoter regions

To narrow down the fragment of DNA within the *ndm* promoter regions that may contain relevant nucleotides for the binding of the transcriptional regulator or RNA polymerase, decreasing fragments of the *ndmA* and *ndmB* promoter regions were amplified and cloned into the wild-type *M. populi* PINKELnc strain. The engineered strains were used in transcriptional promoter fusion assays to identify DNA regions required for upregulation of the *ndmA* and *ndmB* promoters. For the *ndmA* promoter region, upregulation similar to that of the wild-type fragment (658 bp) was observed with fragments of at least 354 base pairs (Figure 3.18A). For the *ndmB* promoter region,

expression no longer occurred with fragments shorter than 206 base pairs (Figure 3.18B).

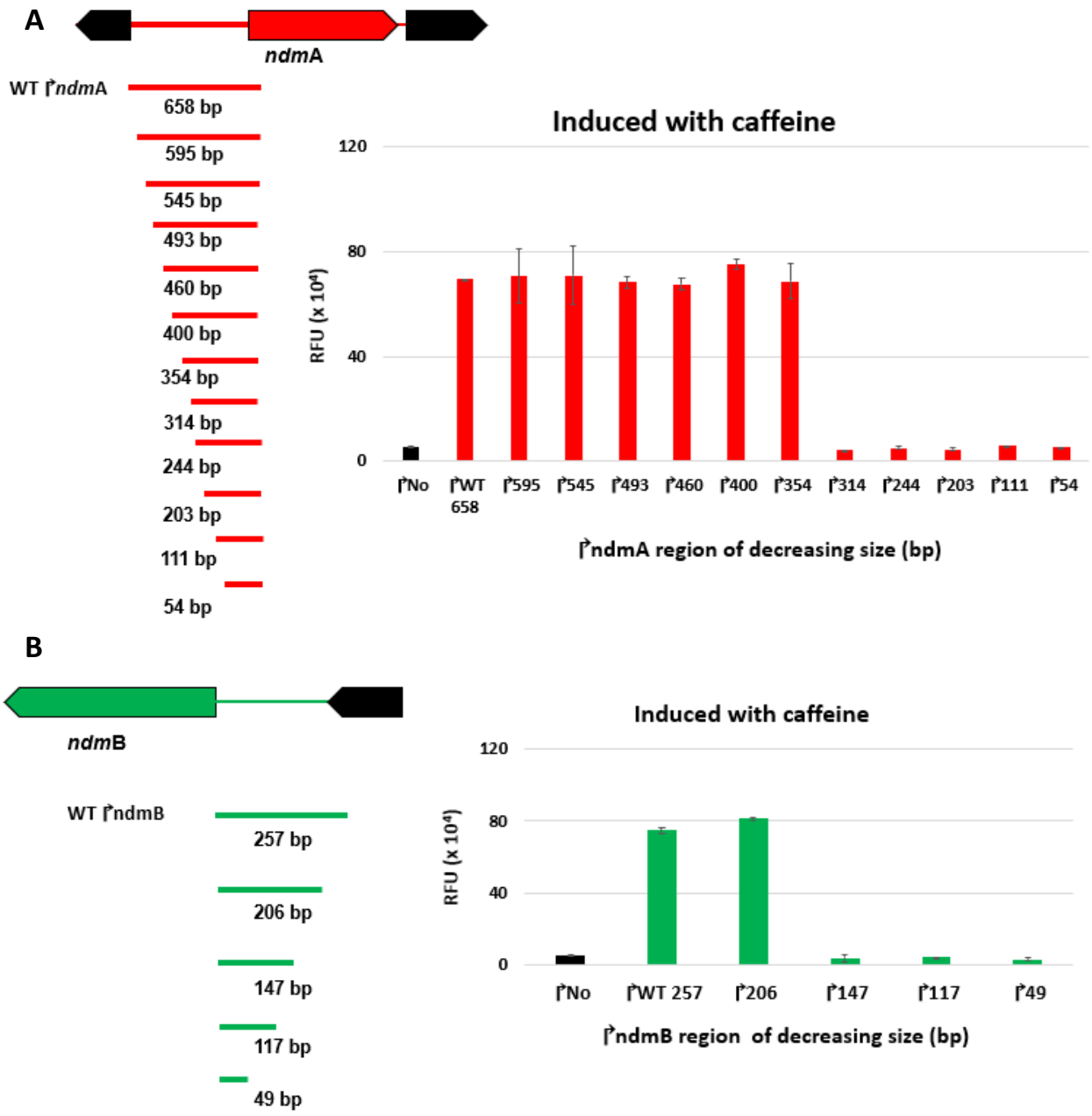



**Figure 3.18**: Transcriptional promoter fusion assay using the wild-type *M. populi* PINKELnc strain containing either (A) the *ndmA* promoter region of decreasing size or (B) the *ndmB* promoter region of decreasing size. The promoterless pAP5 plasmid as well as the wild-type *ndmA* promoter (658 bp) and wild-type *ndmB* promoter (257 bp) were used as controls. Expression was tested in the presence of 5 mM caffeine.

## Discussion

### The *ndm* genes are expressed from three different promoters

Characterizing the *ndm* promoter region(s) as well as the protein(s) that regulate expression of *ndm* genes may aid in the optimization and engineering efforts of caffeine-degrading microbes for the biotechnological applications. The first step taken to make progress towards these efforts was to determine the location and number of promoter regions controlling the expression of the *ndm* genes. Transcriptional promoter fusion experiments using the wild-type *M. populi* PINKELnc as the background strain revealed that there are three *ndm* promoters: the first located upstream of *ndmA*, another upstream of *ndmB*, and the last upstream of *ndmC* (Figure 3.14). Our data indicate that three transcripts are generated in the presence of methylxanthines as shown in Figure 3.19. Expression from all three promoters was found to be inducible, since expression from each was upregulated when cells were grown in the presence of methylxanthines. Additionally, the basal level of expression from the *ndmA* promoter was approximately 3-fold higher than those of the *ndmB* and *ndmC* promoters (Figure 3.14A). The basal expression level of *ndmA* may be higher than those of the other *ndm* genes because NdmA catalyzes the first step of the pathway. Having slightly higher levels of NdmA during uninduced conditions may allow cells to detect and metabolize any caffeine that may be present in the environment, allowing them to generate methylxanthine intermediates that act as stronger inducers than caffeine (discussed in more detail below).

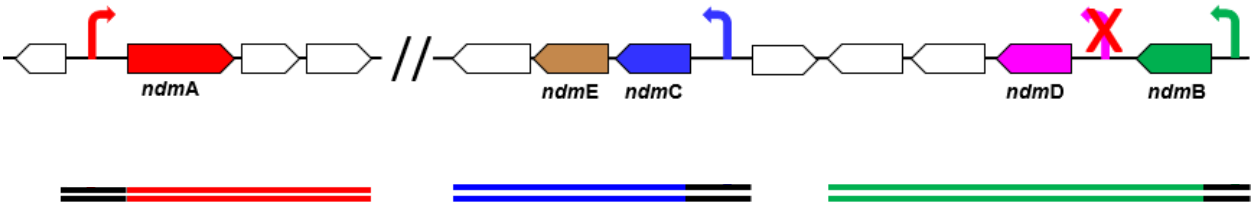



**Figure 3.19**: Predicted *ndm* transcripts generated in the presence of methylxanthines. The red transcript contains the message for NdmA, the blue transcript contains the message for NdmC and NdmE, and the green transcript likely encodes NdmB and NdmD.

## NdmR positively regulates the *ndm* genes

Due to their location in the vicinity of the *ndm* genes (Figure 3.16A), the natural candidates to regulate the *N*-demethylation pathway were NdmR and/or Irr. In contrast to NdmR, deletion of the *irr* gene did not have an effect on the wild-type growth or the expression patterns of *M. populi* PINKEL in the presence of caffeine (Figure 3.15 and Figure 3.16). Thus, it was concluded that Irr does not play a role in the regulation of the *N*-demethylation pathway genes. The first piece of evidence pointing towards NdmR functioning as a positive regulator of the *N*-demethylation pathway was the observation that the *ndmR* deletion mutant was no longer able to grow in on caffeine (Figure 3.15). As expected, complementing the Δ*ndmR* mutant strain with a functional copy of the *ndmR* gene restored the ability of *M. populi* PINKEL to grow on caffeine. Additional evidence came from transcriptional promoter fusion data. These experiments showed that the Δ*ndmR* mutant was unable to upregulate any of the *ndm* genes in the presence of caffeine (Figure 3.16C), whereas wild-type expression patterns from all *ndm* promoters were restored when the Δ*ndmR* mutant was complemented. Together, the growth and transcriptional promoter fusion data provide evidence that NdmR positively

regulates all *ndm* genes. However, whether NdmR directly binds inducer molecules to regulate the *N*-demethylation pathway genes or if regulation is indirect by controlling expression of other regulators still needs to be determined. Determining the details of the mechanism of regulation exerted by NdmR may help in the synthetic biology and metabolic engineering efforts to use bacterial cells as factories to produce medically relevant methylxanthines and bioremediation applications.

## Methylxanthines induce expression of *ndm* genes

To identify the specific metabolite(s) being sensed by NdmR, each of the three *ndm* promoter fusions was individually introduced into the Δ*ndmD* mutant and YFP levels were determined after growth in the presence and absence of each metabolite of the *N*-demethylation pathway. The data showed that all tested xanthines with at least one methyl group attached were capable of acting as inducers. At all three *ndm* promoters, 7-methylxanthine (1 methyl group) was the strongest inducer, whereas methylxanthines with two methyl groups (paraxanthine, theobromine) were slightly weaker inducers, and caffeine (3 methyl groups) was the weakest inducer. In contrast, xanthine itself, which lacks all three methyl groups, did not serve as an inducer. In general, regardless of the methylated xanthine inducer provided, the fold change from the *ndmA* promoter was lower than those of the other two *ndm* promoters. However, this is partly due to the higher basal level of expression from the *ndmA* promoter compared to the other two promoters.

## *ndm* promoter regions

In an effort to identify nucleotides within the identified promoter regions that may serve as recognition sequences for the binding of NdmR, transcriptional factors of RNA polymerase, or the ribosomal binding site, decreasing fragments of the wild-type *ndmA* and *ndmB* promoter regions were analyzed via transcriptional promoter fusion experiments. Since there was no upregulation detected for any fragment measuring less than 354 base pairs for the *ndmA* promoter region, important regulatory nucleotides must be located in the region between 354 and 314 base pairs upstream of the *ndmA* gene. The same must be true for the *ndmB* promoter region between 206 and 147 base pairs upstream of the *ndmB* gene. Since the primers were designed in such a way that the PCR fragment got smaller as coming towards the gene of interest, it is likely that the identified regulatory regions represent binding sites for NdmR. However, binding of NdmR must be verified by electrophoretic mobility gel shift assays with purified NdmR. Although our experiments narrowed down the regions within the *ndmA* and *ndmB* promoter sequences where relevant regulatory nucleotides might be found, the specific nucleotides that participate in regulation still need to be characterized using footprinting experiments in combination with bioinformatic tools.

## References

Baugh AC, Momany C, & Neidle EL. (2023). Versatility and complexity: Common and uncommon facets of LysR-type transcriptional regulators. *Annual Reviews in Microbiology* 77:317-339. doi: 10.1146/annurev-micro-050323-040543. PMID: 37285554.

Bruns RF, & Fergus JH. (1989). Solubilities of adenosine antagonists determined by radioreceptor assay. *Journal of Pharmacy and Pharmacology* 41: 590-594.

Byrne GA, Russell DA, Chen X, & Meijer WG. (2007). Transcriptional regulation of the *virR* operon of the intracellular pathogen *Rhodococcus equi*. *Journal of Bacteriology* 189: 5082–5089.

Cao H, Krishnan G, Goumnerov B, Tsongalis J, Tompkins R, & Rahme LG. (2001). A quorum sensing-associated virulence gene of *Pseudomonas aeruginosa* encodes a LysR-like transcription regulator with a unique self-regulatory mechanism. *Proceedings of the National Academy of Science USA* 98: 14613–14618.

Dash SS, & Gummadi SN. (2008). Inducible nature of the enzymes involved in catabolism of caffeine and related methylxanthines. *Journal of Basic Microbiology* 48(4):227-33. doi: 10.1002/jobm.200800004. PMID: 18720498.

Deghmane AE, Giorgini D, Larribe M, Alonso JM & Taha MK. (2002). Down-regulation of pili and capsule of *Neisseria meningitidis* upon contact with epithelial cells is mediated by CrgA regulatory protein. *Molecular Microbiology* 43: 1555–1564.

Fernandez-López R, Ruiz R, de la Cruz F, & Moncalián G. (2015). Transcription factor-based biosensors enlightened by the analyte. *Frontiers in Microbiology* 1:6:648. doi: 10.3389/fmicb.2015.00648. PMID: 26191047; PMCID: PMC4486848.

Gopishetty, SR, Louie TM, Yu CL, & Subramanian MV. (2011). Microbial degradation of caffeine, methylxanthines, and its biotechnological applications, p 44–67. *In* Thatoi HN, & Mishra BB (eds), *Microbial biotechnology methods and applications*. Narosa Publishing House Pvt, Ltd, New Delhi, India.

He H, Yang M, Li S, Zhang G, Ding Z, Zhang L, Shi G, & Li Y. (2023). Mechanisms and biotechnological applications of transcription factors. *Synthetic and Systems Biotechnology*. 8(4): 565-577. doi: 10.1016/j.synbio.2023.08.006.

Henikoff S, Haughn GW, Calvo JM, & Wallace JC. (1988). A large family of bacterial activator proteins. *Proceedings of the National Academy of Science USA* 85: 6602–6606.

Kim J, Kim JG, Kang Y, Jang JY, Jog GJ, Lim JY, Kim S, Suga H, Nagamatsu T, & Hwang I. (2004). Quorum sensing and the LysR-type transcriptional activator ToxR

regulate toxoflavin biosynthesis and transport in *Burkholderia glumae*. *Molecular Microbiology* 54: 921–934.

Kim JH, Kim BH, Brooks S, Kang SY, Summers RM, & Song HK. (2019). Structural and mechanistic insights into caffeine degradation by the bacterial *N*-demethylase complex. *Journal of Molecular Biology* 431(19): 3647-3661. doi: 10.1016/j.jmb.2019.08.004. PMID: 31412262.

Koentjoro MP, & Ogawa N. (2018). Structural studies of transcriptional regulation by LysR-type transcriptional regulators in bacteria*. Reviews in Agricultural Science*, 6. pp. 105-118. ISSN 2187-090X.

Kovacikova G, & Skorupski K. (1999). A *Vibrio cholerae* LysR homolog, AphB, cooperates with AphA at the *tcpPH* promoter to activate expression of the ToxR virulence cascade. *Journal of Bacteriology* 181: 4250–4256.

Li S, Wen J, He B, Wang J, Hu X, & Liu J. (2020). Occurrence of caffeine in the freshwater environment: Implications for ecopharmacovigilance. *Environmental Pollution* 263(Pt B): 114371. doi: 10.1016/j.envpol.2020.114371. PMID: 32217417.

MacLean AM, Haerty W, Golding BG, & Finan TM. (2011). The LysR–type PcaQ protein regulates expression of a protocatechuate–inducible ABC–type transport system in *Sinorhizobium meliloti*. *Microbiology* 157: 2522-2533.

Maddocks SE, & Oyston PCF. (2008). Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* 154(Pt 12): 3609-3623. doi: 10.1099/mic.0.2008/022772-0. PMID: 19047729.

Minoda A, Weber APM, Tanaka K, & Miyagishima SY. (2010). Nucleus-independent control of the rubisco operon by the plastid-encoded transcription factor Ycf30 in the red alga *Cyanidioschyzon merolae*. *Plant Physiology* 154: 1532-1540.

Parales RE, Sharma G, Zhang X, Subuyuj GA, Langner JT, Wright ME, Ditty JL, & Dawson SC. (2019). Draft genome sequence of the caffeine-degrading methylotroph *Methylorubrum populi* Pinkel. *Microbiology Resource Announcements* 8: e01300-19.

Perez-Rueda E, & Collado-Vides J. (2001). Common history at the origin of the position-function correlation in transcriptional regulators in archaea and bacteria. *Journal of Molecular Evolution* 53: 172–179.

Sperandio B, Gautier C, McGovern S, Ehrlich DS, Renault P, Martin-Verstraete I, & Gue´ don E. (2007). Control of methionine synthesis and uptake by MetR and homocysteine in *Streptococcus mutans*. *Journal of Bacteriology* 189: 7032–7044.

Summers RM, Louie TM, Yu CL, Gakhar L, Louie KC, & Subramanian M. (2012). Novel, highly specific *N*-demethylases enable bacteria to live on caffeine and related purine alkaloids. *Journal of Bacteriology* 194: 2041-2049.

Summers RM, Seffernick JL, Quandt EM, Yu CL, Barrick JE, & Subramanian M. (2013). Caffeine junkie: an unprecedented glutathione *S*-transferase-dependent oxygenase required for caffeine degradation by *Pseudomonas putida* CBB5. *Journal of Bacteriology* 195: 3933-3939.

Summers R, Gopishetty S, Mohanty S, & Subramanian M. (2014). New genetic insights to consider coffee waste as feedstock for fuel, feed, and chemicals. *Central Journal of Chemistry* 12: 1271-1279.

Summers RM, Mohanty SK, Gopishetty S, & Subramanian M. (2015). Genetic characterization of caffeine degradation by bacteria and its potential applications. *Microbial Biotechnology* 8: 369-378.

Sun J, & Klein A. (2004). A LysR-type regulator is involved in the negative regulation of genes encoding selenium-free hydrogenases in the archaeon *Methanococcus voltae*. *Molecular Microbiology* 52: 563–571.

Vega FE, Emche S, Shao J, Simpkins A, Summers RM, Mock MB, Ebert D, Infante F, Aoki S, & Maul JE. (2021). Cultivation and genome sequencing of bacteria isolated from the coffee berry borer (*Hypothenemus hampei*), with emphasis on the role of caffeine degradation. *Frontiers in Microbiology* 12: 644768. doi: 10.3389/fmicb.2021.644768.

# Chapter 4

## Summary of completed research and suggestions for future directions

The worldwide consumption of caffeine and related methylxanthines via various food products and pharmaceuticals has made such compounds ubiquitous in the environment. Hence, it comes to no surprise that caffeine-degrading microorganisms, such as yeast, fungi, and bacteria, have been isolated from the environment. Currently, there is a great interest in engineering caffeine-degrading microbes for the bio-decaffeination of food products, production of medically relevant methylxanthines, and bioremediation of contaminated soil and water systems.

Up until the 1970s, caffeine metabolism in microbes had not received much attention. Since then, approximately 71 bacterial strains capable of degrading caffeine have been reported, the majority of which belong to the genus *Pseudomonas*. Most of what is currently known about the enzymes that degrade caffeine has been determined by using *Pseudomonas putida* CBB1 (C8-oxidation pathway) or *Pseudomonas putida* CBB5. While the genes and enzymes for both the *N*-demethylation and C-8 oxidation pathways have been well characterized, little is known about how these genes are regulated at the transcriptional level or about the chemotaxis response towards caffeine and/or related xanthine alkaloids.

Recently, *Methylorubrum populi* PINKEL, the only known methylotrophic bacterium that can use caffeine as sole carbon, nitrogen, and energy source was isolated via a caffeine enrichment strategy. In addition to caffeine, *M. populi* PINKEL can grow on other methylxanthines, including theobromine (3,7-dimethylxanthine), paraxanthine (1,7-dimethylxanthine), 7-methylxanthine, and xanthine. *In silico* comparison of the *Pseudomonas putida* CBB5 *ndm* genes and the genome of *M. populi* PINKEL helped identify genes that are predicted to encode enzymes of the *N*-

demethylation pathway (amino acid percent identities ranging from 41% to 65%). Unlike *P.* putida CBB5, the *ndm* genes in *M. populi* PINKEL are organized in two separate clusters: the first cluster is approximately 3 kb in length and contains the *ndmA* gene ($N_1$-demethylase; peg_1790); whereas the second cluster measures approximately 11.2 kb and contains the remaining four *ndm* genes: *ndmB* ($N_3$-demethylase; peg_1782), *ndmC* ($N_7$-demethylase; peg_1777), *ndmD* (reductase; peg_1781), and *ndmE* (peg_1776).

Whole cell biodegradation assays in combination with HPLC analysis using wild-type and pathway mutant strains helped identify not only the metabolites formed by *M. populi* PINKEL as it degraded caffeine, but it helped corroborate that PINKEL follows the *N*-demethylation pathway as well. Through this pathway, three sequential *N*-demethylation steps convert caffeine to xanthine. The first step is catalyzed by NdmA, which converts caffeine to a mixture of theobromine and paraxanthine. During the second step, NdmA and NdmB convert paraxanthine and theobromine to 7-methylxanthine, respectively. The last demethylation step is performed by NdmCDE, which convert 7-methylxanthine to xanthine.

To study the transcriptional regulatory mechanism of the *N*-demethylation pathway in *M. populi* PINKEL, four potential *ndm* promoter regions were cloned upstream of a promoterless yellow fluorescent protein gene (*yfp)* and wild-type expression patterns were measured as an output of YFP. The data indicates that only the promoter regions located upstream of *ndmA, ndmB,* and *ndmC* are responsible for the transcriptional regulation of all *ndm* genes. Decreasing fragment lengths of the

confirmed *ndm* promoter regions were cloned and mated into wild-type *M. populi* PINKEL to narrow down the location of relevant nucleotides necessary for regulation.

A mutant strain lacking the putative LysR-type transcriptional regulator, encoded by the *ndmR* gene, which is located in the vicinity of the *ndm* genes, was constructed and differential expression patterns were measured and compared to those of the wild-type strain. The inability of a Δ*ndmR* mutant to upregulate expression from any of the three identified *ndm* promoter regions in the presence or absence of methylxanthines indicates that the NdmR protein positively regulates the *N*-demethylation pathway in *M. populi* PINKEL. Additionally, it was determined that all the intermediates of the *N*-demethylation pathway that contain a methyl group can serve as inducers of the pathway, although at different levels of induction.

In summary, the present study helped determine that:

1. NdmA is responsible for the $N_1$ and $N_3$ demethylation of caffeine during the first step of the *N*-demethylation pathway in *M. populi* PINKEL.
2. There are three promoter regions that control the transcriptional regulation of the *N*-demethylation pathway in *M. populi* PINKEL.
3. NdmR positively regulates all three *ndm* promoter regions
4. Caffeine, theobromine, paraxanthine and 7-methylxanthine can induce expression from the three *ndm* promoter regions.

Future experiments will focus on clarifying whether NdmR exerts its regulation of the *N*-demethylation pathway by directly binding to the regulatory sequence within the *ndm* promoter regions or if the regulation is indirect by controlling expression of other regulators. This can be determined by performing electrophoretic mobility gel shift

assays with purified NdmR and *ndm* promoter regions. Although we were able to narrow down the length of the *ndmA* and *ndmB* promoter regions relevant for regulation, the exact sequence of the promoter as well as the sequence determining the binding site of NdmR need to be determined. This will require a combination of bioinformatic work and DNA footprinting. An area that remains unexplored is the chemotaxis response of caffeine-degrading bacteria towards caffeine and/or related methylxanthines. Identifying a potential methyl-accepting chemotaxis protein (MCP) involved in the chemotaxis response to methylxanthines will aid the efforts to engineer methylxanthine biosensors.

# Appendix A

## DNA and protein sequences

**DNA sequences:** All DNA sequences are listed in 5’ to 3’ directionality.

*ndmA* (peg_1790)

GTGAGCCCGATGCCTGAACCGCGTCCACTGATCAACGACTACACATTCCTTCGGCATT
TCTGGCACCCCGTCGCCACGCTCGACGAATTCGAGCGGGCGAACGCCTCCGGCATC
GGCCCGATGGGGGTGACCCTGCTGAACGAGAGGGTCGTGCTGGCGCGTCTCAACGG
CAAATTCGTCGCGATGAAGGACCAGTGCGTCCATCGGTACGCGAAGCTCTCCATGGG
CCGCGTCAACGAGAACGAGTTGCAGTGCCCCTATCACGGCTGGAAGTTCAACGCCGA
CGGCCAATGCACCCGCATGCCAGCCTGCCCCACTCAGAAGATTTCGGCGCGGGCTCG
CCTCGAGAAGTACGAGTGTGAACTCAAATACGATCTGATCTGGGTGCGCTTGGACAGC
TCGTGGGGCGTGACGGAGATCCCCTACTTCAGCGCCGCCGAAAAGCCGAACATGAAG
GTCGTCGTCCAGGAGCCCTACTGGTGGAACGCCAGCGCTCCGCGCCGCTGGGAGAA
CTTCACGGATTTCTCCCACTTCGCCTTCGTGCATCCAGGCACGCTGTTCGATCCCAAC
AATGCCGAGCCGCCGATTGTCCCGATGGATCGCGTCAACGGGCAATTCCGGTTCATC
TACGACACGCCGGCCGGCATGGACGTTCCGGACGAAGCCCCGATCGGCAAGTTCACC
TATCACTGCAGCATGCCGTTCGCCATCAACCTGGCCGTGATGAAGTACGTCGGAAACA
CCGAGCACATCCTTTTCAACGTTTCCTGCCCTGTGAACGAGAAGCAGACGAAAAACTT
CCTTCTTTTCGCCCGCGACAAGAGTCTCAACGATCCCGACTATCCTCACATCGCCTTC
AACGATCTTGTCTTTGCAGAAGACAAGCCCGTCATCGAATCGCAGTGGCCGGAGGAAA
TCTACGAAGATGAGCTCTCGGTCGTGACCGACAAAGTCTCGATGCAGTACCGCAAGTG
GTTGAAGGAGCTGGAAAAGGGCCATGCGGAAGGCAAGGACGCCTTTCGGAAGGCGTT
GCTCGGCAGCGTCATCGAAAGCGACCGTGTCTATCCGGACATGATGGAAGCCGCCGA
GTAG

*ndmB* (peg_1782)

ATGAACCTGATGCTCAATCCTCGTTCGCTCGTCGAGGATCGCTCCTACCTCAGGCACT
TCTGGCATCCCGTCTGCACCTTGGCGGAGCTTGCGGAAGCCGGACCGGAGGGAGCG
GGGCCGCTCGGCGTCACGCTCCTCGGCGAGCCGCTAGTCATCGTCAAGCTCGAGGG
CGAGATCCGGGCCATGCGGGATCGATGCGCCCATAGGTTTGCCAAGCTCTCAGCGGG
CAAGATCGTCGAGAACCGACTACAATGCCCCTACCACGGCTGGCAGTACAACGCAGA
TGGTCGCTGCGGCAGAATTCCGGCAGCACCGAACGATGCGATCCCGAAGAAGGCCGT
CGTACCCAGCTACGAGTGCGAAGTGAAGTACGACATCGTGTGGGTTCGGCTGGATTC
CTCGTGGGACGCGACCACAATTCCGTATTGCGGCGCTTGGGACAATCCAGAGTACAA
GCGGGTAATCGTAGCAAAGCCGTATAACTGGTCGAGTTCGGCAGAGAGACGGTGGGA
GAACTTCACCGACTTCTCGCATTTCGCCTACGTGCACCCGGGTACGCTCTACGATCCG
GCCTACTCCGAGCCGGCGATCGTCCCGATGGATCGAGTGGATGGAGAACTGCGCTTC
TTCATGGAGCCCGGCAAGGAGATGGTGGATGTCCTTCCGCCCGACAGTCCGCTCGGA
TCTTGGACCTACAGAGCGGCCATGCCGTTCACCATCAACCTGGACATCCGCCTGTACC
GGAACGACAAGCCATTTCTTCTTTGGACCACGTCGAGCCCGATCGACGGAGAAAACTG
CCGCAACTTCATGATCATTGCGCACACCGATCCTGACCTGCCCGACTCGGTGCCGCTT
GATTTCCAAAAAGTCGTTCTCGCGGAGGATCAGCCCGTCATCGAGACTCAGCCCGGA
GATCTGTCGCTCGAAGAAGTGTCTCTTCCGACCGACAAGGTTTCCAACCAATATCGGA
AATGGCTGCGTGAGCTGGCGATGGCTGCCAGCGAGGGACAGGAGAGCTTCCGGAAA
GTTCTGCAGACCAACATCATCGAGAGCGCCTGA

*ndmC* (peg_1777)

ATGCTCCTGGATCAGGACATTGAACTTTCGAACTCGAACGGCGTGATCGCGTCTCGGG
CGCGCACCAAGTCCGGTGCAGCGACGCCGCTTCCGACCCAAGTCCGCTACGGCCAC
ATCTTCACGACCTTGAGCGACAGCCCCCGCGACCTCTTCGCGATGCCTGAATTCGACG
AGCCCGGACGCCGCCTCGTGACTGCCGGTGCCATCGCCGTGTCGTGCTCGGGTCTG
CGCGTGGTCGAGAACTTCCTCGATATGGCCCACTTCCCCTTCGTCCACACGGGCATTC
TCGGCGAGGAGCCGCACACGGACGTGGCTGAGTACGACGTCGAGTTGCGCGAAGAC
GTCGACGAGGTGTGGGCGACGAAGTGCGAATTCTATCAGCCGAAAGCTGCCGCGGC
GGCCGAGGGGGGGCAGCTCACGAGGTACATGTATCGGGTCGCTCAACCCTTCAGCAC
CGTTCTCTACAAGACCTGCCCGATCCGTGAAGGAGCGTGGGACTTGGTCGGGCTGTT
CATCCAGCCCAAGACGGAGACCGAGTGCGATGTACATTCGTTCGTGCTGGTGTTCGA
CGACAGCAATTCCGACAACTCGCTGCTGCAGTTCCAGCAGATGATCTTCATGCAGGAC
CGCATGATTCTCGAGAATCAGGTTCCTGCCAAGTTGCCGCTCGATCCACGCCTCGAGC
TTCCAACCCGGGCCGACGCGAGCTCGATCGCCTTCCGGCGCTGGCTGAAGGCGCAC
GGACAGACCTACGGCGTCTCCCCTCTGGCTTGA

*ndmD* (peg_1781)

ATGAGCCAGCAATTCAAAGTTGCCAGAATTTGGGATGAGACCTCCGAAATCAGGTGTT
TCGAGCTAGTCCCTGAGGATGGCCATACGCTGCATGCGCCCGAACCTGGAGCGCACA
TCGACGTTCACATCGCGGACAAGCTGATCCGTCAGTATTCGCTTTGGAACGGTCCCGA
AGACACCCGTAGCTACTTCATCGGCGTGAAGAAGGAGGCGGAGTCCCGCGGAGGGT
CTGCAGGCATGCACGGTCTCGCGGCTGGCAATCTGCTGACGATCGGCAGCCCGAAGA
ACAACTTCCCGATGGATGCCGATGCCCGGAAAAGCATCCTGATCGCAGGCGGTATCG
GCATCACGCCCCTGCTGTCGATGGCTCGATACCTCGCAGCCGAAGGCCGATCGTACG
AACTGCACCTCTTCGCCAGAAGCGAGACGCACGCTCCGTTCAAGGACCTTCTCGGCA
GCCTCGGAAACGCCAGCTTCCATCTGGGTCTCCAGGCCGAAGCACGAGACGCGCTCG
TGGACGAAATACTGGCCGGTGGAAAGGCGGAAGGCGCGCACGTCTACACGTGCGGT
CCCAAGCCGTTCATGAACCTCATCGTGACCACTGCCGAGAGGCAGGGCTGGAATCCG
GACCGCGTGCATCTGGAGTACTTCTCGGCGGACCTGCCGGAGGCTTCGGCCGACGA
CACCGCGATCGAAGTGGTCCTTGCGAAAAGCGGTAGAGCCATCGTCGTCGCGCCGGA
TCAGACCATCACGGACGCGCTGATCGCGGAGGGCGTGGACATCCTGACTTCCTGCGA
ACAAGGCGTCTGCGGCACCTGCCTGACCACGGTGCTTGAGGGAGAGCCGGATCATCG
CGACGTCTATCTCAACCCGGCGGAGAGGAAGGCCAATAAGACCATGCTGCCCTGCGT
CTCTCGCTGCAGGGGCAAAAGATTGGTACTGGACCTATGA

*ndmE* (peg_1776)

ATGAAGCTTTACGACTACGTGCTGTCGAGCGAGTGCTACAAGGTCAGACTTCTCGGCA
GCTGCCTCGACCTCTCTTTCGAGACCGTTCCGGTGAATGTCTTTCCTGGAGAGGAGCA
CCGCTCCGACCCGTTCCTTGCAATCAGTCCTCAGGGAAGGCTTCCCGTCCTGAAGGA
CGACCGCATCCTCCTGCAGGAGGCGAACGCAATCCTCGTCTATCTGGCCAAGACCTAT
GACGGTGGGGCGTCTTGGTATCCGGATGCGGATCCAGCGACATCCGCGCTCATCCAG
CAGTGGCTGACCTTCTCCATCGAACTGGCCGGGACCGTCGGCGCCGCCCGCCTGCA
CAGCATGCTCGGTCAGCCCCTCGACATCGGCGTCGCGCGCTCGGGCGCGCGAGATC
ATCTCACGATCCTGGACGATCATCTCACGGATCGAGGCTTCGACGGGGCGACATGGC
TGGTCGGCGACCACGCCACCATCGCGGACGTCGCGGTATTTGCTCATGTCGCCGTCG
CAGGAGACGGAGGCATCGACCTGACGCCCTACAAGGCGGTCGGACGATGGCTGAAG
CAGTTCCTGGCACTGCGTGGCTTCGTCCCGATGCCGGGCATCAGCGCGGTGAAGTGA

*ndmR* (peg_1778)

TTGGCGATTTCACCTATGGACTGTATGCAAGTCCAGCCTACCTCTCCAGTGTTCCGAA
GACCCAAGATCTCAAAGCTGTGGAGCGTCTCTCCGGCATTATCTATATGCGAAACAAT
ACGCCTGAGAGATGGGTGCTCGATAAACAGAAGACGAGACAGATCTGCAGCCCGGAG
ACGAAATACACTGTCAACGATTACTGGATGTCAAAGTACTTCTGCGTCAAGGGCATGG
GAGTTTCGTACCTGCCTGACTTCTTTGTCAAATACGAGGTCGAGTCCAAGGCTCTGGT
GCCGATCCTGCCCGATTGGCGCTCGAGAGCGGTCCCCGTCTTCATCATCTATTCTCCA
CAGAAGCACCGCATCGCACGCATTGCGAAGCTGGTCGATATTCTCTGCGAAAATTTCG
AGGGGTTTACGATGCTCCCCGGTTACTCTCTCGTTTCACGGGAGGAGTGATGTCGAG
GCATCCAGGGGGACCGCGCTCTGCGCGACCGACGAAGACGGAGGTCGTGGCACCGA
AACGCCGCATCGTCGCCGCCGATCCGCCTGCTGCGGTCCGTCGGAGGTTCGCGGCA
CGAAGCATCGGGCAGGTCACATGAAGCTGTGA

*irr* (peg_1792)

ATGGGGCTCGGAATTCTGCTGTTCGGAAAGGAGCATCGCCATCTCTCCGCCGAAACTC
TGTTTTCCGAAGCACAGCGCGCAAAAGTGAATGTGTCCTTGGCGACCATCTACAATTG
CCTCCATGCTTTTACGGAAGCGGGACTACTCCGGACGGTCATCGTCGACGGACAGCA
GGTGATCTGGGACACCAATACGTCCCACCACCACCATTTTGTCTGCGATCAAGATGGC
TCGGTCATCGATGTTCCGGCCATGGTTATCGACCAGTCGGCGATCAATGGAGTGCCG
GACGACATGGATGTCGTCGATGTCCAGGTCATCATGCGGGTGCGACCGATCCTGCAG
AGGGGGCATCGGCTTCCCCAATCCAAGTAG

*celA* gene cluster: *celA*, *celB*, *celC*, and *celD*
(peg_1742, peg_1743, peg_1744, and peg_1745)

*celA*

ATGGGAACGACGGTCGCCGGCCTCGTTCTGCTGAGCCAACCGGTCGGGACCCAGAACCA
GCTCGCCATGAGTCTCGCCGCCATGGCGGCGATGATCGTGCTGTGGCTGTTCCTCGACG
GGCCGCGCACCCGCTTCGTGTTCCTGGCGCTCGGGAGCCTCGTGGTGCTCCGCTACATC
CTGTGGCGGGTCACCGACACGCTGCCCTCCCCCGGCGACCCGGTCAGCTTCGGCTTCGG
CCTGCTGCTGCTCGTGGGCGAGTTGTACTGCGTCTTCATCCTGTTCGTCAGCCTGATCATC
AACGCCGACCCGCTCAGGCGCCCCCCGCCCCCCGTGGCGAGCGCGGCCGAACTGCCCA
GCGTCGACGTGTTCGTGCCGACCTACAACGAGGACGCGGCGATCCTGGCGATGACGCTG
GCGGCCGCACGCCAGATCAACTACCCGCCCGACAAGCTCACCGTGTGGCTCCTCGACGA
CGGCGGCACCGACCAGAAATGCGCTGATCCCAACCCGAAGAAGGCCGAGGCCGCCCGC
GAGCGGCGGCGCGACCTGACGGCGCTCGCCGAAGCGCTGGGCTGCCGCTACCTCACCC
GCGCCCGCAACGAGCACGCCAAGGCGGGCAATCTCAACAACGGGCTCGCCTTCGCCACC
GGCGAGATCGTCGTCGTGCTCGACGCCGACCACGTCCCGTTCCGCTCCTTCCTCAGCGA
GACCGTCGGCTACTTCGCCGAGGATCCGAAGCTGTTCCTCGTCCAGACCCCGCACGCCTT
CCTCAACCCGGACCCGATCGAGCGGAATCTGCGCACCTTCGAGCGGATGCCGTCGGAGA
ACGAGATGTTCTACGCGGTGACGCAGCGCGGGCTCGACAAGTGGAACGGCTCGTTCTTCT
GCGGCTCCGCCGCCCTGCTGCGCCGCACCGCACTGGACGAGGCCGGCGGGTTCTCCGG
CATCACCATCACCGAGGATTGCGAGACCGCCTTCGAGCTGCATTCCCGCGGCTGGACCAG
CGCCTATGTCGACAAGCCCCTGATCGCCGGTCTCCAGCCCGAGACGCTCTCGGCCTTCAT
CGGCCAGCGCTCGCGCTGGTGCCAGGGCATGTTCCAGATCCTGCTCCTGAAGAACCCCG
TTTTGCAGAAGGGGCTCAAGCCGATCCAGAAGATCGCCTACCTCTCCAGCATGACGTTCT
GGTTCTTTCCCGTGCCGCGGCTGATCTTCATGTTCGCGCCGCTGCTGCACATCTTCTTCGA
CCTGAAAATCTTCGTCGCCAGCGTCGATGAATCGATTGCCTATACGGCGACCTATATCGTC
ATCAACCTGATGATGCAGAACTACGTCTACGGCAAGTTCCGCTGGCCGTTCGTCTCGGAA
CTCTACGAGTACGTCCAGGGCCTGTACCTGTCGAAGGCGATCGTCTCGGTGATCTGGTCG
CCGCGCAAGCCCACCTTCAACGTCACCGACAAGGGCATCAGCCTCGATCACGACCACCTC
TCGTCGGCCTCGCTCCCGTTCTTCGCCGTCTACGGGCTGCTCGCGGCGGGCTGCGCGGT
GGCGGCGTGGCGCTACCTGTTCGAGCCGGGCGTGACCAACCTGATGCTGGTGGTCGGCC
TGTGGAACTTCTTCAACCTGCTCACCGCCGGCGCCGCCCTCGGGGTCTGCGCCGAGCGG
CGGCAACTGGAGCGCACCCCCTCGCTCGCCATCAACCGGCGCGGCCAGATCACGCTGGC
CGGCCGGGCCATCGACGTGTCGATCGAGCGTGTCTCGGCGGAGGCCTGCACGGTGCGG
CTGCCCGCCGCCCTGCTGCCGACGGGGGCCGGGCACCGCGCGCTCACCGGCGCGCTCA
CCGTCGTGCCGGTGGCCGGCGCGCGCGCGGCCGGCGCCCTGCCGGTGACCCTGGAGG
GGATCGACCGCGCCAAGGACGAGGCCTTCGCCCGCCTGAGCTTCGGCCGCCTGCGCCC
GCAGGACTACGTGGCGCTGGCCGGCCTGATGTACGGCGACGCGGAGGCGATGCGCCGC
TTCCAGATGCGGCGGCGGCGCCACAAGGACATCCTGACCGGCACGCTGCAATTCATCTG
GTGGGGGCTCTCCGAGCCGTTCCGCGCCGTGCGCTACGCGCTCGCCGCCGATGCCCGG
CCGGCCGCGGCGCCCGTGCTCGACGGCACCTCGCCGATCTACGACGCGCCGGAACAGG
CCCCGACCGGGGCGAACCTGCCCGAGCCGCTGCTCCCCTCCGCCCCGCCTCGGCCCGC
GCCGCAGGCCGTGCCGGCGGCGCTCAGGGAAGCCGTCCCGGCCGAGGCGCAGGGCGC
CTCCCAGGCGAGCGTCCAAGCCGCTCCGCAAGTCGCACCACAAGTCGCGCCGCACGCCG
CGCCTCACGCCCCCGCCCGCGCCTCGAACGACTGGGTCGGCATGATGCTCGATTACGAG
AACGAGCGGGCGCTGGCCGGGCGCGCCGGCCGCGGCGCCGGCACGGCGTGA

*celB*

ATGACCCTGCGCCCTCCTTCACGCGCCGGCTTCCCCCGGCACCCGATCGTCCGGGCCCT
GGCCCTCGCGCTCGCCTGCACCGTCGCGGCGGGTCCGGCGGCCCGGGCCCAGAGCTTC
CTCGGGAGCGGTCCGGCCGAGCGCCTGACCGTGCCCCCGACGGGGGACGCCCTGCGTA
AGCAGAGCCCCGCCGCGCCGGCCCAACCCCCTGCGCAATCGGCCGCCGCGCCGGCCCA
GGGCCGCTCCCCGGGCGAGCCGGCCGACACCGTCCGCCGTCCCCAGATCCAGGCCACC
ATCGCCCCGGCCCGGCGCCTGCCCGCGGGCACCCGCGGCTTCCGCCTCTCGGGCGAGG
AGGACATGCTGCAATACCCGGTCTACCTCACCGAGGGGCAGGTGCGGGGCGCGGCGCG
GCTGCGCGTCTCCTACCTCTCGGCGATCTCGGTGGCGCCGGAGGGCTCGGAACTCACCG
GCCTCGTCAACGGCACCAAGGTCGGCTGGACCCGGATCCAGGCGCCCGGCGCCGTCAAG
GTCGTGGAATTCGCGATCCCCGACGGGGTGCTGAAGGCCGGCTACAACGCCGTCACCCT
CACCGCGAGCCAGCGCCACCGGGTCGATTGCGCGATCGAGGCGACCTACGAACTGTGGA
CGCAGATCGACCCCTCCCGCTCCGGGCTGGTGGTCGCACCGGCCACCGATCTCGACCTG
AAGGCCCTGGCCGCCCTGGAGCCCGACGAGAGCGGCGCGCTGCCCGTCGGCGTCCTGC
TCAACGAGAAGCCGACCCTGCCGCGGCTGGAGCGGATGATCGGCGCCGTGCAGGCGCTC
GCCGTGGTGAGCCGTCTCGCCCGGCCCGCGGTCAGCTTCGGGCCGCCGCTCACCGGCC
GCACCGGCGTCAACCTGCTCGTCGGCACCGCCGCCGAGATCCGCGGCACGGACGGCGC
GGAGGATCTCGGGCCGATCACCGGGCCGCGGCTCGCCCTGCTGCCGCCGCGGGGTAAC
CGCGCCCCGACCCTCGTCGTGACCGGCACCGGCCCCGACGACCTGCGCACGGCGACCG
CCCAGCTCGCCGCGGCGGGCGATGCCGGCGGCAGCGGGCCCGGCGCGGCGCTGGCCC
CGCTGATCCGGGGCTACGAGGTGCAGGGTGGCGAGCGGCTCGACCTGGAGCGCCTCGG
CGTCGCCACCCGCGAATTCAACGGGCGCCTGCTGCGCACCCGGATCGAGCTGCGCTTCC
CCGCCGACTTCGTCCCTGCCGACTACGGCAAGGTGATGCTGCATCTGGCCGGCGGCTAC
GCCGCAGGCCTCGATCCGAGCGCCCAAATCGTCGTCGACATCAACGGCCGCAACGCCGC
GAGCGTGCCGCTGCCCTATGCCCGCGGCGAGGTATTCGACGATCAGGCCATCCCGCTGC
CCCTGAGCCTGTGGCGGCCAGGCCTGAACAGCGTCGAGATCAGCGCGCAGCTGCCCAAC
GCGTCGGACAAGGTCTGCGACACCCTCTCCCCGGCCGCACGGCAGCCGCGCTTCCTCTT
CCTCGACCGCACCCGCCTCGAAATCCCGGCGCTCGCGCGCGCCGTGCGCAGCCCCGATC
TCGCGGCGGTCAAGGCGGGAGCGGTGCCGTTCACCGCGGCGGGTCCGCGCCCGCGCCT
CGTTCTGCCGGTCACCGACCGCGAGACGGCGGACGCCGCCGCGACATTGGTCGCCCGC
CTCGCCATCGCCGCCAAACGGGTGATCGATTTCGAACTCGTGCCCGATGCCGGGGGAAC
GGAGGGCGGTGCCAACCTCGTGGTGGCGCCCGCCCGCGCCCTCAACCCGGTGACGATC
CAGGGGGCCGGGCTCGATCCCGGCGAGATCCGCCGCATCTGGGAGGGGCGGGCCGAGA
CCGTGGCCACGCCCGGCCAGTTCGGCACCGAGGGTGTGCTGACCCTCGACCGGCTGCG
CCGCAACCTGCCGATGCGCTGCGCCCTGCCGCCCGCCGCAGTGCCGGTCAGCGTCGCC
GCGCCGGCCTCCGCACCAGTGACCGCTCCGGCCGCGACGCCGGCGCCTGCCGCGGCCT
CGGCGGCGACCGCAACACCGCGGGGGGTTCCCCGGCCGGGCCGCCCGGAGACGGACG
AGGAACTCGTGGCTCGCTGGGACCAGACCCTGTCGGGGGCCGCCTTCGTGCCCACGGTC
CTGCGCGAGGGCTGGACCCGGCTCGCGAACGCGGCCCTGCGGCTGCGCAGCGAGGCGA
CGGGCCTGATGGAGGCGCCGAAGGCCGGGGCCGTTCCGGTCAATCCCCGCGCCTCGCT
GATCATCGCCGAGGACAGCAGCGGAAACGGGCTGGACGAGACCACCGTGCTCGTGACCG
CGCCGAACCCGTCGATGCTCAAGGCCTCCGTCTCCTGCCTCGTCGATCCGGTGGTCTGGA
CCCGCCTCGTCGGGCGGGCCGCCTTCCTCGACGCCTCGGACGGCACCCTCACCACCGTC
GAGCCGGACCACGTCGCCTTGATCGAGACCCGCACGTGGTCGCTCGCGAATTTCCGGCT
GGTCTCGGCAGCGTGGCTCTCCCTCAATCCCGGCTACTACGTCGGCATGACCCTGTTCAT

GGCCCTGTGCCTCGGCCTCGCCACCACGGGCCTCGTGCGCCAGCTCGGCCGGAGGAATT
CCTGA

*celC*

ATGCGGCCCCCCCCGCACGCCGTGCCGTTCCCGCCTCGGCGGCACCCTGCGCCGCGCCG
GATGGGCGCTCGCCCTTGCGGCGGCCAGCCCCGCCGCCGCCCAGCCGACGCCGCAGCC
GGCCGCGCCGCAGCCCGCATCTCCGCAGCAGACCTCGCCGCAGCCGGCGCAGCCGGCT
CCCCAGCAGACGGAGACCAGCACCGTGCCGCCCGCCACGGCCCTGCCCGGCCTGCCGC
GCTCCGAGACGCCGACCCGGACCGATTCCGGCCCTTCCCTCGCCAATACCCTGGACGAC
GACGCGGCCTGGCGCGCCTACCGGGCACGGTTCATCACCGAGCAGGGCCGGATCGTCG
ACACCGCCAACGGCATGATCAGCCACAGCGAGGGCCAGGGCTACGGCATGCTGATCGCG
GTCGCGGCCGGCGACCGGGCCGCCTTCGAGCGGATCTGGGGCTGGACCCGTGCCAACC
TGATGGTGCGCTCCGACGAATTGCTGGCGTGGCGCTGGGCGCCCGACCACCGCCCGGC
GGTGGCGGACATGAACAACGCCACCGACGGCGACATCCTCGTCGCCTGGGCGCTCACCG
AGGCGGCGGAGGCCTGGGGCGAGCCATCCTACCGCACCGCCGCGCGCCGCATCGCCGT
CGAGTTCGGTCGCAAGACCATCCTGTTCAAGGACCCGCACGGACCGCTCCTGCTGCCCG
CCGTCTCCGGCTTCTCGGCCCGCGAACGCGCCGACGGGCCGCTCATCAACCTGTCCTAC
TGGGTCTTCCCCGCCTTCCAGCGGCTGGCGATCGTCGCCCCGGAATACGACTGGGCCTC
GCTGATCCGCAGCGGCCTCGACTTCCTGCGCCAGTCCCGGTTCGGGCCGAGCAGCCTGC
CCACCGAGTGGATCTCGGCCAAGGACAGCCTGCGCCCCGCCGACGGCTTCCCGCCGCTG
TTCTCCTACAACGCGATCCGCGTGCCGCTCTACCTCGCCTGGGCCGGGGTCGGGCGGCC
GGAGGATTACACGCCGTTCAAGACGCTGTGGGGCGGGATCGAGCGCGAGCGCCTGCCGA
TCGTCGACACCCGCGACGGGCAGCCGGTCGAGTGGCTGAGCGAGGTCGGCTACACGGC
GATCCCCGCGCTCACCGCCTGCGCCGCCGACGGCACGCCGTTTCCCGAACCGTTAAGAA
CGGTGCAGGAGAATCAGAACTACTATCCGATGACCCTCCACCTCCTCGCGCTGATCGCGG
CCCGGATGCGGTATCCGTCATGCGTGAAGTCCTGA

*celD*

ATGCGTGAAGTCCTGACGCCCCTGGCCGGCGGCCTGACCGCGCTCCTGCTCGCGAGCGC
GGCGCCCGCCCTCGCCCAGCCCGCCCTCGCCCAGACCGGCCCCCAGGCGGCGCCCGAG
GCCGCCCCGCGCATCGTCTACGGCGACGGCGCCCGCGCCCCCGCGATCGTTGTCGATCC
GGAGACGGGCGGCACGCCCGGCATGGTCGACGAGAGCGCGCTGCGCTACTACGCCTCG
CAGAAGCAGACCGCGCGGATGAAGGCCGAGATCGCCCGGTTGAAGCGGCTCTATCCGGG
CTGGACCGAGCCGGCCGACCTCGACACCCTGCAGCCGAGCCCGCCCGAGGAGGCGCCG
CTGTGGGATCTGTTCACCGCCGGCCGCTTCCAGGACCTGCGCGCCGCCATCGCCACCCG
CCGCTCCGTCGAGCCGAACTGGCAGCCCTCGGAGGAGCTGGCGCGCAAGCTCCGTCGG
GCGGAGTTCCGCAACGGCATCAAGGCGACCGCGGAGAGGAACGGACCGGGCAAGAACG
GTTCCGGCAAAGACGGACTTGGCAGGGACGGACTTGGCAGAAACGCCGAGGAGGTGGTC
GCGCTCTACCGCGCCGATCCCAGCGCCCTCGACACGACCGATCTCGAGAGCCTCTGGAT
CATCGCCGATGCGCTGGCCGTCACGGGCGCGGCCGAGGACGCCTTCGACCTCTACAAGT
CCGTCCTCGACGGCAGCGCCGATGCGGGCGCGCGCCTCGCCACGATCCAGAAGGCGAT
GAGCCATCTCAAGATGGATCAGGCCGAGCGGCTCATTGCCATGGGCCGATCGAATCCGGA
GGGCGCCTCGGAATTCGACGCGATCCGCCTCGACATCACCCGCGCCCGCATCGCCGCCT
TCCTGCACGACGAGCCGGCCCAGACGCCGACCCTGGCCGACCTCACCGCCTTCCAGGTC
TATGCCCGCAAGAGCGGCGACCCGAGCCAGACCGGGCTGCTCGGCTGGTACGCCTACAA

GCGGCGCCAGTTCCGCGAGGCGCTGGAATGGTTCAAGCTCGCCATCGGCCGCGGCGGC
GACGCCATGGTCGCTCACGGCCTCGCCCATACCCTGCGCGAGCTGAACATGCTGCGCGA
GGCGGAAGAGGTCGCCTACGCCTGGCGCGACCGCTTCGTCGGCAACGAGCTGCTGTTCA
TCGACGTGCTCGAACGCAAGCTGACGCTGGCCAACCCGCCCTTCATCGAGCCCGCCCGG
ATCGAGCGCTACGCCAAGGTCGTGATCGCCTCCGCCTCCGGCGAGGGCGCGCAGGGGCT
GGCATGGTACGCCTACAATTCCTGCCAGTTCGACACCGCGCTCGAATGGTTCCAGCGCGC
CGTCGCCTGGATGCCGAGCGAGACGACCGTGTTCGGCTACGCCCTCACGCTCCAGCGGC
TGAAGCGCCAGCGCCCCTATCTCGAGACCGTGAACCGCTACGACGGCTTGTTCCCGAAGG
TCGTCGACATGCTGTTCCGCGAGGATCCCGGCGGCCCGCCACTGCCCTGCGCGGCGCCC
CAGTCGCCCCCGGCCCGGCCGCCCGCTCAGGCGAAACTCGAACCGAAACCCGAGCCGAA
GCCCGACGTCCCCTCCTACGGCCGCGTGCCGCGGCCGGACGCGGCCACGATCGGCGCC
CAGGCCGGCACGGTGCCGGCGGCGGTCCAGCGCGGCGAGTTCCCGCTGAGCGTGCCCC
TCGACAACGCCTTGCGCTACCCGCCGTCCTCGCTGCAGCGGGCGCTGTCCGAGCCCGCC
GCCCCCGGCAGCTATGCCCGCGAGCCGGTCACCGCGCGTCCGCCCCTGGTGGCGCGGC
GGGTGAACGGGGCCGGCCCGATGCCCTACGAGCGCTACGGCTTCGCCCTGCTGCCGGG
CTATAACGGGCTCGACAAGCCGGAAGGCCTCTCCCCCGCCCCGCCCGGCACCCTCTGGC
AGGATCAGCAGGCCGAGCGCGGCCGGGCCGCCACGAGCCCGGAGAGCGACCGCTTCTC
CGTGCCCTCCCGCCTCTCCTCCGGTGCGGGACCGAACGGGAAAGGGTTTCCGTGA

Wild-type *ndmA* promoter region (658 bp)

GTTGATCAGTGGACGCGGTTCAGGCATCGGGCTCACTCCCTGTTTCCTACCGGATCC
GGAGGAGCGTCACTCGAGCGCTCAGCTTCCGGATCAGCGCCGCAATTCTTGATCGCG
GTTACAGGGAATTTTGATTTGGCTTAGTTCCAGTTTGAAAGGTCGAAAATTTAGACACG
CCGTTCTGTCTGATGGACAGCCGACGCCTCGCTGCCCATTTAGGCAGCGCACCATGT
GCGCTGCGGTATGACCGCAGCTTCGACGATTCCGCACTGCATAAAAGACGCGCAAGC
TGCCGAATGGAAGGTCATGGATACGCATGTGCGGTGGCCGGCCGGCGTACTAGGCAA
ACGCCCGCGCTCCTGCTAGCTGGGTGGGTCGCGCTCGTGTCGGACCGATCGTCTTCG
GCAAGCCTCGCTGGATTTGCCAGTCCTTCGATGGACACTCGGAGTTTCCGGCACCGC
AGCCTCGCCTCCGCATCGGCAGCGGAATTGTGAGTGCAGTCGCGCGGTCTCGGAGAA
TGGCCATCCGATGCACGTCCGGCTCCTATGGCGGCGACCTTTCGTGAAATTTTGAAAG
AGCTTCCAATCCCTTCGCATGACACTACGGTGGCGCTCCCCTACTGAATCATGCCTCG
TGTTCGAAGATCGAGAGACGTCGAACG

Wild-type *ndmB* promoter region (256 bp)

CTTCGTGCTCGCTTGGTCAGCCTGGCGACGCACCCATCAAGCTACCGCTCGACAGGC
CCATCTCAGAGCGAAAATGCAACTGTAATGCTAGCCCTCGGGCGCACGCGCTGTTGAT
GCCGAGCGTCCAAAATTCTGGACACCGCATCTCTTGCAGACGCCTACCCGCGCACCA
GCGTCGAGTCTAGCGTCATTCCAAACACTCACATGGGAACGGAAGACGCCACATGAAC
CTGATGCTCAATCCTCGTTCGCTCGTC

Wild-type *ndmC* promoter region (1844 bp)

GCACCCATCTCTCAGGCGTATTGTTTCGCATATAGATAATGCCGGAGAGACGCTCCAC
AGCTTTGAGATCTTGGGTCTTCGGAACACTGGAGAGGTAGGCTGGACTTGCATACAGT
CCATAGGTGAAATCGCCAAGTTTTCTGCGCATCAGAGGCGAGTCCGGCGACTCGCCA
ACATAGATCAAACAATCAAAATCGACCTGACCCGTGATGATTTTTTCAACCGAGCAAAG
TTGGATGTCGAAGTGCACATCGGGATAAGATCGGGAAAAATGTAGGGCGGCGGCTCC
GATGACCGAAGTGCCAAACTCCGATGTCGCGGCCAGCCTGATCGTTCCCCCGACTGC
CGTGTGGGCCCACTGCGCGGCGTTTGCCGCGCTATCGCAGCTTTCGAAGATCGCAGA
GCAGAGTTCGAGATACTTCTTGCCCGCCTCGGTCAGCTCCAGCCCTTTCGCCTTCCGA
ATGAACAGCTGAACCTGGAGGGCGTCTTCCAATTTGCGTAAATGATGACTTAGGCTTG
CCTTCGGAATATTATATATGTTCGAGGCCCTGGAGAGGCTCCCGGCACGAGCGACTTG
GACGAAACAACTGATGTCCTTGAGCTCGTATTCGTAGTTCATGACCGTTCCAAATTGGA
GACACCCAGTTCCGTCCTGCTGCCTAGCCGCGGGGTACCGTGCGGCATACCCTCAAA
ATCTGCGTCATCGCGGCGATAAGCGTGCGCTGTTGCGTACTTATATGCTTAGAACTGG
AATAGTAAAAGATGACAACGAAAAACGTGGGAGGATGGTCACCGGTAGCAATGAGCGA
CGCCGTCGCGCCCGGTGCCGTGTTGCGTGTGGTCGTCGAAGGTAGAGACATCGTAGT
CTGGCGCGGGAACGACGGTGCGGCGCGCGCGTGGGAGAACAGATGCCCCCACCGC
GGGATGCGCCTGTCCTACGGGTTCGTGCGCGGAAACCGTCTCACCTGCCTTTATCAC
GGCTGGGGCTACGACGGGCAGGGCTCCTGCATCTCGATCCCCGCCCATCCGGCGCT
CACGCCTCCGAAGACGATCAAGGTCGCCCGTTACGACTGCGCGGAAGCTGCAGGACT
GCTCTGGGTGGCCGCCCAGGGCACCAACGATCCCCTGCGGGATCCGGGCGGATCCT
GGCTGGCGGTGCAGTCGATTGATATCGCCCTTCCCCTGGACGAAGTTCGATCGGCCG
TCTCGCAGGCCGAACCTCCTTCCGATGAGGACGATGGGTTTCCGCGTCCTGGAGGCG
GTGACGCGGCTCCTCCGGCGTTCCTGCAGGAGACGCTGGTGAATGGCGCGGAGTCC
GCCGGGTCGCTTCTCATGTTCGACGACGTCAGGTTTGGCCGCAAGGTGCTGTGCGCG
CTGCAGGCCGTCGGCGACGATCAAACCGCGCTACACCTCCTCGTCTCTGCAGAAGGA
GATGGTGCGGACGGCGCCGTCCGACGCGCCTACGCACGATGGGGGCGACGGTTCAG
ACGTGGTGCGGGAGCCACGGAGCGCTCTTCTCACCGGGTCGTCTCGGCCACCGCCG
CGGAAAGGCATGCCAGACCGATGGCGTGACGACGCCGGGCAGCGCCGCGGCGGAC
CGTGACCGGGCCGATTGCAACTCTGAAAGCCAGTGGAATCATAGATGTGCGATGATCA
ATTCACCCGCTCTCAATGGTACGCCGTAGCGCGGATTGAGGATCTTCCGGCGGGTGG
GCGCCGGTCCACCATGCTCCTGGATCAGGACATTGAACTTTCGAACTCGAACGGCGT
GATCGCGTCTCGGGCGCGCACCAAGTCCGGTGCAGCGACGCCGCTTCCGACCCAAG
TCCGCTACGGCCACATCT

Wild-type *ndmD* promoter region (1312 bp)

CAACATCATCGAGAGCGCCTGAGGAGAATCGACTTGGACGCGCTCAGGGACGGCGTG
GTTCTGGATGCATGGTATCCGGTCGGGTCGCCGGTCGACATTCGTCGGCACGGCATC
AGCCACACGAAATTGCTGGGGCGTCAGATCGATCTGGCGGTCCACAAGACCTGGATC
TCCCCCAGCATCGATGGGCGGCCGCTTCCGGTAACCGAGCGGCTCGGGCTCGTTTG
GACCACGCTTGGACACCCGAACGGCCCGCCCCAGCCCTTGATCGAGTACTATGAGGT
CGATCGCACAACGCTAAACGTGTCATCGGGACCGATCGAGGCCACGGGCCCGGAGAT
ATTGGAGGACACGCTGCGGATCCTCGACGGATCTGCAGCCGAGGCCGAGCGGTCGC
CGGACCTGAAGGCTGCGTTCGCGCCTGAACGCGACGAGGAGGGATGCCTGCTGGTG
CGGAGCAGCCGGGCGGACCTCGGCGAGTTGGGATCGGACGTCGCCTGCGAAGCACG
TATACTCAGTCCGCAATCGGCGATCCTCAGTTTTCGTCATCCTTCCAGACCGCCGGGT
CAGGACCATCGGAGGGACTTCTTTGCGGTCTTCTGCCAGCCAATCGACCAAAGCGTCA
CGGTGGTCCACGCTTTCGCAGGCATCGTTGACCTATCGACCAGCAGCTTAGACCGAAT
TCGGCGAGGTCGTCCCTTCGTCGGGATCGAAAGCGCGGGCGTGTGCGGAGGGCGAG
CCAGCAAGCCTCCGCCACGGCTCTCGGATCGCGGTTGATGTCTCGGCCCCGGCATCG
CCCGGGCCTCCTGAGGCGGAGTGCACGGCGACCGACCGAGATGGGGGAGGGGAGC
GAGTGCGGTTGGGGAGGGCGCTTGATGATCATCCCGATCTGCAGCCGTGACGGACGT
CAGACGAGACCCTCGAACCTACGCCTGGTGCACGACGGCGCAGCGAGTGGGAAGAT
ATCGGCAGGAAACCGGCGACAAAAAACTTCGAGGCGAGGAGCATGAGCCAGCAATTC
AAAGTTGCCAGAATTTGGGATGAGACCTCCGAAATCAGGTGTTTCGAGCTAGTCCCTG
AGGATGGCCATACGCTGCATGCGCCCGAACCTGGAGCGCACATCGACGTTCACATCG
CGGACAAGCTGATCCGTCAGTATTCGCTTTGGAACGGTCCCGAAGACACCCGTAGCTA
CTTCATCGGCGTGAAGAAGGAGGCGGAGTCCCGCGGAGGGTCTGCAGGCATGCACG
GTCTCGCGGCTGGCAATCTGCTGACGATCGGCAGCCCGAAGAACAACTTCCCGATGG
ATGCC

**Protein Sequences**

NdmA

MSPMPEPRPLINDYTFLRHFWHPVATLDEFERANASGIGPMGVTLLNERVVLARLNGKFVA
MKDQCVHRYAKLSMGRVNENELQCPYHGWKFNADGQCTRMPACPTQKISARARLEKYE
CELKYDLIWVRLDSSWGVTEIPYFSAAEKPNMKVVVQEPYWWNASAPRRWENFTDFSHF
AFVHPGTLFDPNNAEPPIVPMDRVNGQFRFIYDTPAGMDVPDEAPIGKFTYHCSMPFAINL
AVMKYVGNTEHILFNVSCPVNEKQTKNFLLFARDKSLNDPDYPHIAFNDLVFAEDKPVIESQ
WPEEIYEDELSVVTDKVSMQYRKWLKELEKGHAEGKDAFRKALLGSVIESDRVYPDMMEA
AE

NdmB

MNLMLNPRSLVEDRSYLRHFWHPVCTLAELAEAGPEGAGPLGVTLLGEPLVIVKLEGEIRA
MRDRCAHRFAKLSAGKIVENRLQCPYHGWQYNADGRCGRIPAAPNDAIPKKAVVPSYECE
VKYDIVWVRLDSSWDATTIPYCGAWDNPEYKRVIVAKPYNWSSSAERRWENFTDFSHFAY
VHPGTLYDPAYSEPAIVPMDRVDGELRFFMEPGKEMVDVLPPDSPLGSWTYRAAMPFTIN
LDIRLYRNDKPFLLWTTSSPIDGENCRNFMIIAHTDPDLPDSVPLDFQKVVLAEDQPVIETQP
GDLSLEEVSLPTDKVSNQYRKWLRELAMAASEGQESFRKVLQTNIIESA

NdmC

MLLDQDIELSNSNGVIASRARTKSGAATPLPTQVRYGHIFTTLSDSPRDLFAMPEFDEPGR
RLVTAGAIAVSCSGLRVVENFLDMAHFPFVHTGILGEEPHTDVAEYDVELREDVDEVWATK
CEFYQPKAAAAAEGGQLTRYMYRVAQPFSTVLYKTCPIREGAWDLVGLFIQPKTETECDV
HSFVLVFDDSNSDNSLLQFQQMIFMQDRMILENQVPAKLPLDPRLELPTRADASSIAFRRW
LKAHGQTYGVSPLA

NdmD

MSQQFKVARIWDETSEIRCFELVPEDGHTLHAPEPGAHIDVHIADKLIRQYSLWNGPEDTR
SYFIGVKKEAESRGGSAGMHGLAAGNLLTIGSPKNNFPMDADARKSILIAGGIGITPLLSMA
RYLAAEGRSYELHLFARSETHAPFKDLLGSLGNASFHLGLQAEARDALVDEILAGGKAEGA
HVYTCGPKPFMNLIVTTAERQGWNPDRVHLEYFSADLPEASADDTAIEVVLAKSGRAIVVA
PDQTITDALIAEGVDILTSCEQGVCGTCLTTVLEGEPDHRDVYLNPAERKANKTMLPCVSR
CRGKRLVLDL

NdmE

MKLYDYVLSSECYKVRLLGSCLDLSFETVPVNVFPGEEHRSDPFLAISPQGRLPVLKDDRIL
LQEANAILVYLAKTYDGGASWYPDADPATSALIQQWLTFSIELAGTVGAARLHSMLGQPLDI
GVARSGARDHLTILDDHLTDRGFDGATWLVGDHATIADVAVFAHVAVAGDGGIDLTPYKAV
GRWLKQFLALRGFVPMPGISAVK

NdmR

LAISPMDCMQVQPTSPVFRRPKISKLWSVSPALSICETIRLRDGCSINRRRDRSAARRRNTL
STITGCQSTSASRAWEFRTCLTSLSNTRSSPRLWCRSCPIGARERSPSSSSILHRSTASHA
LRSWSIFSAKISRGLRCSPVTLSFHGRSDVEASRGTALCATDEDGGRGTETPHRRRRSAC
CGPSEVRGTKHRAGHMKL

Irr

MGLGILLFGKEHRHLSAETLFSEAQRAKVNVSLATIYNCLHAFTEAGLLRTVIVDGQQVIWD
TNTSHHHHFVCDQDGSVIDVPAMVIDQSAINGVPDDMDVVDVQVIMRVRPILQRGHRLPQ
SK

CelABCD proteins

CelA

MGTTVAGLVLLSQPVGTQNQLAMSLAAMAAMIVLWLFLDGPRTRFVFLALGSLVVLRYILW
RVTDTLPSPGDPVSFGFGLLLLVGELYCVFILFVSLIINADPLRRPPPPVASAAELPSVDVFV
PTYNEDAAILAMTLAAARQINYPPDKLTVWLLDDGGTDQKCADPNPKKAEAARERRRDLT
ALAEALGCRYLTRARNEHAKAGNLNNGLAFATGEIVVVLDADHVPFRSFLSETVGYFAEDP
KLFLVQTPHAFLNPDPIERNLRTFERMPSENEMFYAVTQRGLDKWNGSFFCGSAALLRRT
ALDEAGGFSGITITEDCETAFELHSRGWTSAYVDKPLIAGLQPETLSAFIGQRSRWCQGMF
QILLLKNPVLQKGLKPIQKIAYLSSMTFWFFPVPRLIFMFAPLLHIFFDLKIFVASVDESIAYTA
TYIVINLMMQNYVYGKFRWPFVSELYEYVQGLYLSKAIVSVIWSPRKPTFNVTDKGISLDHD
HLSSASLPFFAVYGLLAAGCAVAAWRYLFEPGVTNLMLVVGLWNFFNLLTAGAALGVCAE
RRQLERTPSLAINRRGQITLAGRAIDVSIERVSAEACTVRLPAALLPTGAGHRALTGALTVV
PVAGARAAGALPVTLEGIDRAKDEAFARLSFGRLRPQDYVALAGLMYGDAEAMRRFQMR
RRRHKDILTGTLQFIWWGLSEPFRAVRYALAADARPAAAPVLDGTSPIYDAPEQAPTGANL
PEPLLPSAPPRPAPQAVPAALREAVPAEAQGASQASVQAAPQVAPQVAPHAAPHAPARA
SNDWVGMMLDYENERALAGRAGRGAGTA

CelB

MTLRPPSRAGFPRHPIVRALALALACTVAAGPAARAQSFLGSGPAERLTVPPTGDALRKQS
PAAPAQPPAQSAAAPAQGRSPGEPADTVRRPQIQATIAPARRLPAGTRGFRLSGEEDMLQ
YPVYLTEGQVRGAARLRVSYLSAISVAPEGSELTGLVNGTKVGWTRIQAPGAVKVVEFAIP
DGVLKAGYNAVTLTASQRHRVDCAIEATYELWTQIDPSRSGLVVAPATDLDLKALAALEPD
ESGALPVGVLLNEKPTLPRLERMIGAVQALAVVSRLARPAVSFGPPLTGRTGVNLLVGTAA
EIRGTDGAEDLGPITGPRLALLPPRGNRAPTLVVTGTGPDDLRTATAQLAAAGDAGGSGP
GAALAPLIRGYEVQGGERLDLERLGVATREFNGRLLRTRIELRFPADFVPADYGKVMLHLA
GGYAAGLDPSAQIVVDINGRNAASVPLPYARGEVFDDQAIPLPLSLWRPGLNSVEISAQLP
NASDKVCDTLSPAARQPRFLFLDRTRLEIPALARAVRSPDLAAVKAGAVPFTAAGPRPRLV
LPVTDRETADAAATLVARLAIAAKRVIDFELVPDAGGTEGGANLVVAPARALNPVTIQGAGL
DPGEIRRIWEGRAETVATPGQFGTEGVLTLDRLRRNLPMRCALPPAAVPVSVAAPASAPV
TAPAATPAPAAASAATATPRGVPRPGRPETDEELVARWDQTLSGAAFVPTVLREGWTRLA
NAALRLRSEATGLMEAPKAGAVPVNPRASLIIAEDSSGNGLDETTVLVTAPNPSMLKASVS
CLVDPVVWTRLVGRAAFLDASDGTLTTVEPDHVALIETRTWSLANFRLVSAAWLSLNPGYY
VGMTLFMALCLGLATTGLVRQLGRRNS

CelC

MRPPRTPCRSRLGGTLRRAGWALALAAASPAAAQPTPQPAAPQPASPQQTSPQPAQPAP
QQTETSTVPPATALPGLPRSETPTRTDSGPSLANTLDDDAAWRAYRARFITEQGRIVDTAN
GMISHSEGQGYGMLIAVAAGDRAAFERIWGWTRANLMVRSDELLAWRWAPDHRPAVAD
MNNATDGDILVAWALTEAAEAWGEPSYRTAARRIAVEFGRKTILFKDPHGPLLLPAVSGFS
ARERADGPLINLSYWVFPAFQRLAIVAPEYDWASLIRSGLDFLRQSRFGPSSLPTEWISAK
DSLRPADGFPPLFSYNAIRVPLYLAWAGVGRPEDYTPFKTLWGGIERERLPIVDTRDGQPV
EWLSEVGYTAIPALTACAADGTPFPEPLRTVQENQNYYPMTLHLLALIAARMRYPSCVKS

CelD

MREVLTPLAGGLTALLLASAAPALAQPALAQTGPQAAPEAAPRIVYGDGARAPAIVVDPET
GGTPGMVDESALRYYASQKQTARMKAEIARLKRLYPGWTEPADLDTLQPSPPEEAPLWD
LFTAGRFQDLRAAIATRRSVEPNWQPSEELARKLRRAEFRNGIKATAERNGPGKNGSGKD
GLGRDGLGRNAEEVVALYRADPSALDTTDLESLWIIADALAVTGAAEDAFDLYKSVLDGSA
DAGARLATIQKAMSHLKMDQAERLIAMGRSNPEGASEFDAIRLDITRARIAAFLHDEPAQTP
TLADLTAFQVYARKSGDPSQTGLLGWYAYKRRQFREALEWFKLAIGRGGDAMVAHGLAH
TLRELNMLREAEEVAYAWRDRFVGNELLFIDVLERKLTLANPPFIEPARIERYAKVVIASASG
EGAQGLAWYAYNSCQFDTALEWFQRAVAWMPSETTVFGYALTLQRLKRQRPYLETVNRY
DGLFPKVVDMLFREDPGGPPLPCAAPQSPPARPPAQAKLEPKPEPKPDVPSYGRVPRPD
AATIGAQAGTVPAAVQRGEFPLSVPLDNALRYPPSSLQRALSEPAAPGSYAREPVTARPPL
VARRVNGAGPMPYERYGFALLPGYNGLDKPEGLSPAPPGTLWQDQQAERGRAATSPES
DRFSVPSRLSSGAGPNGKGFP