Cleaner Water: Genomic Analysis of the Homogentisate Catabolic Pathway in *Pseudomonas putida* F1 for Bioremediation of Aromatic Hydrocarbons

Siddarth Eswarachari and Moira Southern
Breck School

Carcinogenic aromatic hydrocarbons are being released into water supplies in increasing concentrations each year. One technique to bioremediate these carcinogens is to bioengineer bacteria to catabolize (degrade) the aromatic hydrocarbons. The bacteria *Pseudomonas putida* F1 shows promise in bioremediation of aromatic hydrocarbons; however, before *P. putida* F1 can be bioengineered to catabolize aromatic hydrocarbons, ways in which *P. putida* F1 catabolize "more appetizing" aromatic nutrients, such as homogentisate, must be found. We focused on the homogentisate catabolic pathway because this pathway is the central route into which the aromatic amino acids L-phenylalanine and L-tyrosine funnel in the bacteria *P. putida* U and *P. putida* KT2440, which are closely related to *P. putida* F1. Using the online databases FASTA in GenBank, BLAST, T-COFFEE, and WebLogos, we compared protein sequences for *P. putida* F1 to other closely related bacterial strains that included *P. putida* KT2440, *P. aeruginosa* PAO1, *P. fluorescens* Pf0-1, and *Comamonas testosterone* CNB-1 to identify substrates and enzymes in the pathway by which homogentisate is catabolized by *P. putida* F1. We then used TMHMM and PSORTb to determine the transporter and regulatory proteins for the homogentisate pathway in *P. putida* F1. E-values were used to validate best-fit protein sequences, showing 100% accuracy for choices of enzymes identified for the homogentisate pathway in *P. putida F1*. Consequently, our results were published on GENI-ACT, which is a database that contains genomic information on 2775 bacteria. Eswarachari S, Southern M. Cleaner water: Genomic analysis of the homogentisate catabolic pathway in *Pseudomonas putida* F1 for bioremediation of aromatic hydrocarbons. *Minnesota Academy of Science Journal of Student Research* 2015; **3:**19-25.

Keywords: *Pseudomonas putida* F1, homogentisate, catabolic pathway, bioinformatics, aromatic hydrocarbons

Advisor: Ms. L. Fruen
The University of St. Thomas funded the project. There was no grant associated with this project.
This paper received a 2015 Minnesota Academy of Science STEM Communicator Award. The judging process for this award satisfied the function of manuscript review:
Abbreviations: BLAST, Basic local alignment search tool; EPA, Environmental protection agency; E-value, Expected value; GENI-ACT, Genomics education national initiative- annotation collaboration toolkit; IMG JGI, Integrated microbial genomes joint genome institute; NCBI, National center for biotechnology information; PDB, Protein data bank; T-COFFEE, Tree-based consistency objective function for alignment evaluation; TMHMM, Transmembrane helices hidden markov models

INTRODUCTION

As human industrialization advances on a global scale, greater concentrations of carcinogenic compounds are being released into water supplies each year[1]. Many of these pollutants consist of aromatic hydrocarbons, which are chemicals that enter the environment through oil spills and through incomplete combustion of coal, wood, and other organic substances[2]. Since aromatic hydrocarbons can cause cancer and other chronic illnesses, cleaning up these compounds is crucial[1]. One cleanup technique is to use bacteria that catabolize (degrade)

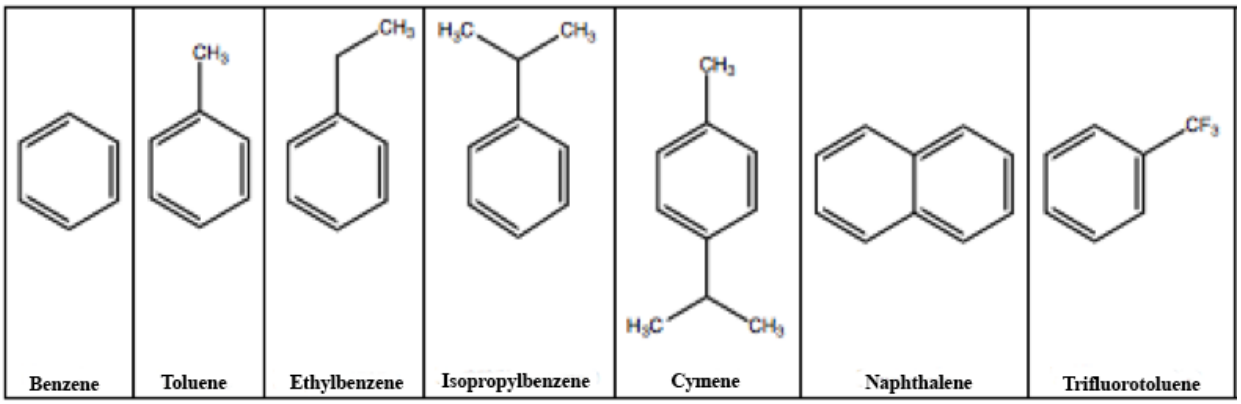

Figure 1. Carcinogenic aromatic hydrocarbons to which *P. putida* shows chemotaxis[2,3]

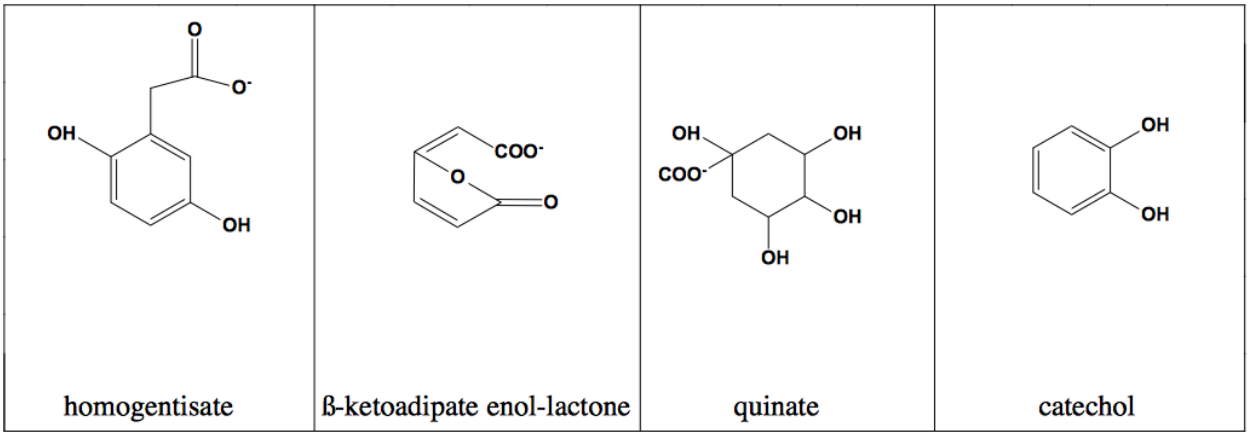

Figure 2. Aromatic nutrients that *P. putida* catabolize for energy through the Krebs Cycle, which produces energy for the bacteria

aromatic hydrocarbons - a process by which hydrocarbons are broken down by the bacteria into simpler compounds to produce energy. One bacteria that shows promise in bioremediation of aromatic hydrocarbons is *Pseudomonas putida* F1 - a nonpathogenic bacteria that occurs naturally in soil and aquatic environments.

P. putida F1 shows promise in bioremediation to clean up aromatic hydrocarbon pollutants for two reasons. First, studies have shown that *P. putida* F1 is able to respond to chemical signals from aromatic nutrients in its environment, a process known as chemotaxis[2]. Second, work by Parales *et al.*[2] and Harwood, Rivelle, & Ornston[3] showed that *P. putida* F1 is chemotatic towards the following carcinogenic aromatic hydrocarbons: benzene, toluene, ethylbenzene, isopropylbenzene, cymene, naphthalene, and trifluorotoluene (Figure 1).

However, before *P. putida* F1 can be use in bioremediation efforts to remove carcinogenic aromatic hydrocarbons from the environment, it is important to determine ways in which *P. putida* F1 catabolize "more appetizing" aromatic nutrients, such as homogentisate, quinate, B-ketoapdipate enol-lactone, and catechol[4] (Figure 2). This is because *P. putida* F1 bacteria are naturally attracted to these more aromatic nutrients, so the bacteria will not catabolize aromatic hydrocarbons until they are bioengineered to prefer the aromatic hydrocarbons. The purpose of our study was to identify enzymes and substrates involved in the pathway by which *P. putida* F1 metabolizes the aromatic nutrient homogentisate, with the long-term goal of knocking out genes that encode receptor proteins responsible for chemotaxis of *P. putida* F1 to homogentisate in order to bioengineer the bacteria for bioremediation

of aromatic hydrocarbons. We were inspired to use data-mining techniques instead of laboratory workups because the large magnitude of genomic data available on genomic databases enabled us to work independently to produce results that are highly accurate.

Previous work has shown that there are four known catabolic pathways for aromatic compounds in *P. putida KT2440* and *P. putida U*, which are *Pseudomonas* strains that are closely related to *P. putida* F1. Previous research has shown that the catabolic pathways in *P. putida KT2440* and *P. putida U* are the phenylacetate, quinate, catechol, and the homogentisate pathways[4]. We focused on the homogentisate catabolic pathway because this pathway is the central route into which the aromatic amino acids L-phenylalanine and L-tyrosine funnel from a peripheral pathway in the bacteria *P. putida* U and *P. putida* KT2440, which are closely related to *P. putida* F1[5] (Figure 3).

Our specific goals were to:

1. Use nucleotide accession numbers of genes that encode enzymes in the homogentisate pathway of *P. putida* U;
2. Match the protein sequences of enzymes in the homogentisate pathway of *P. putida* U to the genome of *P. putida* F1 to locate best-fit protein sequences of enzymes in the genome of *P. putida* F1;
3. Validate best-fit protein sequences by matching the protein sequences of enzymes identified in the genome of *P. putida* F1 with best-fit protein sequences of enzymes in the genomes of other closely related bacterial strains, including *P. putida* KT2440, P. *aeruginosa* PAO1, *P. fluorescens* Pf0-1, and *Comamonas testosterone* CNB-1;
4. Find the amino acid sequences for each enzyme identified in homogentisate pathway of *P. putida* F1.
5. Match identified enzymes with substrates involved in the homogentisate pathway of *P. putida* F1;
6. Identify the transporter and regulators proteins in the homogentisate pathway of *P. putida* F1; and
7. Publish to GENI-ACT, which is a database that contains genomic information on 2775 bacteria, for the enzymes involved in the homogentisate pathway for *P. putida* F1.

Because the bacteria *P. putida* U, *P. putida* KT2440, *P. aeruginosa* PAO1, *P. fluorescens* Pf0-1, and *Comamonas testosterone* CNB-1 are closely related to *P. putida* F1, we hypothesized that protein sequences of enzymes that regulate the homogentisate pathway in the related bacteria could be used to identify enzymes involved in the homogentisate pathway of *P. putida* F1.

.

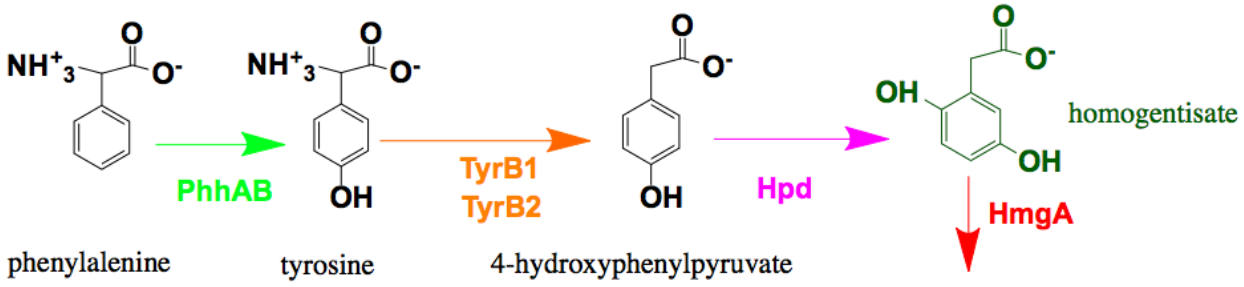

Figure 3. The peripheral phenylalanine pathway in *P. putida* U and *P. putida* KT2440 that funnels into the homogentisate pathway[5]

MATERIALS AND METHODS

Software programs and online databases that were used in this study included: FASTA in GenBank, Basic Local Alignment Search Tool (BLAST), Tree-based Consistency Objective Function for alignment Evaluation (T-COFFEE), National Center for Biotechnology Information (NCBI), Integrated Microbial Genomes Joint Genome Institute (IMG JGI) gene-search database, Protein Data Bank (PDB), WebLogos, prediction using Hidden Markov Models (TMHMM), the Protein Sorting Database version b (PSORTb), and the Genomics Education National Initiative- Annotation Collaboration Toolkit (GENI-ACT)

Determining protein sequences for P. putida U: First, nucleotide accession numbers reported by Arias-Barrau *et al*.[5] for genes that encode enzymes in the homogentisate pathway of the closely related bacteria *P. putida* U were entered into GenBank to which the study by Arias-Barrau *et al*. had submitted the accession numbers[5]. Then, FASTA in GenBank was used to access the protein sequences of the enzymes involved in the homogentisate pathway of *P. putida* U.

Next, the genome for *P. putida* F1 was accessed in BLAST, and then, BLAST was used to compare the protein sequences for *P. putida* U to the genome of *P. putida* F1. E-values on the list of possible protein sequence matches produced by BLAST were evaluated to find the best-fit protein sequence (enzyme) matches, where an E-value of 0.0 indicates 100% match.

To validate the protein sequences of the enzymes identified for *P. putida* F1, the BLAST method was repeated to match protein sequences in the genome of *P. putida* U with best-fit protein sequences in the genomes of each of the following closely related bacteria: *P. putida* KT2440, *P. aeruginosa* PAO1, *P. fluorescens* Pf0-1, and *Comamonas testosterone* CNB-1.

E-values were used to select all best-fit sequences. Then, T-COFFEE was used to compare all best-fit protein sequences among each bacteria to those best-fit protein sequences found for *P. putida* F1.

The outputted best-fit sequence alignment identified using T-COFFEE for each enzyme in the related bacteria as well as in *P. putida* F1 were entered into WebLogos. WebLogos was run to match amino acids in the enzymes of related bacteria to *P. putida* F1 to identify fits that aligned most closely. Identified enzymes were then matched to substrates in the homogentisate pathway for *P. putida* U identified by Arias-Barrau[5] to identify substrates in the homogentisate pathway of *P. putida* F1.

Determining transport protein in P. putida F1: After an extensive literature search, only one study was found that gave information about a transport protein for homogentisate. This study, done by Jiménez *et al*.[4], suggested that phhT serves as the transport protein for homogentisate in the bacteria *P. aeruginosa* PAO1. To determine if phhT is also the transport protein in *P. putida* F1, the accession number of phhT reported by Jiménez *et al*.[4] was entered into FASTA in GenBank to access the protein sequence of phhT in *P. aeruginosa* PAO1. The protein sequence for phhT was then entered into both PSORTb and TMHMM to run analyses that showed there is a 100% chance that the function of phhT is to transport homogentisate across the cytoplasmic membrane, which validated that phhT is a transport protein. Next, the protein sequence for phhT was entered into BLAST to compare it to the genome sequence of *P. putida* F1 and to determine the E-value for matches in *P. putida* F1.

Determining regulatory protein in P. putida F1**:** To identify the regulatory protein in *P. putida* F1**,** the nucleotide accession number for the regulatory protein hmgR, which Arias-Barrau *et al*.[5] showed is involved in the homogentisate pathway of *P. putida* U, was entered into FASTA in Genbank to access the protein sequence of hmgR. Next, BLAST was used to compare the protein sequence of hmgR to the genomes of *P. putida* KT2440, *P. aeruginosa* PAO1, *P. fluorescens* Pf0-1, and *Comamonas testosterone* CNB-1, as well as to the genome of *P. putida* F1.

The E-value was used to identify the best-fit protein sequence in each bacteria. Then, T-COFFEE was used to compare the protein sequence of hmgR to best-fit protein sequences in *P. putida* KT2440, *P.*

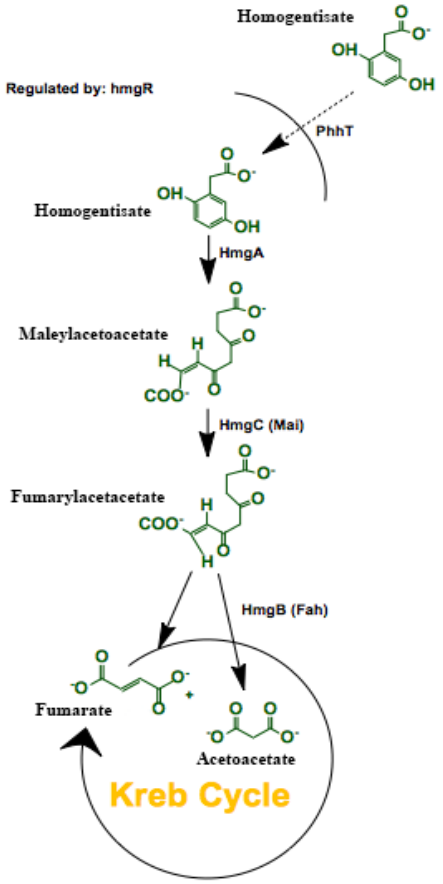

Figure 4. Substrates and enzymes found in the homogentisate of *P. putida* F1

aeruginosa PAO1, *P. fluorescens* Pf0-1, and *Comamonas testosterone* CNB-1. The outputted best-fit sequence alignment was entered into WebLogos to determine amino acid sequences of the regulatory protein in *P. putida* F1.

Updating GENI-ACT: First, locus tags located at the beginning of each protein sequence of each enzyme involved in the homogentisate pathway of *P. putida* F1 were inputted into GENI-ACT. GENI-ACT outputted locations on the *P. putida* F1 genome of each locus tag. Then, each DNA sequence, each protein sequence, a screenshot of BLAST results that included E-values, T-COFFEE results with E-values, and WebLogo information were compiled on a GENI-ACT information page for each enzyme in the homogentisate pathway of *P. putida* F1.

RESULTS

Figure 4 shows the substrates and the enzymes found in the homogentisate pathway of *P. putida* F1. The transporter protein phhT transports homogentisate into *P. putdia* F1 while hmgR regulates the enzymes involved in the homogentisate pathway. The regulatory protein hmgA converts homogentisate to maleylacetoacetate. Then, the enzyme hmgC converts maleylacetoacetate to fumarylacetaceatete, and hmgB converts fumarylacetaceatet to acetoacetate and fumarate, which enter the Krebs Cycle to produce energy for the bacteria.

Table 1 shows the accuracy of enzymes and substrates found in the homogentisate pathway of *P. putida* F1. The E-values indicate 100% accuracy for choices of enzymes identified in the homogentisate pathway of *P. putida* F1.

Table 1. Accuracy of enzymes and substrates found in the homogentisate pathway of *P. putida* F1. E-value of 0.0 indicates 100% accuracy.

Enzyme	Substrate	Product	Function	E-Value
phhT	NA	NA	transport protein	7.0×10^{-150}
hmgR	NA	NA	regulatory protein	2.0×10^{-173}
hmgA	homogentisate	maleylacetoacetate	catalyst	0.0
hmgC	maleylacetoacetate	fumarylacetacetate	catalyst	7.0×10^{-128}
hmgB	fumarylacetoacetase	acetoacetate and fumarate	catalyst	0.0

DISCUSSION

In conclusion, we identified enzymes and substrates that are involved in the degradation of homogentisate by *P. putida* F1. The hypothesis was supported. We were able to use protein sequences of enzymes that regulate the homogentisate pathway in the related bacteria that are closely related to *P. putida* F1 to identify enzymes involved in the homogentisate pathway of *P. putida* F1. Using the software programs and the online databases FASTA in GenBank, BLAST, T-COFFEE, and WebLogos, protein sequences for *P. putida* F1 were compared to other closely related bacterial strains that included *P. putida* KT2440, *P. aeruginosa* PAO1, *P. fluorescens* Pf0-1, and *Comamonas testosterone* CNB-1 to identify substrates and enzymes in the pathway by which homogentisate is catabolized by *P. putida* F1. PSORTb and TMHMM were then used to validate the transporter protein for the homogentisate pathway in *P. putida* F1. The best-fit BLAST sequences of phhT were entered into TMHMM and PSORTb in order to determine the validity of the enzyme and its function as well as its location in the bacteria. Then, BLAST was used to identify the regulatory protein for the homogentisate pathway in *P. putida* F1. E-values that were used to validate best-fit protein sequences showed 100% accuracy for choices of enzymes identified in the homogentisate pathway in *P. putida F1*. Consequently, we were invited to publish our results to GENI-ACT, which is a database that contains genomic information on 2775 bacteria.

Future work will involve two steps in shutting down the chemotactic response of *P. putida* F1 towards homogentisate. The first step will be to knock out the genes that encode the enzymes identified in this paper in order to shut down the homogentisate catabolic pathway. Even with the homogentisate catabolic pathway shut down, *P. putida* F1 will be chemotactic towards homogentisate. Therefore, the second step will be to identify receptor proteins for chemotaxis towards homogentisate using bioinformatics. Then, the genes that encode the receptor proteins can be knocked out. Next, the enzymes and substrates in the phenylacetate, quinate, and catechol catabolic pathways must be identified for *P. putida* F1, using the same protocol that was used for homogentisate.

Once the bacteria are engineered, the Environmental Protection Agency (EPA) must approve the bioengineered *P. putida* before it can be introduced into the environment. Because the bacteria will be used for degrading aromatic hydrocarbons, it will be subject to the Toxic Substances Control Act[6]. Steps that are required by the EPA for a genetically engineered bacteria to be released into the environment are a taxonomy report, a chemistry report, a construct analysis, ecological hazard assessment, a human health assessment, an exposure assessment, and a risk assessment. The taxonomy report includes the genus and species of the bacteria, and the chemistry report identifies genetic mutations made to the bacteria. The construct analysis identifies knocked-out receptor proteins, and the ecological hazard assessment identifies potential environmental impacts of the genetically engineered bacteria, including impacts of catabolic byproducts on organisms in the environment, such as aquatic and terrestrial vertebrates, invertebrates, and plants. The human health assessment identifies potential impacts of the genetically engineered bacteria and catabolic byproducts on human health. The exposure assessment identifies concentrations of the genetically engineered bacteria that are safe for release, while the risk assessment identifies hazard and exposure concerns for field testing[7]. Once EPA approval is granted, the next step will to run a controlled field test in a contained area where unexpected consequences to nontargeted organisms and the environment are closely studied.

Because homogentisate and other aromatic compounds, such as phenylacetate, quinate, and catechol, are more appetizing to *P. putida* than aromatic hydrocarbons are, the pathways by which *P. putida* F1 catabolize these aromatic nutrients must be fully understood before the bacteria can be engineered for use in bioremediation efforts against the aromatic hydrocarbons. Our study is an important step in this process.

ACKNOWLEDGEMENTS

Dr. J. Ditty served as our mentor. L. Fruen oversaw our study and reviewed our paper. The University of St. Thomas funded the project.

REFERENCES

1. Pandey G, Jaln RK. Bacterial chemotaxis toward environmental pollutants: role in bioremediation. *Applied and Environmental Microbiology*. 2002; 68(12):5789-5795.
2. Parales RE, Ju KS, Rollefson JB, Ditty JL. Bioavailability, transport, and chemotaxis towards organic pollutants. In: Microbial Biodegradation: Genomics and Molecular Biology. Diaz E, eds. Caister Academic Press, Norfolk, United Kingdom, 2008, pp. 145-187.
3. Harwood CS, Rivelli M, Ornston LN. Aromatic acids are chemoattractants for *Pseudomonas putida*. *Journal of Bacteriology*. 1984; 160(2):622-628.
4. Jiménez JI, Miñambres B, García JL, Díaz E. Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environmental Microbiology*. 2002; 4(12):824-841.
5. Arias-Barrau E, Olivera ER, Luengo JM, Fernández C, Galán B, García JL, Díaz E, Miñambres B. The homogentisate pathway: A central catabolic pathway involved in the degradation of a L-phenylalanine, L-tyrosine, and 3-hydroxyphenylacetate in *Pseudomonas putida*. *Journal of Bacteriology*. 2004; 186(15):5062-5077.
6. United States Environmental Protection Agency. Statement of policy; microbial products subject to the federal insecticide, fungicide, and rodenticide act and the toxic Substances control act. http://www.epa.gov/oppt/biotech/pubs/pdf/coordinated-framework-1986.pdf. Published June 26, 1986, accessed July 18, 2014.
7. Bioremediation and biosensing using bacteria. http://www.whitehouse.gov/files/documents/ostp/Issues/ceq_ostp_study7.pdf. Published, not available accessed August 13, 2014.