

**Molecular and Bioinformatic Characterization of the Megaplasmiids of
Pseudomonas maltophilia, strain DI-6, and their Role in Degradation of the
Herbicide dicamba**

by

Matthew C. Hibberd

A Senior Honors and Distinction Thesis submitted to the faculty of

The University of Nebraska – Lincoln

in partial fulfillment of the requirements for the degree of

Bachelor of Science

Department of Biochemistry

The University of Nebraska – Lincoln

Submitted March 27, 2007

**Molecular and Bioinformatic Characterization of the Megaplasmiids of
Pseudomonas maltophilia, strain DI-6, and their Role in Degradation of the
Herbicide dicamba**

Matthew C. Hibberd

Adviser: Dr. Donald P. Weeks

ABSTRACT:

Biological degradation mechanisms are important bases for developing efficient and effective genetically-modified crop systems. Understanding of degradative machinery in environmental isolates capable of degrading xenobiotic compounds aids efforts to manage and control environmental contamination. One such notable microorganism is *Pseudomonas maltophilia*, strain DI-6, which is capable of exploiting the broadleaf herbicide dicamba as a sole carbon source. It is known that all genes necessary for the metabolism of dicamba are located on this organism's high molecular weight megaplasmiids.

Megaplasmiid DNA was isolated from *P. maltophilia*, strain DI-6 grown on dicamba by a commercial alkaline lysis preparative method. Isolated DNA was analyzed by contour-clamped, homogeneous field (CHEF) electrophoresis and consisted of 8 distinctly-sized plasmids. A high molecular weight DNA size marker and a collection of known-size BACs were run on further gels and compared to the *P. maltophilia* megaplasmiids to facilitate size determination. Statistical regression analysis revealed megaplasmiids of sizes ranging from 26 kilobases to 181 kilobases.

Total megaplasmiid DNA from this organism was also sequenced and subjected to initial bioinformatics analysis. Contigs from the sequencing operation were imported into the VectorNTI software suite (Invitrogen, Carlsbad, CA) and potential open reading frames were predicted. ORFs were translated and compared by BLAST to protein sequences of known function. Positively-identified genes (high BLAST scores) were listed and qualified by elementary statistical analysis, while unknown genes were screened for conserved domains or folding patterns in efforts to predict possible function. Multiple sequence alignment was performed against dicamba monooxygenase, the first enzyme in the dicamba degradation pathway, in attempts to determine gene copy number.

Individual megaplasmiids from *P. maltophilia* were "tagged" with an inducible origin of replication/antibiotic resistance marker and subsequently transformed into competent *E. coli*. Positive transformants of DI-6 megaplasmiids were obtained, but further analysis is needed to determine the nature of the plasmids re-isolated from these transformants.

This work contributes to the base of knowledge necessary for further work in elucidating the metabolic pathway responsible for dicamba degradation in *P. maltophilia*, DI-6. Further work in understanding this pathway is warranted by the need to develop more efficient and effective herbicide-resistant crop systems for use in agricultural production.

ACKNOWLEDGEMENTS:

I would like to extend my most sincere gratitude to Dr. Donald Weeks and Mr. Mark Behrens, M.S. for their advice, support, and day-to-day guidance throughout this project. Thanks also to Dr. Razvan Dumitru for his direct supervision and enthusiasm in light of favorable results. Thanks also to all the members of the Weeks Lab group and to the Department of Biochemistry for creating a positive, challenging, and intellectually-stimulating atmosphere in which to work and learn.

Mrs. Joan Krush also deserves particular thanks, for her constant support and timely advice on deadlines, classes, and careers in Biochemistry. Without her help, the field of Biochemistry may never have established itself in my interests as strongly as it has.

I would like to express my heartfelt thanks to my family: Chuck, Janel, and Katie Hibberd. My father has provided for me the scientific, academic, and moral example toward which I have always strived, and instilled in me the scientific curiosity and perspective which has led me toward an interest in scientific research. Thanks so much to my mother, for inspiring me with creativity and a positive attitude, even in times of extreme doubt regarding both my present and my future. Thanks also to my sister for constantly challenging my thoughts and actions as I matured, yet also for providing constant sibling camaraderie. Without her, I may never have learned to hold as strong and defend as enthusiastically my beliefs and ideas.

This work was funded in part by a grant from the Undergraduate Creative Activities and Research Program (UCARE) at UNL.

TABLE OF CONTENTS:

Heading	Page
A. Literature Review	1-12
A.1 <i>Biological degradation of chemicals of environmental concern</i>	1
A.2 <i>Isolation and identification of microorganisms for the</i>	3
<i>degradation of dicamba</i>	
A.3 <i>Role of Pseudomonas megaplastids in the degradation of</i>	5
<i>dicamba and other xenobiotic compounds</i>	
A.4 <i>Summary and relevancy</i>	11
B. Materials and Methods	13-21
B.1 <i>Source and growth of microorganisms</i>	13
B.2 <i>Isolation of megaplastids from P. maltophilia, strain DI-6</i>	14
<i>and from BAC Clones</i>	
B.3 <i>Electrophoretic separation and size determination of megaplastids</i> . . .	14
<i>by Pulse-Field Gel Electrophoresis</i>	
B.4 <i>Extraction and purification of individual megaplastids</i>	17
<i>from agarose gel</i>	
B.5 <i>Antibiotic resistance marker “tagging” of megaplastids and</i>	18
<i>transformations into competent E. coli</i>	
B.6 <i>Sequencing and bioinformatic analysis of total P. maltophilia</i>	20
<i>megaplastids</i>	
C. Results	21-38
C.1 <i>Electrophoretic separation and size determination of megaplastids</i> . . .	21
<i>by Pulse-Field Gel Electrophoresis</i>	
C.2 <i>Antibiotic resistance marker “tagging” of megaplastids and</i>	29
<i>transformations into competent E. coli</i>	
C.3 <i>Sequencing and bioinformatics analysis of DNA from pooled</i>	30
<i>P. maltophilia megaplastids</i>	
D. Discussion and Implications	38-45
D.1 <i>Molecular characterization of P. maltophilia, strain DI-6</i>	38
<i>megaplastids</i>	
D.2 <i>Demonstration of the ability to isolate, separate, “tag” with</i>	39
<i>antibiotic resistance, and transform P. maltophilia megaplastids</i>	
D.3 <i>Initial bioinformatics analysis of the P. maltophilia</i>	40
<i>extrachromosomal gene complement</i>	
D.4 <i>Proposal for and implications of future research</i>	42
E. References	46-48

LIST OF FIGURES:

<i>Figure</i>	<i>Page</i>
Figure 1: Enzymatic Activity of dicamba O-Demethylase	7
Figure 2: Proposed scheme for the enzymatic conversion of dicamba to 3,6-dichlorosalicylic acid by the dicamba monooxygenase of <i>P. maltophilia</i> , strain DI-6.	9
Figure 3: CHEF gel, indicating high molecular weight sizing ladders and samples of <i>Pseudomonas maltophilia</i> , strain DI-6 megaplasms.	23
Figure 4: Calibration curve and data for sizing megaplasmsid DNA, includes BAC Clones (BACPAC Resources, Oakland, CA) and a high MW DNA marker (Invitrogen, Carlsbad, CA).	25

LIST OF TABLES:

<i>Table</i>	<i>Page</i>
Table 1: Optimized CHEF gel electrophoresis conditions for separation of <i>Pseudomonas maltophilia</i> , strain DI-6 megaplasmsids.	15
Table 2: Table of <i>P. maltophilia</i> megaplasmsid sizes and relative abundance estimates. (Gel image from a preparative CHEF Gel for illustration)	27
Table 3: Table(s) of predicted ORFs, estimated/determined function, and elementary statistical analysis, classified by functional category	31

A. LITERATURE REVIEW:

A.1 Biological degradation of chemicals of environmental concern

In the increasingly environment-conscious perspective of modern science and society, avoiding potentially dangerous contamination of natural resources is a prime consideration in the minds of many. A large number and diversity of synthetic chemicals are present in the air, inland waters, marine waters, and in both agricultural-use and nonagricultural-use soils, having been introduced by intentional and directed application, deliberate “dumping,” or inadvertent contamination (Alexander, 1981). Some of these synthetic compounds pose a threat to the naturally-existing ecosystem, as manifest in primary toxicity or the potential for chemical conversion into toxic compounds (Alexander, 1981). Danger to the immediate vicinity is compounded by the ability of such contaminants to be further distributed by evaporation, surface water flow, leaching downward through the soil into the subterranean water table, and uptake by animal species with subsequent redistribution in waste products (Newton, 1990). The ease by which such contaminants are incorporated into the soil, runoff irrigation water, and groundwater reserves by careless agricultural practice makes a consideration of the situation of environmental contamination with synthetic chemicals particularly relevant to agriculturally-focused areas such as the Midwest United States.

The ability of synthetic compounds to be incorporated into the ecosystem creates a particular difficulty in controlling their deleterious effects after deliberate or inadvertent environmental contamination. However, synthetic organic chemicals which have been introduced into the environment are subject to the same natural forces that act on

environmental chemicals of biological origin. On one side, abiotic reactions such as photolysis or chemical hydrolysis provide avenues by which potentially-harmful synthetic chemicals can be inactivated in water and soil (Newton, 1990). Such abiotic degradation can be significant under limited circumstances and dependent on the contaminant itself; however, the ability of heterotrophic bacteria and fungi to act on such compounds far surpasses the degradative ability of abiotic processes in many cases (Alexander, 1981).

The deactivation of contaminant form and mineralization of an organic molecule into its simplest inorganic constituents is nearly always a consequence of the activity of environmental microbial communities (Alexander, 1981). Populations of heterotrophic bacteria and fungi, *via* evolving to participate in the catabolism of new sources of nourishment, reap significant benefits from degradative processes; otherwise, such selection-developed processes would not be biologically-favored. Some of the carbon in the synthetic organic substrate can be used by the microbes to construct and replenish cell constituents (Alexander, 1981). In addition, the breakdown of new nourishment compounds as they feed into the catabolic pathways of the organism releases previously untapped energy, which can then be used to support increases in metabolism, organism number, and biomass (Alexander, 1981). Examples of such biologically-mediated degradation are apparent through studies of the ability of nonsterile environmental samples to degrade compounds such as DDT, 2,4-D, aldrin, heptachlor, and many other chlorinated and non-chlorinated organic compounds (Alexander, 1981). These studies further supported the essential role of biotic degradation in environmental samples by

establishing the inability of sterilized environmental samples to produce significant catabolic effects of the relevant synthetic compounds (Alexander, 1981).

A.2 *Isolation and identification of microorganisms for the degradation of dicamba*

The synthetic compound 3,6-dichloro-2-methoxybenzoic acid, commonly named dicamba, has been used as a pre- and post-emergent herbicide for the control of broad-leaf weeds and many grassy weeds since 1965 (Krueger, 1989). Dicamba acts preferentially on broadleaf weeds by mimicking auxins (e.g. indole-3-acetic acid), important plant hormones which are fundamental plant cellular development regulators when present at low concentration. Application of dicamba and similar herbicides at high concentration effectively unbalances plant cell developmental regulation, leading to plant damage and death by promoting uncontrolled cell growth and division. The herbicide is commonly used to control dicotyledon weeds in fields of corn, small grains, sugarcane, and turf grass (Krueger, 1991). Thus, dicamba is a widely used and agriculturally important herbicide; however, its broad phytotoxic ability also can cause parallel damage to broad-leaf crops if not correctly applied or if the herbicide persists in surplus in the soil on a long term basis. More importantly, dicamba is an important complement to crop treatment with glyphosate (Roundup®), an herbicide so ubiquitous that environmental selection pressure is driving increasing development of resistance in target weed species (Owen *et al.*, 2005).

Significant evidence suggests that the mineralization of dicamba in aerobic soil and water is microbiologically mediated (Krueger, 1989; Caux *et al.*, 1993; NPIC, 2002). Under normal, ambient conditions in soil, dicamba is stable to oxidation and hydrolysis.

Subsequently, half-life for dicamba can range from four to 555 days, with an average half-life of 14-28 days (Caux et al., 1993; NPIC, 2002). Time required for clearance of dicamba from soil is primarily dictated by the catabolic activity of soil microorganisms (Caux et al., 1993; NPIC, 2002). In water, this chemical is somewhat susceptible to photolysis, although this type of breakdown does not occur on the surface of soil (NPIC, 2002). Leaching is a particularly notable possibility for contamination of groundwater by dicamba, as this chemical has been shown to be highly mobile in soil due to poor adsorption characteristics (NPIC, 2002). Although this herbicide is persistent in soil ($t_{1/2}$ = 14-28 days) and water ($t_{1/2}$ = 7 days), it exhibits very low toxicity in humans and wildlife including birds, fish, and insects (NPIC, 2002). Therefore, the main consideration for controlling levels of dicamba contamination in the environment is with the intent of limiting inadvertent concurrent damage to crops in the immediate vicinity of herbicide application.

The evidence that degradation of dicamba is microbiologically-mediated led in turn to research efforts directed at isolating environmental organisms capable of degrading dicamba and studying their unique metabolic ability. Isolation of such species was pursued so as to provide both a tool optimizing environmental contaminant clearance and as a supply of genetic material that could be utilized in engineering dicamba-resistant higher plants, most significantly crop species (Krueger, 1989). Krueger et. al. was successful in isolating dicamba-degrading organisms from soil/water samples obtained from a storm water retention pond at a dicamba manufacturing plant at Beaumont, Texas (Krueger, 1989). Significant isolates from this study included members of genus

Pseudomonas, which are known for the ability to perform unique and diverse catabolic reactions of novel compounds (Krueger, 1989).

A.3 *Role of Pseudomonas megaplasmids in the degradation of dicamba and other xenobiotic compounds*

The *Pseudomonas* taxonomic group is a bacterial genus known for well-documented activity and versatility in processes of mineralizing synthetic and natural organic compounds in nature (Palleroni, 1975). Their widespread presence in soil and water, coupled with catabolic diversity and relative ease of isolation, have made this genus particularly common in enrichment cultures intended to isolate bacteria with novel degradative pathways (Palleroni, 1975). By the mid-20th century, researchers had concluded that often the degradative ability of the *Pseudomonas* was conferred by the presence of partial or complete catabolic pathways located on extrachromosomal plasmids, a trait confirmed in part by plasmid curing studies performed by A. M. Chakrabarty in the early 1970s (Chakrabarty, 1972; Holloway, 1975). Further, members of this genus are known to carry multiple compatible degradative plasmids in each cell, thus giving the particular microorganism the ability to deactivate and/or utilize a relatively large number of uncommon nutrients, for example, camphor, naphthalene, salicylate, and crude oil, among others (Chakrabarty, 1976).

The species *Pseudomonas maltophilia*, strain DI-6 is notable for its effective role in degrading the commercial broadleaf herbicide dicamba (Krueger, 1989; Krueger, 1991; Cork *et al.*, 1995). Further study of the ability of strain DI-6 to mineralize dicamba led to the identification of the enzyme responsible for the conversion of dicamba (3,6-dichloro-2-methoxybenzoic) to DCSA (3,6-dichlorosalicylic acid), the first step in the

degradation pathway of this compound (Cork *et al.*, 1995; Wang *et al.*, 1997). (Figure 1)

The initial enzymatic component of this degradation pathway was determined to function as a three-component system, with successful enzymatic activity dependent on the interplay of the Rieske [2Fe-2S] cluster-containing dicamba monooxygenase, a small ferredoxin, a flavin-containing reductase, and the essential enzyme cofactors Oxygen (O₂), NADH, and Mg²⁺ (Wang *et al.*, 1997; Chakraborty *et al.*, 2005; Herman *et al.*, 2005). (Figure 2) Most relevant to this particular consideration of the xenobiotic degradation roles of extracellular genetic elements was the discovery that the monooxygenase component of the dicamba O-demethylase is located strictly on cytosolic megaplastids in *P. maltophilia*, strain DI-6, likely in multiple copies (Chrastil, 2000; Herman *et al.*, 2005). Further, experiments in which total megaplastid samples from *Pseudomonas* strain BRW-3 (a dicamba-degrading strain closely related to strain DI-6) were transformed into an *Escherichia coli* host suggested that the ability to utilize dicamba as a sole carbon source was conferred specifically and wholly by the megaplastids themselves with minor contribution by the host cell chromosomal DNA (Herman, 1998; Chrastil, 2000). Additional analysis by scientists at Amman University, Jordan, and the Illinois Institute of Technology supported the conclusion that megaplastids are essential for the degradation of dicamba in *P. maltophilia*, strain DI-6 (Cork *et al.*, 1995; Khalil *et al.*, 2000). In summary, all genes strictly necessary for the use of dicamba as sole carbon source (i.e. the complete mineralization pathway) are located on the megaplastids of the *Pseudomonas* species studied, and these genes therefore would be available to isolate and analyze in pursuit of understanding the mineralization pathway of dicamba.

Figure 1: Enzymatic Activity of dicamba monooxygenase, the initial step in the catabolic pathway of dicamba in *P. maltophilia*, strain DI-6

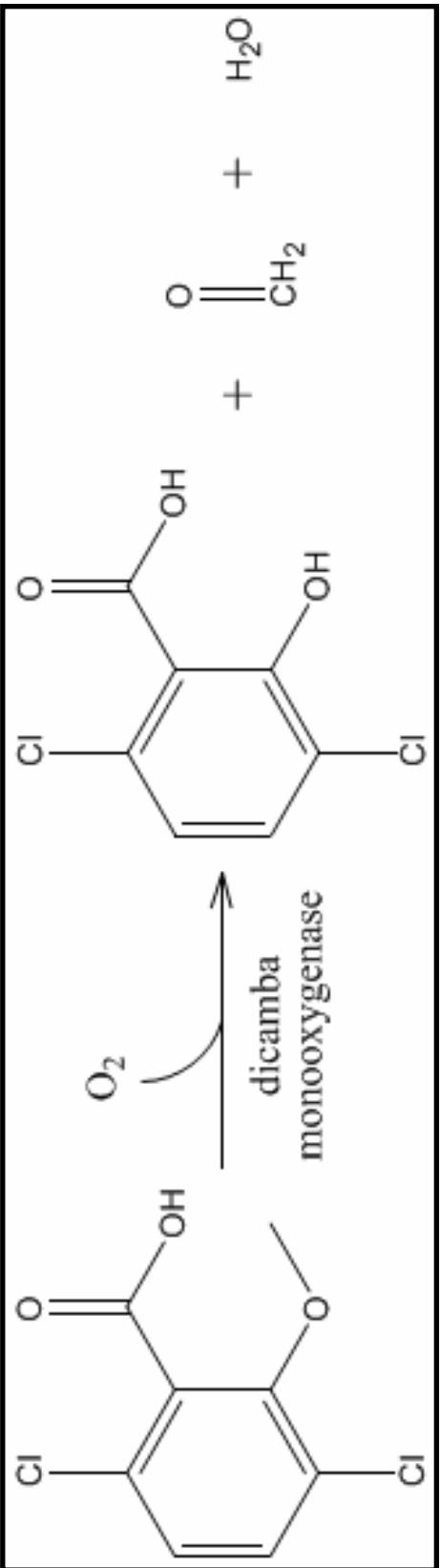
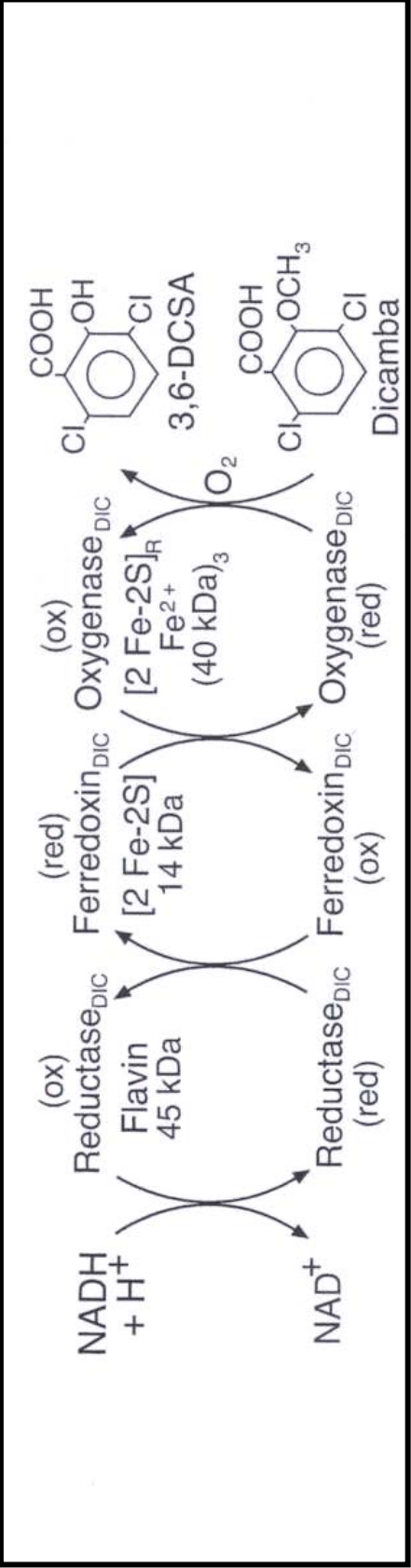


Figure 2: Proposed scheme for the enzymatic conversion of dicamba to 3,6-dichlorosalicylic acid by the dicamba monooxygenase of *P. maltophilia*, strain DI-6 (Chakraborty et al., 2005). This figure indicates the participation of a flavin associated reductase and a [2Fe-2S] iron-sulfur cluster associated ferredoxin which participate with dicamba monooxygenase in performing O-demethylase activity.



A.4 Summary and Relevancy

A primary goal of engineering genetically-modified crop organisms capable of withstanding treatment with synthetic herbicides is, in summary, to decrease the amount of synthetic chemicals which enter the environment as contamination. Whether the herbicide enters the environment by illegal dumping or intentional application to crops and subsequent runoff, leaching, and concomitant contamination, taxation on the environment can be decreased by efficient, directed, and specific herbicide application. Positive strategies for supporting such conscientious action would be supported by the development of efficient herbicide-resistant crops or by the engineering of organisms able to clear xenobiotic contaminants from the soil.

Recent research also suggests that many common agricultural weeds, including horseweed (*Conyza canadensis* (L) Cronq), wild buckwheat (*Polygonum convolvulus* L), and the ubiquitous common lambsquarters (*Chenopodium album* L) have developed increasing resistance to herbicide application (Owen *et al.*, 2005). Interestingly, observation of herbicide resistant weeds actually contributed to the academic insight that herbicide resistant crops could potentially be developed (Duke, 2005). Particularly in the case of one of the most commonly-applied herbicides, glyphosate (commonly used under the trade name Roundup®), developing resistance of weed species urges the agricultural users of the chemical to either apply elevated quantities in attempts to control the more resistant weeds or to utilize additional herbicides. In addition, the nonspecific phytotoxic activity of glyphosate, to which no plants are known to be naturally resistant, creates much more danger to adjacent crops which may have contact with increased environmental contamination concomitant with over-application (NPIC, 2000).

Probable environmental dangers of this type have encouraged research geared toward creating genetically-modified crops which are resistant to more species-selective, less broadly phytotoxic synthetic herbicides, like dicamba. In fact, researchers have already engineered soybeans, tomato, and tobacco with the dicamba monooxygenase gene from *P. maltophilia*, strain DI-6, which are able to withstand 10-20 fold higher dicamba levels than the standard commercial application rate (Weeks Lab, Unpublished Research). The use of specific herbicide application in conjunction with effectively-engineering resistant crops would allow agricultural producers more efficiently target weed species. Thus, the ability to complement application of a nonspecific herbicide such as glyphosate with a specific, nonhazardous herbicide like dicamba would potentially allow for an increased specific effectiveness, decreased application amount, and subsequently decreased environmental contamination.

In summary, investigating the source of the herbicide inactivation and mineralization genes used to engineer herbicide-resistant crops is exceptionally relevant to efforts aimed at developing and improving such crop systems. In the pursuit of this goal, identifying and characterizing the genes involved in the mineralization pathway of dicamba from microbiological sources would allow researchers to potentially optimize the resistance mechanisms developed for and engineered into crop systems. It is known that the megaplasmids of *Pseudomonas maltophilia*, strain DI-6, hold the essential genes for dicamba mineralization. Thus, a molecular and bioinformatic characterization of this extrachromosomal DNA provides a fundamental resource of data and experience for engineering more effective dicamba-resistant crop species and for understanding enzymatic systems involved in the degradation of xenobiotic compounds.

B. MATERIALS AND METHODS:

B.1 Source and growth of microorganisms

Cultures of *Pseudomonas maltophilia* strain DI-6 were obtained from a storm water retention pond at the dicamba manufacturing plant at Beaumont, Texas and provided to the Weeks Laboratory by Dr. Douglas Cork. (Krueger, 1989) Solid medium containing 5 mM dicamba and Gelrite (10g/L), rather than agar, was used to start *P. maltophilia* DI-6 cultures from -80° C stocks before inoculation into liquid medium. Starter cultures of 5 mL sterile reduced chloride medium, abbreviated RCL (Krueger, 1989) were inoculated from very low temperature (-80° C) glycerol stocks and allowed to grow approximately 12 hours at 30° C with shaking at 240 RPM. Liquid cultures of this type were supplemented with either 5 mM dicamba or 2 mg/mL glucose and 2 mg/mL casamino acids. At saturation density, the starter cultures exhibited the presence of a bright pink byproduct, which seemed to signal the achievement of maximum culture growth density. Starter cultures were pelleted at 4800 RPM at 4° C for 10 minutes in a Beckman GS-15R centrifuge and resuspended in fresh RCL, supplemented with a suitable carbon source. After a second 12 hour incubation, 2 mL of the starter culture was used to inoculate 1 L of sterile RCL media containing the same carbon source supplement. A growth curve was constructed for the 1 L cultures of *P. maltophilia*, strain DI-6 with OD₆₀₀ measurements taken approximately every 6 hours. Based on the growth curve, the large cultures were grown to an OD₆₀₀ of 0.600 ODU, corresponding to late-logarithmic stage, which was achieved approximately 48 hours after the 2 mL inoculation as described above. One liter cultures of DI-6 cells were pelleted at 6000 RPM at 4° C for

15 minutes in the Beckman J25I centrifuge. The supernatant was decanted and the pellets stored at -20° C.

B.2 Isolation of megaplasמידs from P. maltophilia, strain DI-6 and from BAC Clones

In order to isolate total megaplasמידs from fresh or frozen (-20° C) pellets of DI-6 cells, a Qiagen® Plasmid Maxi Kit was used in the QIAGEN tip-500 format. The Very Low-Copy Plasmid/Cosmid protocol was followed, with an additional 70% Ethanol wash performed after the first isopropanol precipitation. Very gentle handling was required at each step so as to avoid megaplasמיד contamination with chromosomal DNA and mechanical shearing of the large megaplasמידs themselves. Wide-bore pipette tips were used in most handling steps requiring pipetting. Plasmid DNA was resuspended in Qiaprep EB buffer (10 mM Tris, pH 8.5) and stored at -20° C.

BAC clone DNA from BACPAC Resources of Oakland, CA was isolated by a modified Qiagen-Tip protocol, as supplied by the manufacturer of the clones. This procedure relied on isopropanol precipitation rather than the standard column-based preparation protocol to collect BAC DNA of very high molecular weight.

B.3 Electrophoretic separation and size determination of megaplasמידs by Pulse-Field Gel Electrophoresis

Due to the large anticipated sizes of the megaplasמידs isolated from *P. maltophilia*, strain DI-6, a low percentage agarose Pulse-Field Electrophoretic separation was performed. Contour-clamped, homogeneous electric field electrophoresis was performed with a Bio-Rad CHEF-DR® III system. Optimized conditions for the electrophoretic separation are listed in Figure 3.

Table 1: Optimized CHEF gel electrophoresis conditions for separation of *Pseudomonas maltophilia*, strain DI-6 megaplasms.

<i>Bio-Rad CHEF-DR[®] III system Run Conditions (optimized)</i>	
Gel	0.7% standard agarose gel, 100 mL
Initial Switch Time	2 seconds
Final Switch Time	2 seconds
Set Volts	3.5 V/cm
Included Angle	120°
Buffer	TAE @ 14° C
Run Time	24-26 hours

Gels were stained by soaking in an ethidium bromide solution at approximately 1 $\mu\text{g}/\text{mL}$ concentration and photographed by UV-based imaging on a FOTODYNE gel documentation apparatus.

In order to accurately size the large bacterial megaplasמידs, for which no suitable commercially-produced DNA sizing ladders were available, a ladder was constructed by the use of bacterial artificial chromosomes (“BACs”) of known size. BAC clones of known size (BACPAC Resources, Oakland, CA) from 35 kilobases to 220 kilobasepairs were run individually on a CHEF gel (see above description) in either a NotI-digested (linearized) or undigested form. In addition, a high MW DNA marker (Invitrogen, Carlsbad, CA) with band sizes from 8.3 kilobases to 48.5 kilobases was used to cover the lower range of the megaplasמיד size spectrum. Relative migration distances (R_f) from the BAC clones of known size and high MW marker were used to construct size calibration curves for each subset of sizing standards (R_f vs. log basepairs). A curve fit equation generated from each subset of standards was used to determine the size of each high molecular weight megaplasמיד from *P. maltophilia*, strain DI-6, by interpolation. Finally, sizing gels were run in duplicate and predicted sizes were averaged to obtain a more accurate estimate.

B.4 Extraction and purification of individual megaplasמידs from agarose gel

Standard methods for purifying common size DNAs from gels are relatively ineffective above approximately 50 kilobases size, and are more-suited for purifying DNA in the lower molecular weight range range (e.g. PCR-amplified fragments).

Therefore, in order to obtain isolated samples of each individual megaplasמיד for further

analysis, electro-elution was performed on excised bands from CHEF gels with a Model 422 Electro-Elutor apparatus. (Bio-Rad, Hercules, CA)

Bands representing an individual megaplasmid size from each lane on a CHEF gel were excised and pooled. Pooled gel slices of each size were gently minced and loaded into the apparatus. The manufacturer's protocol was followed for loading, buffer concentration, and power settings, although run times were much longer, and varied with megaplasmid size from 3 hours to 12 hours. The electro-elution cell was kept in an ice bath, and elution buffer (40 mM Tris, 20 mM Glacial Acetic Acid, 1 mM EDTA, 0.1% SDS) was changed approximately every 4 hours, when current readings on the power source suggested buffer failure. After elution for an appropriate time, DNA samples were collected from the apparatus, then purified by phenol-chloroform extraction, ethanol precipitation, and resuspension in either 10 mM Tris Buffer, pH 8.0 or sterile, deionized H₂O, for storage at -20° C until further analysis. Samples for later transformation were stored in sterile, deionized H₂O rather than buffer, as the chemical contents of the buffer tended to interfere with electroporative transformation in later procedures.

B.5 Antibiotic resistance marker "tagging" of megaplasms and transformations into competent E. coli

The natural megaplasms of *P. maltophilia*, strain DI-6 do not possess a suitable selectable marker which would allow identification of competent *E. coli* transformants. Early experiments performed in which potential transformants of natural megaplasms were plated on 5 mM dicamba were inconclusive, therefore, a transposition-based EZ-Tn5™ <oriV/KAN-2> inducible origin of replication/Kanamycin resistance system

(EPICENTRE[®] Biotechnologies, Madison, WI) was used to “tag” megaplasמידs for subsequent transformation into EPI300[™] Electrocompetent *E. coli*. (EPICENTRE[®] Biotechnologies, Madison, WI) Manufacturer’s protocol was followed for this procedure.

Successfully tagged megaplasמידs were transformed by electroporation in a Bio-Rad GenePulser[®] II system with settings 2.5 kV, 25 μ F, 200 Ω and a 2 mm gap cuvette. (Bio-Rad, Hercules, CA) After recovery for 1 hour in SOC medium (Tryptone 2% (w/v), Yeast extract 0.5% (w/v), NaCl 8.6 mM, KCl 2.5 mM, MgSO₄ 20 mM, Glucose 20 mM), electroporated cells were plated on Luria-Bertani media plates ((10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl, pH 7.5) supplemented with 50 μ g/mL Kanamycin plates and incubated overnight at 30° C. A growth temperature of 30° C was used in an effort to more easily facilitate the replicative machinery of the megaplasמיד source organism, as *Pseudomonas* grows preferentially at a lower optimal temperature than *E. coli*.

Further treatment of transformants is an adaptation of the EPICENTRE[®] EPI300[™] electrocompetent cells and <oriV/KAN-2> induction procedure.

Transformation events from each selection plate were inoculated into 50 mL cultures of Luria-Bertani broth media supplemented with 50 μ g/mL Kanamycin and grown overnight at 37° C overnight with shaking at 220 rpm. All 50 mL of starter culture were inoculated into 500 mL of the same media and incubated 30 min at 37° C with 220 rpm shaking. After the 30 min incubation, 0.5 mL of the manufacturer-provided 1000x Induction Solution was added and the 550 mL cultures were incubated at 37° C for 2 hours with 250 rpm shaking. Cultures of this treatment were then centrifuged to pellet cells and

plasmid DNA was isolated by the Qiagen Very Low Copy protocol described above or the pellets were stored at -20° C for later DNA isolation.

B.6 *Sequencing and bioinformatic analysis of total P. maltophilia megaplasמידs*

A sample of total megaplasמיד DNA was sent for sequencing by MWG Biotech, Inc. of High Point, NC. Upon return of sequence data in text file format, data was imported by individual contig into Vector NTI (Invitrogen, Carlsbad, CA). Each contig was analyzed initially by scanning for predicted open reading frames on both the direct and complementary strand with size stringency limited to ORFs with greater than 100 codons. Predicted ORFs were translated with the Vector NTI software and compared by Protein-Protein BLAST to existing proteins in the National Center for Biotechnology Information databases. The most significant match, if found, for each predicted ORF was tabulated along with predicted or determined function, homology information, and source organism data to create a list of predicted ORFs localized to the total megaplasמיד pool of *P. maltophilia*, strain DI-6.

In order to identify the dicamba monooxygenase gene copy number in the *Pseudomonas maltophilia* megaplasמיד complement, which was one goal of particular interest, multiple sequence alignment was performed with the VectorNTI-associated AlignX tool, using the ClustalW algorithm. Individual contigs were compared to both the direct and complementary strand of the previously-sequenced dicamba monooxygenase gene to identify sequence homology. In addition, incomplete or fragmentary ORFs, as predicted by the VectorNTI software on the borders of each contig, were aligned against the direct and complementary sequences of the DMO. This

experiment was performed for the sake of completeness in considering all available sequence data as potentially containing copies of the DMO gene.

C. RESULTS:

C.1 *Electrophoretic separation and size determination of megaplasmiids by Pulse-Field Gel Electrophoresis*

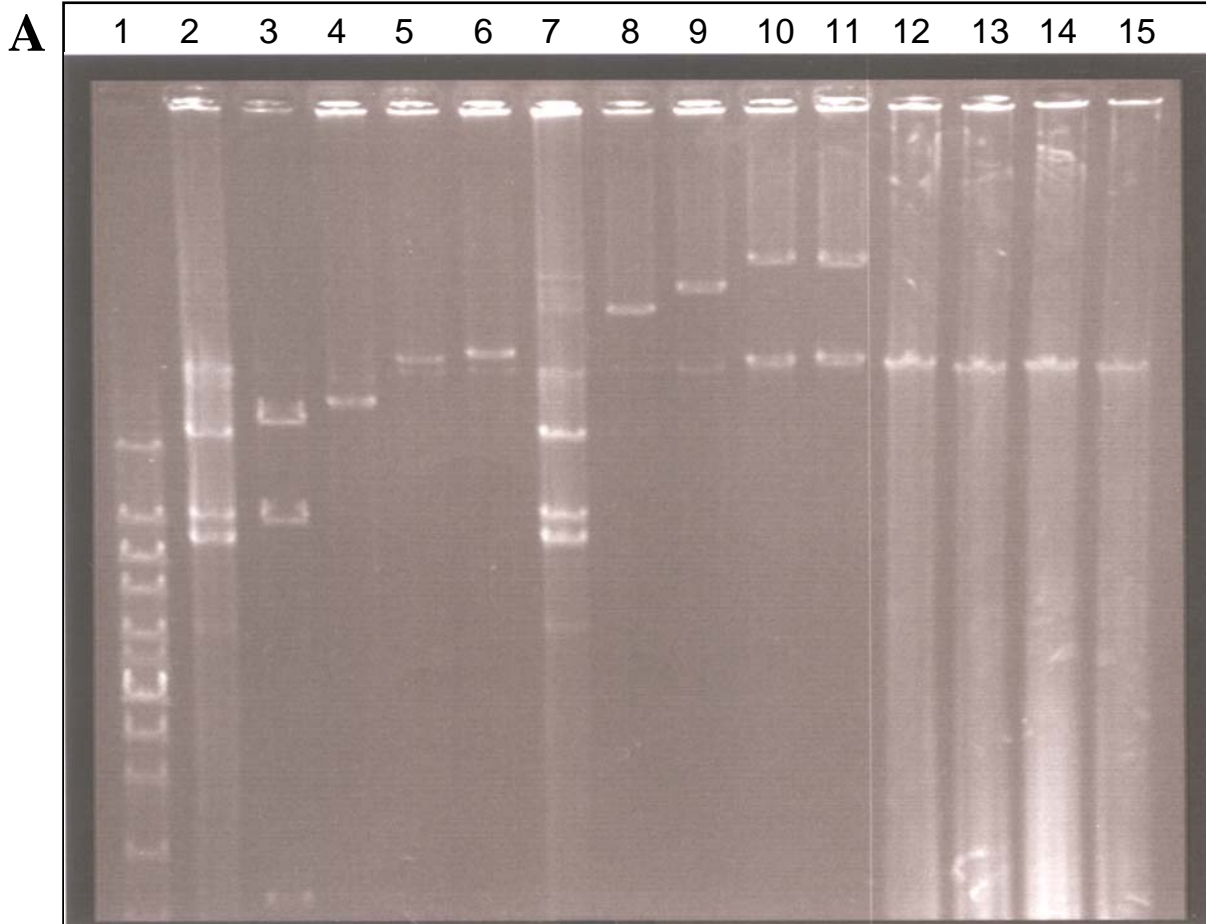
Through pulse-field gel electrophoresis under optimized conditions, the megaplasmiids of *Pseudomonas maltophilia* DI-6 were separated into a unique, discrete banding pattern. The use of bacterial artificial chromosomes and a high molecular weight ladder, both of known size, allowed the construction of a size-determination calibration curve, against which lanes of megaplasmiids from *P. maltophilia* were compared.

Previous work involving pulse-field electrophoretic separation had suggested that the megaplasmiid complement of *P. maltophilia*, strain DI-6 consisted of four distinct extrachromosomal plasmids (Chrastil, 2000). However, running an analogous separation under optimized electrophoresis conditions, in conjunction with an improved alkaline lysis preparatory isolation method, suggested that the megaplasmiid complement of *P. maltophilia* consisted not of four megaplasmiids, but rather eight unique, distinct plasmids of varying size.

A size determination was performed for each of the *P. maltophilia* megaplasmiids, based on comparison to a high molecular weight standard constructed by a combination of bacterial artificial chromosomes of known size and an additional commercially available high molecular weight ladder. In addition, relative abundance estimates were

made for the proportions of the megaplasmid pool belonging to each size based on band intensity.

Figure 3: CHEF gel, run under optimized conditions. Figure 4A indicates the ethidium bromide stained gel image visualized by UV-imaging. Figure 4B indicates the content and treatment of each lane. The Invitrogen High Molecular Weight Ladder (10 kB to 48.5 kB size) is in Lane 1. Lanes 3-6, 9-11 include BACs of known size (49.9 kB to 220 kB sizes). Combination of the data from these two sets of known-size data were used to generate a calibration curve for sizing the total isolated megaplasms of *P. maltophilia*, strain DI-6 contained in Lanes 2 and 7. Lane 8 contains DNA from a control transformant from the EZ-Tn5 “tagging” reaction, and Lanes 12-15 contain DNA from transformations with tagged *P. maltophilia* megaplasms. Specifically, DNA in Lanes 12 and 13 was obtained from *E. coli* transformed with tagged samples of pooled *P. maltophilia* megaplasms. DNA in Lanes 14 and 15 was isolated from *E. coli* transformed with the p38 megaplasms from *P. maltophilia*, DI-6.

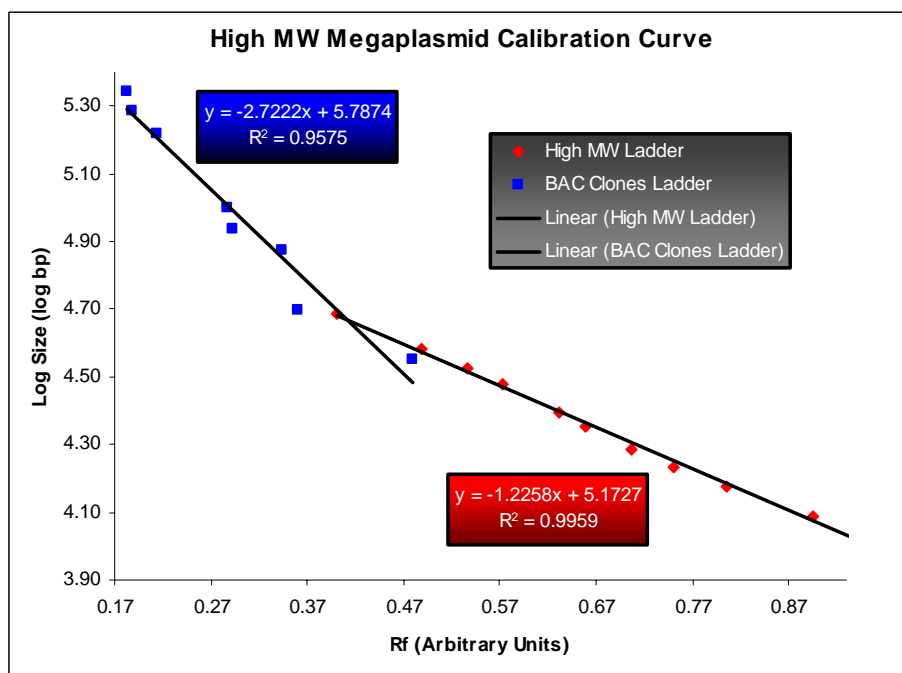


B

Lane	Load
1	Invitrogen High Molecular Weight DNA Markers
2	<i>P. maltophilia</i> , strain DI-6 Total Megaplasms
3	BAC Clone 193*, NotI Digest
4	75 kB BAC*, No Digest
5	BAC Clone 193*, No Digest
6	100 kB BAC*
7	<i>P. maltophilia</i> , strain DI-6 Total Megaplasms
8	EPICENTRE EZ-Tn5 Kit Control BAC (provided)**
9	BAC Clone 202*, No Digest
10	BAC Clone 241*, No Digest
11	BAC Clone 248*, No Digest
12	EZ-Tn5 Transformant Tot-2, Isolated Plasmids
13	EZ-Tn5 Transformant Tot-4, Isolated Plasmids
14	EZ-Tn5 Transformant p38-1, Isolated Plasmids
15	EZ-Tn5 Transformant p38-3, Isolated Plasmids
	*BAC Clones courtesy of BACPAC Resources, Oakland, CA
	**EZ-Tn5 Transposon "tagging" kit from EPICENTRE® Biotechnologies, Madison, WI

Figure 4: Calibration curve and data for sizing megaplasmid DNA. Figure 5A indicates the calibration curve, which was generated from data obtained from the CHEF gel pictured in Figure 4. Figure 5B contains the data used to generate the calibration curves, including the Invitrogen High MW DNA ladder, the known size BAC clones from BACPAC Resources. Due to migration effects manifest differently at high versus low molecular weight, the higher size BACs were plotted separately than the lower size ladder, and used to generate two distinct calibration curves. R_f data from the bands containing DNA from *P. maltophilia*, DI-6 total megaplasms and *E. coli* megaplasmid transformants was compared to the relevant calibration curve to interpolate an accurate size for each band.

A



B

	Size (bp)	Log Size	Rf
High MW Ladder	48,502	4.6858	0.4015
	38,416	4.5845	0.4894
	33,498	4.5250	0.5361
	29,942	4.4763	0.5732
	24,776	4.3940	0.6308
	22,621	4.3545	0.6597
	19,399	4.2878	0.7063
	17,057	4.2319	0.7503
	15,004	4.1762	0.8052
	12,220	4.0871	0.8957
	10,086	4.0037	0.9699
BAC Clones Ladder			
BAC 193/NotI lane 3	49,980	4.6988	0.3603
BAC 193/NotI lane 3	35,740	4.5532	0.4799
75 kB BAC lane 4	75,000	4.8751	0.3439
BAC 193 Uncut lane 5	86,000	4.9345	0.2931
100 kB BAC lane 6	100,000	5.0000	0.2876
EZ-Tn5 Control (~165 kB) lane 8	165,000	5.2175	0.2409
BAC 202 Uncut lane 9	165,000	5.2175	0.2135
BAC 241 Uncut lane 10	220,000	5.3424	0.1833
BAC 248 Uncut lane 11	192,000	5.2833	0.1874
DI-6 Plasmids			
	181,189	5.2581	0.1997
	160,157	5.2045	0.2190
	141,657	5.1512	0.2382
	85,924	4.9341	0.3164
	53,469	4.7281	0.3906
	37,679	4.5761	0.4867
	34,801	4.5391	0.5169
	25,872	4.4128	0.6199
Transformants			
Tot-2	92,008	4.9638	0.3057
Tot-4	89,627	4.9524	0.3098
p38-1	90,375	4.9560	0.3085
p38-3	90,375	4.9560	0.3085

Table 2: Table of *P. maltophilia* megaplasmid sizes and relative abundance estimates. (Gel image from a preparative CHEF Gel for illustration) Abundance estimates based on intensity must take into account the effect of increased ethidium bromide binding to larger size DNA, and thus higher intensity, even though true stoichiometric evaluation may not indicate the same abundance comparison.

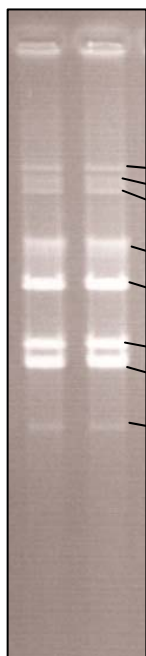


Table of P. maltophilia megaplasmid calibration-determined sizes and estimates of relative megaplasmid abundance

Arbitrary Plasmid Designation	Size (basepairs)	Relative Abundance [Band Intensity Rated 1 (Low) to 10 (High)]
p181	181,189	2
p160	160,157	2
p142	141,657	2
p86	85,924	4
p53	53,469	8
p38	37,679	8
p35	34,601	8
p26	25,872	2

C.2 *Antibiotic resistance marker “tagging” of megaplasמידs and transformations into competent E. coli*

Control test reactions performed with the 165 kilobase BAC provided by the kit manufacturer produced Kanamycin resistant transformants (at 50 µg/mL in LB), positive response to the induction procedure, and relatively easy isolation of the BAC from induced transformant culture.

Tagging reactions with the EZ-Tn5 kit performed with total megaplasמידs from *P. maltophilia* and a purified sample of the 38 kilobase plasmid from strain DI-6 suggested successful transformations into electrocompetent EPI300 *E. coli*. Prospective transformants in both a tagging reaction using a preparation of all megaplasמידs and a tagging reaction with an isolated, purified p38 plasmid grew successfully on LB + 50 µg/mL Kanamycin plates when 200 µL of undiluted liquid culture was plated and incubated overnight at 30°C. Between 20 and 30 colonies were recovered per plate. The numbers of transformants and growth characteristics on selective medium plates indicated successful transformation, however, both sets of transformants did not seem to respond particularly well to the induction procedure. Cells from both tagged megaplasמיד sources and transformation events were grown in a 500 mL format, and plasmid isolation with the Qiagen Very Low Copy protocol yielded distinct bands on multiple CHEF gels. Comparison to the established high MW ladder/BAC system determined a size for these bands of approximately 90 kB size in transformants containing plasmids from both the total megaplasמיד transposon tagging reaction and 38 kB megaplasמיד transposon tagging reaction. (Figure 4)

C.3 Sequencing and bioinformatics analysis of DNA from pooled *P. maltophilia* megaplasms

Sequencing of DNA from the strain DI-6 megaplasms was successful, with a 94% pass rate, an average of 637 high quality bases per read (Phred20 score), and a total of 3.2x coverage over the 1920 clones generated from the pooled *P. maltophilia* megaplasms. Initial assembly of the sequence data by MWG Biotech, Inc. generated 87 contigs, with a total additive contig length of 236,415 basepairs.

Analysis of the assembled contig sequences with the Vector NTI software suite (Invitrogen, Inc.) identified 181 potential open reading frames, with a size range of 452 to 4061 basepairs. Of these ORFs, 82 (45% of total ORFs predicted, average size of 1,170 bp) were identified by function through significant homology to known polypeptides in the NCBI databases through translation and Protein-Protein BLAST analysis. Another 6 ORFs (3.3% of total ORFs predicted, average size 1,547 bp) were identified through BLAST as possessing conserved domains, but exhibited little homology to known proteins in the bioinformatics database other than through these conserved domains. With a total of 93 predicted ORFs (53% of total ORFs predicted, average size 726 bp) it was not possible to provide a positive identity using Protein-Protein BLAST analysis. Of the 236,415 bp of DNA sequence represented by the assembled contigs, sequence analysis allowed positive identification of 97,121 bp, representing approximately 41.1% of the total sequence. The domain-containing and function-identified ORFs from this sequence analysis were classified by category and listed in Figure 7.

Table 3: Table(s) of predicted ORFs, estimated/determined function, and elementary statistical analysis, classified by functional category:

- Replication and Stable Inheritance
- Conjugative Transfer
- Transposase and Resolvase
- Regulation
- Metabolism
- Other
- Unknown

Replication and Stable Inheritance										
Contig	ORF Desig.	Gene Identification and Determined/Predicted Function	Strand	Start	Stop	Size	% Identity	Source	ID # of Closest Relative	
1	9	Replication Initiation Protein	F	19809	20831	1022	62/236 (26%)	Plasmid pLB1	YP_740299	
6	176	Putative Plasmid Stabilization Protein	C	221	1999	1778	236/595 (39%)	Plasmid pLB1	YP_740302	
10	17	Putative DNA Primase (pLB1)	C	3440	4636	1196	281/365 (76%)	Plasmid pLB1	YP_740334	
18	23	Replication Initiation Protein (pLB1)	F	10593	11462	869	62/236 (26%)	Plasmid pLB1	YP_740299	
18	24	Putative Primase (pLB1)	F	11529	12455	926	289/293 (98%)	Plasmid pLB1	YP_740300	
22	35	ArdC-like Antirestriction Protein (pLB1)	F	35	1345	1310	150/291 (51%)	Plasmid pLB1	YP_740303	
28	45	ParA-like Partitioning Protein (pLB1)	F	1059	1844	785	243/252 (96%)	Plasmid pLB1	YP_740330	
49	91	Putative Replication Protein (contains Plasmid replication initiation protein domains)	F	961	1719	758	92/265 (34%)	Sphingomonas sp. SK458	ZP_01304914	
57	107	Putative methylase/helicase	F	175	1887	1712	363/572 (63%)	Sphingomonas wittichii RW1	ZP_01609720	
62	122	ArdC-like Antirestriction Protein (pLB1)	C	10128	10964	836	227/233 (97%)	Plasmid pLB1	YP_740303	
62	126	Replication Primase	C	3642	4589	947	179/310 (57%)	Xanthobacter autotrophicus Py2	ZP_01200680	
62	128	Hypothetical Protein pLB1_p06 (Fic Protein Family (cAMP-induced filamentation))	C	1383	2663	1280	412/426 (96%)	Plasmid pLB1	YP_740304	
65	132	Partition Protein (parA domain, ATPase domain)	F	1788	2456	668	157/219 (71%)	Zymomonas mobilis	AAF23806	
69	145	Probable Partitioning Protein ParA	F	321	992	671	141/223 (63%)	Agrobacterium tumefaciens	YP_086772	
79	152	Putative Methylase/Helicase	C	520	1206	686	138/220 (62%)	Sphingomonas wittichii RW1	ZP_01609720	
81	154	DNA Adenine Modification Methylase	F	43	714	671	189/223 (84%)	Xanthobacter autotrophicus Py2	ZP_01199225	
82	162	Plasmid Replication Initiator Protein	C	3633	4544	911	79/257 (30%)	Sphingomonas wittichii RW1	ZP_01609711	
						Average Size =	1001.5			
						Standard Deviation =	346.0			

Conjugative Transfer									
Contig	ORF Desig.	Gene Identification and Determined/Predicted Function	Strand	Start	Stop	Size	% Identity	Source	ID # of Closest Relative
1	3	Conjugation TrbI Family Protein	C	4778	6046	1268	170/410 (41%)	Sphingomonas wittichii RW1	ZP_01610042
1	4	Conjugal Transfer Protein TrbG/VirB9/CagX	C	6051	7055	1004	129/279 (46%)	Sphingomonas wittichii RW1	ZP_01610043
1	6	TrbL/VirB6 Plasmid Conjugal Transfer Protein	C	8147	9181	1034	111/340 (32%)	Sphingomonas wittichii RW1	ZP_01610045
1	1	Bacterial Conjugative Coupling Protein	C	1859	3736	1877	298/540 (55%)	Sphingomonas wittichii RW1	ZP_01610040
1	5	VirB8 Type IV Secretion Protein	C	7060	8037	977	94/221 (42%)	Sphingomonas wittichii RW1	ZP_01610044
8	181	Putative Transmembrane traG Homologue/Component of Type IV Secretion System	C	185	1267	1082	78/180 (43%)	Rhizobium leguminosarum bv. viciae 3841	YP_766071
10	15	Relaxase (pLB1)	F	1497	2660	1163	264/348 (75%)	Plasmid pLB1	YP_740333
10	16	VirD4 Type IV Secretion Protein (pLB1)	C	4611	6422	1811	596/608 (98%)	Plasmid pLB1	YP_740335
18	26	VirB1 Type IV Secretion Protein (pLB1)	C	8596	9261	665	218/221 (98%)	Plasmid pLB1	YP_740347
18	27	VirB4 Type IV Secretion Protein (pLB1)	C	6714	7952	1238	360/387 (93%)	Plasmid pLB1	YP_740344
18	28	VirB4 Type IV Secretion Protein (pLB1)	C	5510	6671	1061	353/353 (100%)	Plasmid pLB1	YP_740344
18	29	VirB5 Type IV Secretion Protein (pLB1)	C	4783	5662	779	233/237 (98%)	Plasmid pLB1	YP_740343
18	30	VirB6 Type IV Secretion Protein (pLB1)	C	3418	4434	1016	336/338 (99%)	Plasmid pLB1	YP_740341
18	31	VirB8 Type IV Secretion Protein (pLB1)	C	2394	3077	683	227/227 (100%)	Plasmid pLB1	YP_740339
18	32	VirB9 Type IV Secretion Protein (pLB1)	C	1555	2397	842	280/280 (100%)	Plasmid pLB1	YP_740338
18	33	VirB10 Type IV Secretion Protein (pLB1)	C	285	1781	1496	369/391 (94%)	Plasmid pLB1	YP_740337
21	34	Conjugal Transfer Protein TraD	C	1017	2558	1541	295/422 (69%)	Sphingomonas wittichii RW1	ZP_01609818
40	81	unknown (virD2 relaxase domain - Type IV secretory system)	C	82	1461	1379			
65	134	Relaxase/Mobilization Nuclease Domain	F	4685	6094	1409	99/306 (32%)	Aeromonas punctata	YP_067824
82	160	Partition Protein A/Flip plus assembly protein (ATPase CpaE domain)	F	2653	3288	635	116/213 (54%)	Laribacter hongkongensis	ABC70161
						Average Size =	1148.0		
						Standard Deviation =	357.3		

Transposase and Resolvase									
Contig	ORF Desig.	Gene Identification and Determined/Predicted Function	Strand	Start	Stop	Size	% Identity	Source	ID # of Closest Relative
1	7	Transposase (Tn3 Family)	F	12241	15177	2936	927/979 (94%)	Plasmid pLB1	YP_740316
7	177	Recombinase (Resolvase) - like	F	587	1147	560	146/183 (79%)	Comamonas testosteroni KF-1	ZP_01518522
29	46	Transposase	F	11	1612	1601	495/501 (98%)	Sphingomonas sp. SKA58	ZP_01301684
29	47	Transposase	F	2105	3625	1520	283/494 (57%)	Aurantimonas sp. S185-9A1	ZP_01227137
29	48	IS21 family Transposase/ATP-binding Protein	F	3385	4437	1052	177/245 (72%)	Sphingomonas sp. KA1	YP_718203
29	52	Tn3 Transposase	C	5530	8439	2909	969/969 (100%)	Sphingomonas wittichii RW1	ZP_01610104
29	53	Resolvase, N-terminal domain	C	4584	5501	917	288/288 (100%)	Sphingomonas wittichii RW1	ZP_01610103
33	75	Transposase (Class 7)	C	430	3522	3092	396/985 (40%)	Pseudomonas aeruginosa	YP_758661
38	80	Transposase (Class 7)	C	123	1964	1841	162/512 (31%)	Pseudomonas putida	YP_709357
56	103	Transposon Resolvase	F	1688	2545	857	125/189 (66%)	Burkholderia thailandensis E264	YP_440197
62	125	Phage Integrase	C	4678	6843	2165	412/704 (58%)	Mesorhizobium sp. BNC1	YP_665854
62	123	Phage-DNA Integrase-like SAM-like	C	8624	9979	1355	292/460 (63%)	Mesorhizobium sp. BNC1	YP_665852
65	130	Resolvase/Recombinase/Invertase	F	309	887	578	90/187 (48%)	Stappia aggregata IAM 12614	ZP_01550083
66	140	Putative Transposase	C	345	1445	1100	332/365 (90%)	Rhizobium leguminosarum bv. viciae 3841	YP_764867
66	139	Phage Integrase/Site-specific Recombinase	C	1506	2468	962	282/320 (88%)	Sphingomonas wittichii RW1	ZP_01609885
74	148	Integrase/Recombinase (Phage Integrase and SAM Domains)	C	269	1216	947	156/317 (49%)	Sphingomonas wittichii RW1	ZP_01609828
						Average Size = 1524.5			
						Standard Deviation = 840.8			
Regulation									
Contig	ORF Desig.	Gene Identification and Determined/Predicted Function	Strand	Start	Stop	Size	% Identity	Source	ID # of Closest Relative
9	13	RNA polymerase sigma-70 factor domain	C	1073	1831	758	26/71 (36%)	Bacillus weihenstephanensis KBAB4	ZP_01187840
32	59	Putative Transcriptional Repressor	F	51	686	635	74/183 (40%)	Sphingomonas sp. SKA58	ZP_01304563
33	72	Heavy Metal Resistance Transcriptional Regulator	F	5710	6183	473	81/124 (65%)	Sphingopyxis alaskensis RB2256	YP_617724
						Average Size = 622.0			
						Standard Deviation = 142.9			

Metabolism										
Contig	ORF Desig.	Gene Identification and Determined/Predicted Function	Strand	Start	Stop	Size	% Identity	Source	ID # of Closest Relative	
1	2	Type II Secretion System Protein E	C	3748	4771	1023	199/330 (60%)	Sphingomonas wittichii RW1	ZP_01610041	
9	14	Sulphate Transporter	C	4335	5852	1517	372/494 (75%)	Sphingopyxis alaskensis RB2256	YP_617822	
14	20	Phosphoadenosine Phosphosulfate Reductase	C	506	1243	737	102/244 (41%)	Polaromonas naphthalenivorans C.J2	YP_973329	
26	41	TerC Integral Membrane Protein (Tellurium Resistance)	C	218	718	500	121/165 (73%)	Sphingopyxis alaskensis RB2256	YP_615993	
27	42	Chromate Transporter	F	30	920	890	183/294 (62%)	Mesorhizobium sp. BNC1	YP_672723	
27	43	Manganese and Iron Superoxide Dismutase	F	932	1663	731	153/238 (64%)	Sphingopyxis alaskensis RB2256	YP_616826	
31	58	Sulfatase	C	1	792	791	80/251 (31%)	Pseudoalteromonas atlantica T6c	YP_660469	
32	65	Acetyltransferase	C	6095	6730	635	104/171 (60%)	Stigmatella aurantiaca DW4/3-1	ZP_01461054	
32	66	Fusarc Acid Resistance Protein (integral membrane protein)	C	3584	5560	1976	191/668 (28%)	Brucella melitensis biovar Abortus 2308	YP_418673	
32	67	Secretion Protein HlyD (includes Multidrug Resistance Efflux Pump)	C	2466	3374	908	136/286 (47%)	Brucella suis 1330	NP_699935	
32	68	RND Efflux System, Outer Membrane Lipoprotein, NodT	C	1036	2469	1433	203/455 (44%)	Ralstonia metallidurans CH34	YP_696921	
33	73	unknown (Mercury Transporter Domain)	C	5289	5828	539	74/129 (57%)	Sphingopyxis alaskensis RB2256	YP_617723	
33	74	Mercury Reductase MerA	C	3491	4909	1418	297/462 (64%)	Sphingopyxis alaskensis RB2256	YP_617721	
34	77	DdmC Dicamba Monooxygenase	C	515	1534	1019	339/339 (100%)	Stenotrophomonas maltophilia	AAV53699	
42	82	Glycosyltransferase (low homology)	C	24	740	716	92/212 (43%)	Burkholderia mallei ATCC 23344	YP_103663	
56	106	Hypothetical Protein Involved in Quinolone Resistance	C	352	1959	1607	194/471 (41%)	Klebsiella pneumoniae pK245	ABG56873	
59	111	Putative Monooxygenase (Flavin-containing)	F	234	1667	1433	192/335 (57%)	Pseudomonas putida KT2440	NP_744078	
60	113	Protein Disulfide Isomerase	F	5	937	932	145/238 (60%)	Sphingomonas sp. KA1	YP_718066	
61	115	Phosphotricin (glufosinate) Acetyltransferase	C	333	881	548	88/174 (50%)	Mesorhizobium loti MAFF303099	NP_105235	
66	138	Putative Outer Membrane Phospholipase A	C	8034	9404	1370	201/391 (51%)	Sphingomonas sp. KA1	YP_718007	
67	142	Asparagine Synthase	C	145	1041	896	115/300 (38%)	Sphingopyxis alaskensis RB2256	YP_617573	
81	156	hypothetical protein pLB1_p23 (ABC-type Multidrug/Bacteriocin Transporter Domains)	F	3656	6460	2804	627/936 (66%)	Plasmid pLB1	YP_740321	
84	166	Benzyl Alcohol Dehydrogenase (Zn Dependent)	F	74	1153	1079	209/358 (58%)	Pseudomonas sp. ND6	NP_943147	
85	171	Chromate Resistance Exported Protein (chrB)	C	261	1211	950	137/318 (43%)	Desulfotribrio vulgaris	YP_009134	
						Average Size =	1102.2			
						Standard Deviation =	527.1			

Other									
Contig	ORF Desig.	Gene Identification and Determined/Predicted Function	Strand	Start	Stop	Size	% Identity	Source	ID # of Closest Relative
18	25	Nuclease (pLB1)	F	12236	13036	800	156/163 (95%)	Plasmid pLB1	YP_740301
69	146	Excalibur Nuclease	C	1393	2082	689	102/206 (49%)	Sphingopyxis alaskensis RB2256	YP_615191
81	155	Type III Restriction-Modification Enzyme, helicase subunit (pLB1)	F	711	3647	2936	841/977 (86%)	Plasmid pLB1	YP_740322
				Average Size =		1475.0			
				Standard Deviation =		1266.5			
Unknown									
Contig	ORF Desig.	Gene Identification and Determined/Predicted Function	Strand	Start	Stop	Size	% Identity	Source	ID # of Closest Relative
4	173	hypothetical protein SwitDRAFT_1073	F	786	1568	782	74/165 (44%)	Sphingomonas wittichii RW1	ZP_01609744
44	87	unknown (transcriptional Regulator Domains)	C	321	1928	1607			
54	101	unknown (DNA Integrase Domain)	C	6	1040	1034			
76	150	unknown (Carboxypeptidase Domain)	C	536	1447	911			
85	170	Similar to LpqO Lipoprotein	F	1380	2270	890	166/289 (57%)	Bradyrhizobium japonicum USDA 110	NP_774707
9	12	unknown (phenyl acetyl CoA hydratase domain similarity)	F	1338	5399	4061			
				Average Size =		1547.5			
				Standard Deviation =		326.8			

Initial functional analysis of the predicted ORFs gleaned from the 87 contigs identified a single dicamba monooxygenase gene in the sequence data analyzed. Gene copy number is an interesting consideration in the analysis of the dicamba mineralization pathway in *P. maltophilia*, therefore, further analysis was needed to confirm the single gene copy number observation. Multiple sequence alignment was performed between each contig and the known DNA sequence of the (1020 bp) dicamba monooxygenase gene (Sequence ID # AY786443). Both forward and reverse-sequence MSA was performed with the Vector NTI AlignX tool, using the ClustalW algorithm, which found no further evidence of multiple copies of DMO in the sequence data. Finally, predicted ORF fragments from the termini of each contig were forward and reverse-sequence aligned against the DMO gene sequence, but with no significant homology found.

Due to the fact that the megaplastids of this organism were sequenced in total, rather than individually, it is difficult to make statements about the nature, purpose, and features of each individual plasmid. However, certain generalizations can be made regarding features of the *P. maltophilia* megaplastids as a whole.

One of the larger contigs (Contig 18, 13,527 bp) possessed nearly all the genes generally responsible for the Type IV secretory system/conjugal transfer characteristic of conjugative plasmids. This cluster of conjugative genes was complemented by the presence of other genetic participants in the conjugal transfer pathway in other areas of the analyzed sequence. In addition, this sequence segment showed 96% homology over 94% coverage of the conjugal transfer region of another biodegradative plasmid, pLB1, which is capable of mediating γ -Hexachlorocyclohexane degradation in the soil bacterium *Sphingobium japonicum* UT26 (Miyazaki et al., 2006).

In addition to the expected replicative, plasmid maintenance, conjugative, and transposon elements on the *P. maltophilia* plasmids, this strain possesses a number of diverse metabolic genes likely correlated to the diverse metabolic repertoire of the *Pseudomonas* genus. Included are a number of transport-assist proteins, including a member of the HlyD family (Sequence ID # NP_699935), a Resistance Nodulation Cell Division (RND) efflux protein (Sequence ID # YP_586921), and an ABC-type multidrug/bacteriocin transporter (Sequence ID # YP_740321). Also included are chromate and tellurium resistance proteins, a fusaric acid resistance protein, a glufosinate acetyltransferase (Sequence ID # NP_105235), and a notable benzyl alcohol dehydrogenase (Sequence ID # NP_943147). Additional metabolic genes are listed above (Figure 7).

D. DISCUSSION AND IMPLICATIONS:

D.1 Molecular characterization of P. maltophilia, strain DI-6 megaplasmids

The study of the molecular structure of the extrachromosomal DNA of *P. maltophilia* is particularly relevant to the intent of this series of experiments, as the dicamba monooxygenase gene has been localized by southern blot to this organism's megaplasmids. (Chrastil, 2000) Therefore, better understanding of the megaplasmid structure of this dicamba-degrading organism is a logical starting point for further analysis of the dicamba mineralization pathway.

Determination that the megaplasmids of *P. maltophilia* are of eight distinct sizes, ranging from 26 kB to 181 kB (Figure 6) suggests a large and important reservoir of functional, important metabolic and resistance genes. This idea is first of all supported

by the vast metabolic diversity of the *Pseudomonas* genus, and the understanding that lateral conjugative gene transfer is a particularly important means by which novel metabolic genes are shared and transferred between species. The existence of plasmids in an organism, although sometimes transient in the case of lysogenic or transposon systems, is an essential feature of conjugative gene transfer, and *P. maltophilia* certainly exhibits this phenotype. In addition, by simple logic, the large amount of genetic information present in *P. maltophilia* in megaplasmid form suggests a large number of potential genes maintained extra-chromosomally. Hence, the large megaplasmid complement of this organism can be speculatively correlated to its high level of metabolic diversity.

A further consideration of the number and size of these megaplasmids is the localization of whole or partial degradative pathways to each megaplasmid. In the case of this study, one major goal is the determination whether the dicamba mineralization pathway in *P. maltophilia* is localized to one or multiple megaplasmids in the extrachromosomal genome. If all genes necessary for dicamba metabolism were located on a single megaplasmid, the study and characterization of the degradative pathway would be much simplified. The ability to separate and size the *P. maltophilia* plasmids provides a starting point for the in-depth study of individual plasmids.

D.2 Demonstration of the ability to isolate, separate, “tag” with antibiotic resistance marker, and transform P. maltophilia megaplasmids

Due to the fact that the complete degradative pathway for dicamba mineralization may or may not be localized to a single megaplasmid, there exists no suitable selectable

marker inherent to each plasmid which would allow selection after transformation with a single sized plasmid. Therefore, unless the ability to metabolize dicamba is conferred by a single plasmid, the study of single megaplasms from *P. maltophilia* is very difficult. Even in the case of a single dicamba-degradative megaplasms, that particular genetic element would be the only one possessing a suitable selection marker. Therefore, using a transposon-based antibiotic “tagging” system supports a two-fold benefit. First, the presence of an engineered Kanamycin resistance cassette on each megaplasms would allow transformation and selection of positive single-megaplasms transformants, thus facilitating their study. Second, the transposon strategy used also included an inducible origin of replication, which should allow induction of the single-megaplasms transformation events to higher copy number. Thus, the inducible origin of replication and Kanamycin resistance combination should facilitate further research by allowing overproduction of single megaplasms in their *E. coli* host. Upon DNA isolation, the availability of relatively pure samples of each tagged megaplasms would facilitate potential molecular or sequencing/bioinformatic characterization to a much greater extent than in the case of the “natural” megaplasms.

D.3 Initial bioinformatics analysis of the P. maltophilia extrachromosomal gene complement

Perhaps the most immediately relevant result of the experiments described above is the increased knowledge of the extrachromosomal genetic complement of this particular organism, a known bioremediator of xenobiotic compounds. Information from the sequencing, ORF prediction, and BLAST analyses provided immediate data regarding

the dicamba degradation machinery of *P. maltophilia*, strain DI-6, but a consideration of the megaplasmid sequence data as a whole can be used in analysis of the machinery responsible for the vast metabolic diversity of the *Pseudomonas* genus.

It is known that the megaplasms of *P. maltophilia* are sufficient to confer the ability to metabolize dicamba for use as a sole carbon source in host organisms. Therefore, contained within the sequence data for the total megaplasms of this organism is all the genetic information necessary to construct and understand the dicamba degradative pathway. Although the process of elucidating said pathway is not easy nor straightforward, the first step was taken toward this goal by sequencing and performing an initial characterization of the extrachromosomal genetic material. Now the analysis and elucidation of this pathway is a matter of, starting with the dicamba monooxygenase (Contig 34, ORF 77, *see above*), identifying genes which give functional hints as to the stepwise progression of the degradative pathway until it feeds into the conserved metabolic machinery encoding TCA cycle enzymes of *P. maltophilia*.

In addition to laying the groundwork for further study of the dicamba degradation pathway through prediction of gene identification and function, the bioinformatic analysis performed here provided other evidence in support of other hypothesis regarding the phenotype of dicamba degradation. For example, one can hypothesize as to the origin of the metabolic machinery responsible for dicamba utilization in nature. The presence of a complete Type IV conjugative transfer system in the megaplasmid complement of *P. maltophilia* suggests the potential for dicamba utilization as spreading through the soil or water microbial community by interspecies gene transfer. The metabolic machinery responsible for utilization of this herbicide as a sole carbon source was most likely not

initially evolved for this strict purpose, thus, both the original function and origin of the responsible genetic elements remain uninvestigated at present. However, the presence of intact conjugative machinery, along with a wide variety of genes derived from known soil- and water-inhabiting bacteria suggests that the genes responsible for dicamba utilization may have originated in a particularly adaptable environmental microorganism and spread throughout the immediately-resident bacterial population. Gene sharing between genera as diverse as *Agrobacterium*, *Bradyrhizobium*, and *Xanthomonas* (of which *P. maltophilia* was formerly a member) exists and is well-documented (Don *et al.*, 1981; Kinkle *et al.*, 1993; Daane *et al.*, 1996; Phillips *et al.*, 2004). Horizontal transfer between species like these in nature likely created the population of xenobiotic bioremediator organisms from which *P. maltophilia*, DI-6 was isolated.

D.4 Proposal for and implications of future research

The molecular and bioinformatic characterizations described above have laid only a most basic groundwork for understanding the biodegradative ability of *P. maltophilia* in the context of specific utilization of the herbicide dicamba as a sole carbon source. In hopes of elucidating this metabolic pathway for the benefit of scientific advancement and development of improved transgenic crops, much more research on the subject is warranted.

An intriguing direction of research could be investigating the origin of the degradation pathway responsible for dicamba metabolism in *P. maltophilia*, DI-6. Perhaps a first step might be to use the sequence data already obtained to compare the conjugative machinery contained on megaplasmids to other known conjugative transfer

systems from various organisms. High homology to a particular species may suggest relationships between two similar species, or may give important hints as to the dynamics of lateral gene transfer between divergent species. Another consideration related to the origin of this degradative pathway would be determining the original substrate of the dicamba monooxygenase enzyme. Work is already underway in the Weeks Lab to screen for other chemicals able to be chemically modified by DMO, but further work would certainly provide unique insights. As dicamba was a relatively recent environmental contaminant (introduced around 1965), there was assuredly another natural substrate for DMO long before it ever acted on dicamba. The discovery of such a substrate might provide insight as to the environmental origin and construction of this unique degradative pathway.

Elucidation of the dicamba-degrading metabolic pathway in *P. maltophilia* will require extensive study of the sequence data obtained previously and initially described in these experiments. The first step will be identifying genes of known function and genes possessing known domains which might participate in the degradative pathway. Sequence analysis tools similar to those used in this study will need to be used to predict gene function or conserved domains, and identification will be qualified by homology information and the presence of gene-associated sequences (e.g. Shine-delgarno, etc.). In conjunction with the wealth of existing data regarding the metabolism of aromatic xenobiotic compounds, specific information derived from the sequence data for *P. maltophilia* megaplasmids should give hints as to the enzymes and associated chemical transformations required to feed dicamba into the conserved host metabolic machinery. Once potential degradative genes have been identified, they will be cloned out of their

individual host megaplasmid of origin, over-expressed in a suitable host, and assayed for potential activity in the pathway.

Additional research is necessary to obtain the complete megaplasmid DNA sequence data for *P. maltophilia*. Complete sequence data and analysis will allow for construction of individual plasmid vectors with suspected degradative genes. With clones of potential dicamba-associated metabolic genes in hand and assayed for activity, the suspected degradative genes can be clustered together in combinatorial, novel individual constructs. Such constructs can be placed into host cells by transformation to potentially re-create the dicamba degradative pathway *in vivo* and, thus, characterize all participating enzymes, cofactors, etc.

With the dicamba metabolic pathway elucidated and characterized, researchers will have the tools to work toward optimizing certain features of the pathway through substituting individual genes to create novel combinations and assaying host cells containing the constructed vectors for improved degradative ability. Experiments of this type will have immediate value within this project when used to optimize current systems and work toward creating more efficient transgenic broadleaf crops. The definitive goal in this effort will be to affect a more positive herbicide-treatment strategy by giving agricultural producers more diverse and more effective options for transgenic crop systems.

Ultimately, efforts directed at optimizing degradative machinery systems for engineering crop plants and biodegradative species will be supplemented by knowledge gained from degradative pathway research. Through further understanding of the processes and pathways by which biological degradation occurs, scientists can contribute

to both conscientious and proactive strategies for decreasing environmental contamination with xenobiotic compounds. Increased options for the end-users of herbicides and other chemical treatments, particularly in agricultural applications, will expectedly stimulate more directed, selective, and environmentally-conscious applied use of such potential contaminant compounds. In addition, research directed at understanding novel degradative pathways holds the potential to extend applications beyond the current focus of creating improved resistant crop systems. Complementary to the benefits afforded by stimulating superior applied use of herbicides and chemical treatments is the potential for targeting the opposite temporal side of environmental contamination. With knowledge gained from degradative pathway research, application will increase scientists' ability to effectively engineer microorganism, fungi, or plant bioremediators capable of treating and clearing environmental contaminants after intentional or unintentional contamination has occurred. The potential for improving such a diverse and adaptable tool has very real implications for the future of managing environmental disruption and damage caused by contamination with xenobiotic compounds.

E. REFERENCES

- Alexander, M. (1981). "Biodegradation of chemicals of environmental concern." Science **211**(4478): 132-8.
- Caux, P. Y., R. A. Kent, et al. (1993). "Environmental fate and effects of dicamba: a Canadian perspective." Rev Environ Contam Toxicol **133**: 1-58.
- Chakrabarty, A. M. (1972). "Genetic Basis of the Biodegradation of Salicylate in Pseudomonas." J Bacteriol **112**: 815-823.
- Chakrabarty, A. M. (1976). "Plasmids in Pseudomonas." Annu Rev Genet **10**: 7-30.
- Chakraborty, S., M. Behrens, et al. (2005). "A three-component dicamba O-demethylase from Pseudomonas maltophilia, strain DI-6: purification and characterization." Arch Biochem Biophys **437**(1): 20-8.
- Chrastil, B. M. (2000). The Role of Plasmids in the Degradation of the Herbicide, dicamba, by Pseudomonas maltophilia, strain DI-6. Department of Biochemistry. Lincoln, NE, University of Nebraska - Lincoln. **B.S.**
- Cork, D. J. and A. Khalil (1995). "Detection, isolation, and stability of megaplasmid-encoded chloroaromatic herbicide-degrading genes within Pseudomonas species." Adv Appl Microbiol **40**: 289-321.
- Cork, D. J. and J. P. Krueger (1991). "Microbial transformations of herbicides and pesticides." Adv Appl Microbiol **36**: 1-66.
- Daane, L. L., J. A. Molina, et al. (1996). "Influence of earthworm activity on gene transfer from Pseudomonas fluorescens to indigenous soil bacteria." Appl Environ Microbiol **62**(2): 515-21.
- Don, R. H. and J. M. Pemberton (1981). "Properties of six pesticide degradation plasmids isolated from Alcaligenes paradoxus and Alcaligenes eutrophus." J Bacteriol **145**(2): 681-6.
- Duke, S. O. (2005). "Taking stock of herbicide-resistant crops ten years after introduction." Pest Manag Sci **61**(3): 211-8.
- Herman, P. L., M. Behrens, et al. (2005). "A three-component dicamba O-demethylase from Pseudomonas maltophilia, strain DI-6: gene isolation, characterization, and heterologous expression." J Biol Chem **280**(26): 24759-67.

- Holloway, B. W. (1975). Genetic Organization of Pseudomonas. Genetics and Biochemistry of Pseudomonas. P. H. Clarke, Richmond, M. H. New York, John Wiley & Sons.
- Khalil, A. and D. J. Cork (2000). "Plasmid-mediated catabolism of dicamba by Pseudomonas species strain PXM." Microbios **102**(403): 183-91.
- Kinkle, B. K., M. J. Sadowsky, et al. (1993). "Plasmids pJP4 and r68.45 Can Be Transferred between Populations of Bradyrhizobia in Nonsterile Soil." Appl Environ Microbiol **59**(6): 1762-1766.
- Krueger, J. P., Butz, R. G., Atallah, Y. H., Cork, D. J. (1989). "Isolation and Identification of Microorganisms for the Degradation of dicamba." J Agric Food Chem **37**: 534-538.
- Krueger, J. P., Butz, R. G., Cork, D. J. (1991). "Use of dicamba-Degrading Microorganisms to Protect dicamba Susceptible Plant Species." J Agric Food Chem **39**: 1000-1003.
- Miyazaki, R., Y. Sato, et al. (2006). "Complete nucleotide sequence of an exogenously isolated plasmid, pLB1, involved in gamma-hexachlorocyclohexane degradation." Appl Environ Microbiol **72**(11): 6923-33.
- Newton, M., Roberts, F., Allen, A., Kelsas, B., White, D., Boyd, P. (1990). "Deposition and Dissipation of Three Herbicides in Foliage, Litter, and Soil of Brushfields of Southwest Oregon." J Agric Food Chem **38**(574-583).
- NPIC (2000). Glyphosate: National Pesticide Information Center Technical Fact Sheet. E. P. Agency.
- NPIC (2002). dicamba: National Pesticide Information Center Technical Fact Sheet. E. P. Agency.
- Owen, M. D. and I. A. Zelaya (2005). "Herbicide-resistant crops and weed resistance to herbicides." Pest Manag Sci **61**(3): 301-11.
- Palleroni, N. J. (1975). General Properties and Taxonomy of the Genus Pseudomonas. Genetics and Biochemistry of Pseudomonas. P. H. Clarke, Richmond, M. H. New York, John Wiley & Sons.
- Phillips, G. and B. E. Funnell (2004). Plasmid biology. Washington, D.C., ASM Press.
- Wang, X., B. Li, et al. (1997). "A Three-Component Enzyme System Catalyzes the O Demethylation of the Herbicide dicamba in Pseudomonas maltophilia DI-6." Appl Environ Microbiol **63**(4): 1623-1626.

Yang, J., X. Z. Wang, et al. (1994). "Analysis of dicamba degradation by *Pseudomonas maltophilia* using high-performance capillary electrophoresis." Anal Biochem **219**(1): 37-42.