Reproduced with permission from Environmental Science and Technology, Sleight, T.; Khanna, V.; Gilbertson, L.; Ng, C. Network Analysis for Identifying High Impact Biodegradation Metabolites: A PAH Case Study. Environ. Sci. Technol. Rev. 2020. Copyright 2021, American Chemical Society

Supporting Information for

Network Analysis for Identifying High Impact Biodegradation Metabolites: A PAH Case Study

Authors:

Trevor W. Sleight, Vikas Khanna, Leanne M. Gilbertson and Carla A. Ng

108 Pages, 11 Figures, 12 Tables

Contents

[**Supporting Information Section 1: Acenaphthene** S2](#_Toc37225766)

[**Figure S1.1:** Acenaphthene High Throughput Plot with Pathway Match Metabolites S2](#_Toc37225767)

[**Table S1.1:** Acenaphthene Yellow Node (Pathway Matches) Smiles Codes S3](#_Toc37225768)

[**Figure S1.2:** Maximum Common Substructures (MCS) in the Top 1% of Betweenness: Acenaphthene. S7](#_Toc37225769)

[**Table S1.2:** Acenaphthene Empirical Literature Review S8](#_Toc37225770)

[**Table S1.3:** Acenaphthene Empirical Review Summary S13](#_Toc37225771)

[**Supporting Information Section 2: Anthracene** S15](#_Toc37225772)

[**Figure S2.1**: Anthracene High Throughput Plot with Pathway Match Metabolites S15](#_Toc37225773)

[**Table S2.1:** Anthracene Yellow Node (Pathway Matches) Smiles Codes S16](#_Toc37225774)

[**Figure S2.2:** Maximum Common Substructures (MCS) in the Top 1% of Betweenness: Anthracene. S17](#_Toc37225775)

[**Table S2.2**: Anthracene Empirical Literature Review S18](#_Toc37225776)

[**Table S2.3:** Anthracene Empirical Review Summary S32](#_Toc37225777)

[**Section 3: Fluorene** S36](#_Toc37225778)

[**Figure S3.1**: Fluorene High Throughput Plot with Pathway Match Metabolites S36](#_Toc37225779)

[**Table S3.1:** Fluorene Yellow Node (Pathway Matches) Smiles Codes S37](#_Toc37225780)

[**Figure S3.2:** Maximum Common Substructures (MCS) in the Top 1% of Betweenness: Fluorene. S39](#_Toc37225781)

[**Table S3.2**: Fluorene Empirical Literature Review S40](#_Toc37225782)

[**Table S3.3** Fluorene Empirical Review Summary S54](#_Toc37225783)

[**Section 4: Phenanthrene** S57](#_Toc37225784)

[**Figure S4.1:** Phenanthrene High Throughput Plot with Pathway Match Metabolite S57](#_Toc37225785)

[**Table S4.1:** Phenanthrene Yellow Node (Pathway Matches) Smiles Codes S58](#_Toc37225786)

[**Figure S4.2:** Maximum Common Substructures (MCS) in the Top 1% of Betweenness: Phenanthrene S59](#_Toc37225787)

[**Table S4.2**: Phenanthrene Empirical Literature Review S60](#_Toc37225788)

[**Table S4.3:** PhenanthreneEmpirical Review Summary S91](#_Toc37225789)

[**Section 5: Analysis and Plotting Code** S98](#_Toc37225790)

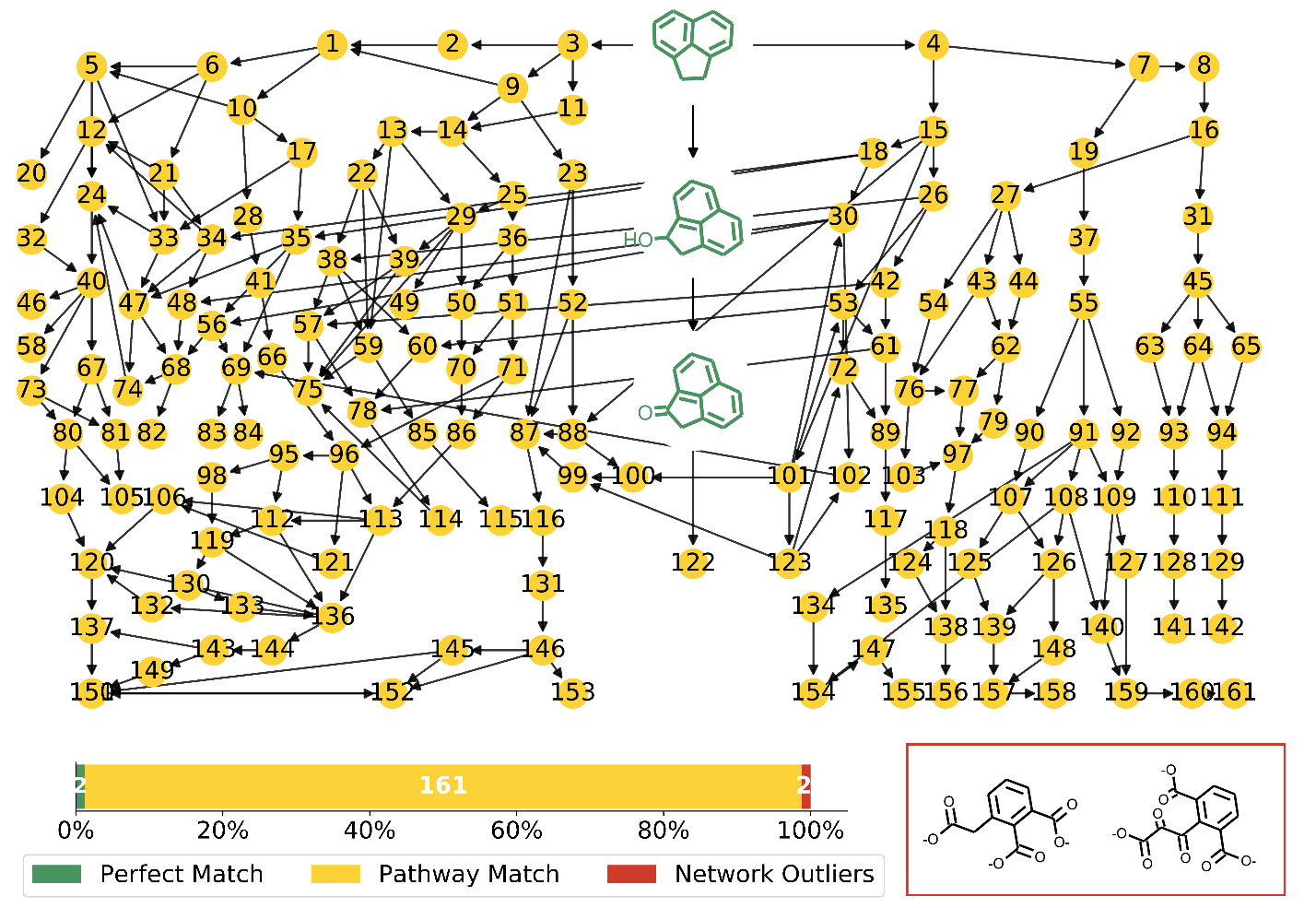
[**Section 6. Sensitivity Analysis** S100](#_Toc37225791)

[**Figure S6.1:** Network Metrics at Different Aerobic Likelihood Spreads S102](#_Toc37225792)

[**Figure S6.2:** Sensitivity Analysis of Different Weighting Schemes S102](#_Toc37225793)

[**Figure S6.3**: Selection of the Threshold for High Node Throughput S103](#_Toc37225794)

# Supporting Information Section 1: Acenaphthene

****

**Figure S1.1:** Acenaphthene High Throughput Plot with Pathway Match Metabolites. SMILES codes for the nodes are presented in Table S1.1.

Figure S1.1 shows a network constructed from only the High Node Throughput compounds. Reconstructing the network with only this subset of compounds yields a much smaller more manageable network.

Perfect Match compounds match a literature reference exactly. Pathway Match compounds are High Node Throughput compounds that are found on a degradation pathway to or from a Perfect Match compound. Network Outliers are High Node Throughput compounds that do not have a Perfect Match compound either above or below them in their degradation chains. Table S1.1 contains the SMILES codes for the Pathway Match compounds.

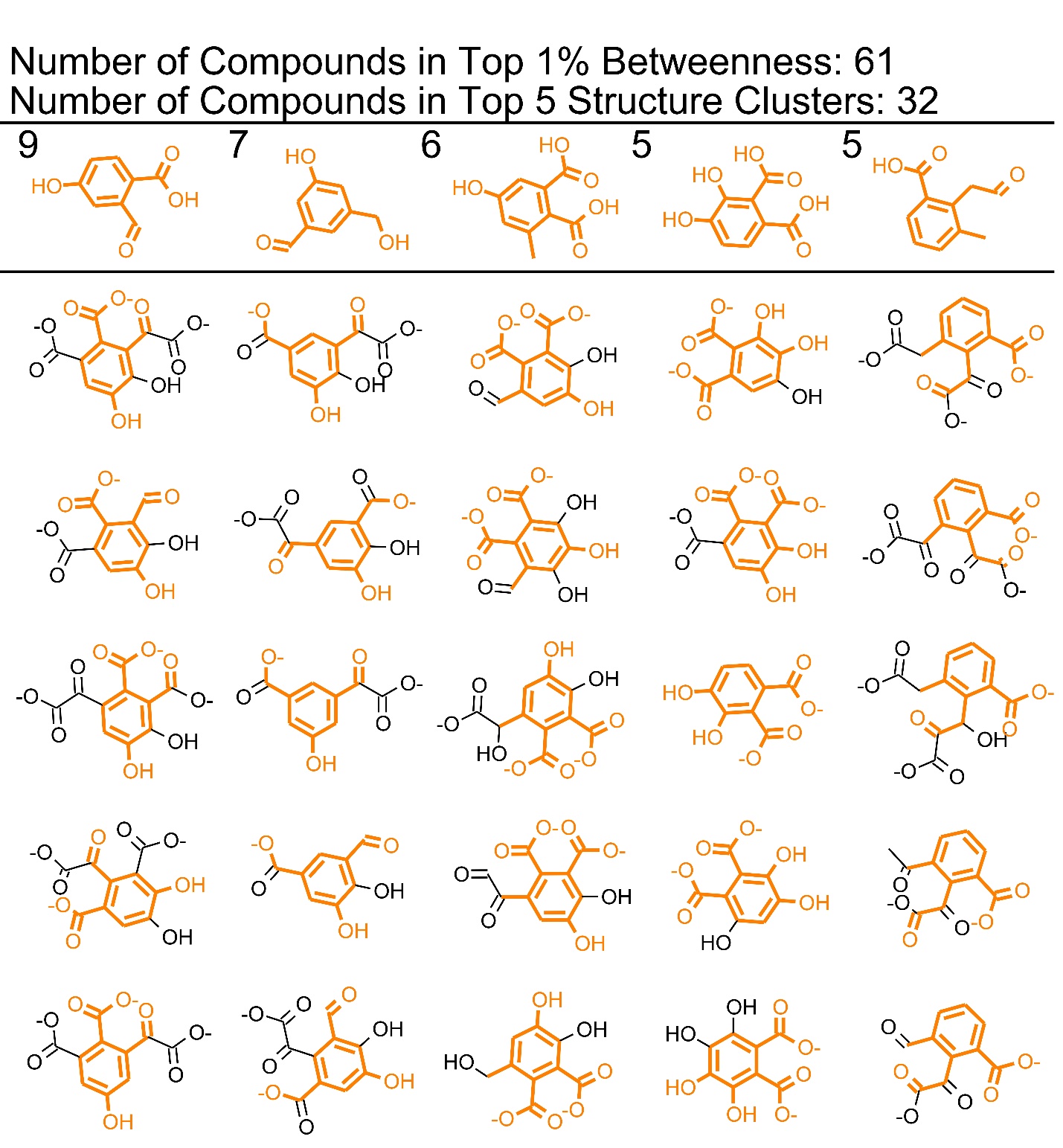
Acenaphthene, and to a lesser extent, Fluorene, both have an aliphatic 5-carbon ring in their structure, which breaks open early in the degradation pathway, contributing to a high number of metabolites with a High Node Throughput Value.

**Table S1.1:** Acenaphthene Yellow Node (Pathway Matches) Smiles Codes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | OC1Cc2cccc(C([O-])=O)c2\C1=C/C([O-])=O |  | 81 | Oc1cc(CC([O-])=O)c(C([O-])=O)c(c1)C([O-])=O |
| 2 | OC1Cc2cccc3c(O)c(O)cc1c23 |  | 82 | OC1(OC(=O)Cc2cccc(C([O-])=O)c12)C([O-])=O |
| 3 | Oc1cc2CCc3cccc(c1O)c23 |  | 83 | [O-]C(=O)c1cccc2COC(=O)Cc12 |
| 4 | Oc1cc2cccc3CCc(c1O)c23 |  | 84 | [O-]C(=O)c1cccc2CC(=O)OCc12 |
| 5 | OC1(CC([O-])=O)C(=O)Cc2cccc(C([O-])=O)c12 |  | 85 | OC1COC(=O)c2c1cccc2C([O-])=O |
| 6 | OC1Cc2cccc(C([O-])=O)c2C1(O)CC([O-])=O |  | 86 | OC(CC([O-])=O)(C=O)c1c(cccc1C([O-])=O)C([O-])=O |
| 7 | [O-]C(=O)\C=c1\cccc2CCC(C([O-])=O)=c12 |  | 87 | [O-]C(=O)c1cccc2CCC(=O)c12 |
| 8 | OC1Cc2ccc\c(=C\C([O-])=O)c2=C1C([O-])=O |  | 88 | OC1(CCc2cccc(C([O-])=O)c12)C(=O)C([O-])=O |
| 9 | [O-]C(=O)\C=C1\CCc2cccc(C([O-])=O)c12 |  | 89 | [O-]C(=O)c1cccc2C(=O)CCc12 |
| 10 | [O-]C(=O)\C=C1\C(=O)Cc2cccc(C([O-])=O)c12 |  | 90 | Oc1cc2C(=O)OCC(C([O-])=O)=c2\c(=C/C([O-])=O)c1O |
| 11 | OC1Cc2cc(O)c(O)c3cccc1c23 |  | 91 | OC\C(C([O-])=O)=c1/c(ccc/c/1=C/C([O-])=O)C([O-])=O |
| 12 | OC1Cc2cccc(C([O-])=O)c2C1=O |  | 92 | Oc1c\c(=C\C([O-])=O)c2=C(COC(=O)c2c1O)C([O-])=O |
| 13 | OC1CC(O)(CC([O-])=O)c2c1cccc2C([O-])=O |  | 93 | OCc1cc(O)c(O)\c(=C\C([O-])=O)\c\1=C\C([O-])=O |
| 14 | OC1C\C(=C\C([O-])=O)c2c1cccc2C([O-])=O |  | 94 | OCc1c(O)c(O)c\c(=C\C([O-])=O)\c\1=C\C([O-])=O |
| 15 | OC(=O)C(=O)C1CCc2cccc(C(O)=O)c12 |  | 95 | OC(C(C([O-])=O)c1c(cccc1C([O-])=O)C([O-])=O)C([O-])=O |
| 16 | [O-]C(=O)\C=c1\cccc2CC(=O)C(C([O-])=O)=c12 |  | 96 | [O-]C(=O)\C=C(\C([O-])=O)c1c(cccc1C([O-])=O)C([O-])=O |
| 17 | OC(C1C(=O)Cc2cccc(C([O-])=O)c12)C([O-])=O |  | 97 | OC(C([O-])=O)c1c(O)c(O)c(O)\c(=C\C([O-])=O)\c\1=C(\O)C([O-])=O |
| 18 | OC1Cc2cccc(C([O-])=O)c2C1C(=O)C([O-])=O |  | 98 | [O-]C(=O)C(C(=O)C([O-])=O)c1c(cccc1C([O-])=O)C([O-])=O |
| 19 | OC1CC(C([O-])=O)=c2c1ccc\c2=C\C([O-])=O |  | 99 | OC1CCc2cccc(C([O-])=O)c12 |
| 20 | OC1(CC([O-])=O)OC(=O)Cc2cccc(C([O-])=O)c12 |  | 100 | OC1(CCc2cccc(C([O-])=O)c12)C([O-])=O |
| 21 | OC(C([O-])=O)C1(O)C(O)Cc2cccc(C([O-])=O)c12 |  | 101 | [O-]C(=O)C1CCc2cccc(C([O-])=O)c12 |
| 22 | OC(C([O-])=O)C1(O)CC(O)c2cccc(C([O-])=O)c12 |  | 102 | OC1Cc2cccc(C([O-])=O)c2C1 |
| 23 | OC1(CC([O-])=O)CCc2cccc(C([O-])=O)c12 |  | 103 | OC(C([O-])=O)c1c(O)c(O)c\c(=C\C([O-])=O)\c\1=C(\O)C([O-])=O |
| 24 | [O-]C(=O)c1cccc2CC(=O)C(=O)c12 |  | 104 | [O-]C(=O)C(=O)c1cccc(C([O-])=O)c1C([O-])=O |
| 25 | [O-]C(=O)\C=C1\CC(=O)c2cccc(C([O-])=O)c12 |  | 105 | OC(C([O-])=O)c1cc(O)cc(C([O-])=O)c1C([O-])=O |
| 26 | OC1CC(C(=O)C([O-])=O)c2c1cccc2C([O-])=O |  | 106 | OC(CC([O-])=O)c1c(cccc1C([O-])=O)C([O-])=O |
| 27 | [O-]C(=O)\C=c1\cccc2CC(=O)OC(C([O-])=O)=c12 |  | 107 | OC\C(C([O-])=O)=c1/c(cc(O)c(O)/c/1=C/C([O-])=O)C([O-])=O |
| 28 | [O-]C(=O)\C=C1\C(=O)OCc2cccc(C([O-])=O)c12 |  | 108 | [O-]C(=O)\C=c1\cccc(C([O-])=O)\c\1=C(/C=O)C([O-])=O |
| 29 | OC1(CC([O-])=O)CC(=O)c2cccc(C([O-])=O)c12 |  | 109 | OC\C(C([O-])=O)=c1/c(C([O-])=O)c(O)c(O)c/c/1=C/C([O-])=O |
| 30 | OC1Cc2cccc(C([O-])=O)c2C1C([O-])=O |  | 110 | Oc1cc(C=O)\c(=C/C([O-])=O)\c(=C/C([O-])=O)\c1O |
| 31 | [O-]C(=O)\C=c1\cccc2CC(=O)C=c12 |  | 111 | Oc1c\c(=C\C([O-])=O)\c(=C\C([O-])=O)\c(C=O)c1O |
| 32 | OC1Cc2cccc(C([O-])=O)c2C(=O)O1 |  | 112 | OC(C([O-])=O)C(O)(C([O-])=O)c1c(cccc1C([O-])=O)C([O-])=O |
| 33 | OC(C([O-])=O)C1(O)C(=O)Cc2cccc(C([O-])=O)c12 |  | 113 | OC(CC([O-])=O)(C([O-])=O)c1c(cccc1C([O-])=O)C([O-])=O |
| 34 | OC1Cc2cccc(C([O-])=O)c2C1(O)C(=O)C([O-])=O |  | 114 | OC1CC(=O)c2cccc(C([O-])=O)c12 |
| 35 | [O-]C(=O)C(=O)C1C(=O)Cc2cccc(C([O-])=O)c12 |  | 115 | OCC(O)c1cccc(C([O-])=O)c1C([O-])=O |
| 36 | [O-]C(=O)\C=C1\COC(=O)c2cccc(C([O-])=O)c12 |  | 116 | [O-]C(=O)c1cccc2CCC(=O)Oc12 |
| 37 | [O-]C(=O)\C=c1\cccc2C(=O)CC(C([O-])=O)=c12 |  | 117 | [O-]C(=O)c1cccc2OC(=O)CCc12 |
| 38 | OC1CC(O)(C(=O)C([O-])=O)c2c1cccc2C([O-])=O |  | 118 | O\C(C([O-])=O)=c1/c(C(=O)C([O-])=O)c(O)c(O)c(O)/c/1=C/C([O-])=O |
| 39 | OC(C([O-])=O)C1(O)CC(=O)c2cccc(C([O-])=O)c12 |  | 119 | OC(C([O-])=O)(C(=O)C([O-])=O)c1c(cccc1C([O-])=O)C([O-])=O |
| 40 | [O-]C(=O)c1cccc2CC(=O)OC(=O)c12 |  | 120 | [O-]C(=O)c1cccc(C([O-])=O)c1C([O-])=O |
| 41 | OCc1cccc(C([O-])=O)c1\C(=C/C([O-])=O)C([O-])=O |  | 121 | [O-]C(=O)\C=C/c1c(cccc1C([O-])=O)C([O-])=O |
| 42 | [O-]C(=O)C(=O)C1CC(=O)c2cccc(C([O-])=O)c12 |  | 122 | O=C1OCc2cccc3cccc1c23 |
| 43 | O\C(C([O-])=O)=c1/c(CC([O-])=O)ccc/c/1=C/C([O-])=O |  | 123 | [O-]C(=O)c1cccc2CCCc12 |
| 44 | Oc1cc2CC(=O)OC(C([O-])=O)=c2\c(=C/C([O-])=O)c1O |  | 124 | O\C(C([O-])=O)=c1/c(C=O)c(O)c(O)c(O)/c/1=C/C([O-])=O |
| 45 | [O-]C(=O)\C=c1\cccc2COC(=O)C=c12 |  | 125 | OC\C(C([O-])=O)=c1/c(O)c(O)c(O)c(O)/c/1=C/C([O-])=O |
| 46 | Oc1cc(C([O-])=O)c2C(=O)OC(=O)Cc2c1O |  | 126 | Oc1cc(C([O-])=O)\c(=C(/C=O)C([O-])=O)\c(=C/C([O-])=O)\c1O |
| 47 | OC1(C(=O)C([O-])=O)C(=O)Cc2cccc(C([O-])=O)c12 |  | 127 | OC\C(C([O-])=O)=c1/c(O)c(O)c(O)c/c/1=C/C([O-])=O |
| 48 | OC1Cc2cccc(C([O-])=O)c2C1(O)C([O-])=O |  | 128 | Oc1cc(C([O-])=O)\c(=C/C([O-])=O)\c(=C/C([O-])=O)\c1O |
| 49 | CC(=O)c1cccc(C([O-])=O)c1C(=O)CC([O-])=O |  | 129 | Oc1c\c(=C\C([O-])=O)\c(=C\C([O-])=O)\c(C([O-])=O)c1O |
| 50 | OC1(CC([O-])=O)COC(=O)c2cccc(C([O-])=O)c12 |  | 130 | OC(C([O-])=O)(C([O-])=O)c1c(cccc1C([O-])=O)C([O-])=O |
| 51 | OC\C(=C\C([O-])=O)c1c(cccc1C([O-])=O)C([O-])=O |  | 131 | Oc1c(CCC([O-])=O)cccc1C([O-])=O |
| 52 | OC(C([O-])=O)C1(O)CCc2cccc(C([O-])=O)c12 |  | 132 | [O-]C(=O)c1cccc(C([O-])=O)c1C=O |
| 53 | OC1CC(C([O-])=O)c2c1cccc2C([O-])=O |  | 133 | OC(C([O-])=O)c1c(cccc1C([O-])=O)C([O-])=O |
| 54 | Oc1c\c(=C\C([O-])=O)c2=C(OC(=O)Cc2c1O)C([O-])=O |  | 134 | OC\C(C([O-])=O)=c1/c(O)c(O)cc/c/1=C/C([O-])=O |
| 55 | [O-]C(=O)\C=c1\cccc2C(=O)OCC(C([O-])=O)=c12 |  | 135 | Oc1cccc(C([O-])=O)c1CCC([O-])=O |
| 56 | [O-]C(=O)C1C(=O)Cc2cccc(C([O-])=O)c12 |  | 136 | [O-]C(=O)C(=O)c1c(cccc1C([O-])=O)C([O-])=O |
| 57 | OC1(CC(=O)c2cccc(C([O-])=O)c12)C(=O)C([O-])=O |  | 137 | Oc1cc(C([O-])=O)c(C([O-])=O)c(C([O-])=O)c1O |
| 58 | Oc1cc2CC(=O)OC(=O)c2c(C([O-])=O)c1O |  | 138 | O\C(C([O-])=O)=c1/c(C([O-])=O)c(O)c(O)c(O)/c/1=C/C([O-])=O |
| 59 | OC1CC(=O)c2c1cccc2C([O-])=O |  | 139 | Oc1c(O)c(O)\c(=C(/C=O)C([O-])=O)\c(=C/C([O-])=O)\c1O |
| 60 | OC1CC(O)(C([O-])=O)c2c1cccc2C([O-])=O |  | 140 | Oc1c\c(=C\C([O-])=O)\c(=C(\C=O)C([O-])=O)\c(C([O-])=O)c1O |
| 61 | [O-]C(=O)C1CC(=O)c2cccc(C([O-])=O)c12 |  | 141 | Oc1c(O)c(O)\c(=C/C([O-])=O)\c(=C/C([O-])=O)\c1O |
| 62 | O\C(C([O-])=O)=c1/c(CC([O-])=O)cc(O)c(O)/c/1=C/C([O-])=O |  | 142 | Oc1c\c(=C\C([O-])=O)\c(=C\C([O-])=O)\c(O)c1O |
| 63 | Oc1cc2COC(=O)C=c2\c(=C/C([O-])=O)c1O |  | 143 | Oc1cc(C([O-])=O)c(C(=O)C([O-])=O)c(C([O-])=O)c1O |
| 64 | OCc1ccc\c(=C\C([O-])=O)\c\1=C\C([O-])=O |  | 144 | Oc1cc(C([O-])=O)c(C(=O)C([O-])=O)c(c1)C([O-])=O |
| 65 | Oc1c\c(=C\C([O-])=O)c2=CC(=O)OCc2c1O |  | 145 | Oc1cccc(C([O-])=O)c1O |
| 66 | [O-]C(=O)\C=C(\C([O-])=O)c1c(C=O)cccc1C([O-])=O |  | 146 | Oc1c(cccc1C([O-])=O)C([O-])=O |
| 67 | OC(=O)Cc1cccc(C([O-])=O)c1C([O-])=O |  | 147 | Oc1cc\c(=C\C([O-])=O)c(=C(C([O-])=O)C([O-])=O)c1O |
| 68 | OC1(C([O-])=O)C(=O)Cc2cccc(C([O-])=O)c12 |  | 148 | Oc1cc(C([O-])=O)c(=C(C([O-])=O)C([O-])=O)\c(=C/C([O-])=O)c1O |
| 69 | [O-]C(=O)c1cccc2CC(=O)Cc12 |  | 149 | Oc1cc(C([O-])=O)c(C(=O)C([O-])=O)c(O)c1O |
| 70 | OCC(O)(CC([O-])=O)c1c(cccc1C([O-])=O)C([O-])=O |  | 150 | Oc1cc(C([O-])=O)c(C([O-])=O)c(O)c1O |
| 71 | [O-]C(=O)\C=C(\C=O)c1c(cccc1C([O-])=O)C([O-])=O |  | 151 | Oc1ccc(O)c(O)c1O |
| 72 | OC1CCc2c1cccc2C([O-])=O |  | 152 | Oc1ccc(C([O-])=O)c(O)c1O |
| 73 | OC(=O)c1c(CC([O-])=O)cccc1C([O-])=O |  | 153 | Oc1cc(C([O-])=O)c(O)c(C([O-])=O)c1O |
| 74 | OC1C(=O)Cc2cccc(C([O-])=O)c12 |  | 154 | Oc1cc\c(=C\C([O-])=O)\c(=C(\C=O)C([O-])=O)\c1O |
| 75 | [O-]C(=O)c1cccc2C(=O)CC(=O)c12 |  | 155 | Oc1cc\c(=C\C([O-])=O)c(=CC([O-])=O)c1O |
| 76 | O\C(C([O-])=O)=c1/c(CC([O-])=O)c(O)c(O)c/c/1=C/C([O-])=O |  | 156 | O\C(C([O-])=O)=c1/c(O)c(O)c(O)c(O)/c/1=C/C([O-])=O |
| 77 | O\C(C([O-])=O)=c1/c(CC([O-])=O)c(O)c(O)c(O)/c/1=C/C([O-])=O |  | 157 | Oc1c(O)c(O)c(=C(C([O-])=O)C([O-])=O)\c(=C/C([O-])=O)c1O |
| 78 | OC1(CC(=O)c2cccc(C([O-])=O)c12)C([O-])=O |  | 158 | Oc1c(O)c(O)\c(=C\C([O-])=O)c(=CC([O-])=O)c1O |
| 79 | OC(C([O-])=O)c1cc(O)c(O)\c(=C\C([O-])=O)\c\1=C(\O)C([O-])=O |  | 159 | Oc1c\c(=C\C([O-])=O)\c(=C(\C=O)C([O-])=O)\c(O)c1O |
| 80 | OC(C([O-])=O)c1cccc(C([O-])=O)c1C([O-])=O |  | 160 | Oc1c\c(=C\C([O-])=O)c(=C(C([O-])=O)C([O-])=O)c(O)c1O |
|  |  |  | 161 | Oc1c\c(=C\C([O-])=O)c(=CC([O-])=O)c(O)c1O |

**High Betweenness Maximum Common Substructures:**

The compounds with the highest betweenness centrality display similar substructures. Code was adapted from the RDKit Cookbook, referenced in the main text. The Tanimto Similarities between the Morgan Fingerprints were first calculated, and then clusters were developed with Butina Clustering using a cutoff of 0.4. For full details and the actual python code, see the github linked in the main text.



**Figure S1.2:** Maximum Common Substructures (MCS) in the Top 1% of Betweenness: Acenaphthene. The first row shows the MCS in top 5 largest clusters. The numbers indicate how many compounds are contained in each cluster. The compounds in the vertical columns with the substructure highlighted in orange are presented as examples of which compounds are included in the cluster. In the case of acenaphthene, each cluster has a unique substructure.

**Table S1.2:** Acenaphthene Empirical Literature Review

|  |  |  |  |
| --- | --- | --- | --- |
| Structure: | Network Metrics | Empirical Studies finding the same thing | Similar Compounds |
| Acenaphthene |  |  |  |
| OC1Cc2cccc3cccc1c23  1-Acenaphthenol | **Perfect Match**  Throughput: 0.015 | Ghosal et al, 2013 [1](#_ENREF_1) Komatsu et al, 1993[2](#_ENREF_2)  Mallick, 2019[3](#_ENREF_3)  Schocken et al, 1984[4](#_ENREF_4)  Birolli et al, 2018[5](#_ENREF_5)  Liu et al, 2006[6](#_ENREF_6)  Birolli et al, 2018[5](#_ENREF_5) |  |
| OC1c2cccc3cccc(c23)C1O  cis-1,2-Acenaphthenediol  1,2-Acenaphthenediol | **Partial Match**    Throughput: 0.015 | Mallick, 2019[3](#_ENREF_3)  Schocken et al, 1984[4](#_ENREF_4) Kouzuma et al, 2006 |  |
| C1=Cc2cccc3cccc1c23  acenaphthylene | **Partial Match**    OC1Cc2cccc3cccc1c23  Throughput: 0.015 | Birolli et al, 2018[5](#_ENREF_5) |  |
| OC1=C(O)c2cccc3cccc1c23  1,2-dihydroxyacenaphthylene | **No Match (Single paper)** | Schocken et al, 1984[4](#_ENREF_4) |  |
| O=C1Cc2cccc3cccc1c23  1-Acenaphthenone | **Perfect match**  Throughput: 0.015 | Ghosal et al, 2013[1](#_ENREF_1)  Komatsu et al, 1993[2](#_ENREF_2)  Mallick, 2019[3](#_ENREF_3)  Schocken et al, 1984[4](#_ENREF_4)  Chapman et al, 1995[7](#_ENREF_7) |  |
| O=c1c2cccc3cccc(c23)c1=O  Acenaphthenequinone | **Partial Match**    Throughput: 0.015  OC1Cc2cccc3cccc1c23 | Ghosal et al, 2013[1](#_ENREF_1)  Mallick, 2019[3](#_ENREF_3)  Chapman et al, 1995[7](#_ENREF_7)  Poothrigpun et al, 2006[8](#_ENREF_8) |  |
| [O-]C(=O)c1cccc2ccccc12  1-Naphthoic acid | **Low Throughput Match**  Throughput: 0.000002 | Ghosal et al, 2013[1](#_ENREF_1)  Mallick, 2019[3](#_ENREF_3)  Poothrigpun et al, 2006[8](#_ENREF_8)  Nayak et al, 2009[9](#_ENREF_9) |  |
| O=C(O)/C=c1\cccc2c1=C(O)C(=O)C=C2  or    O=C(O)Cc1cccc2c1C(=O)C(=O)C=C2  7,8-diketonaphthyl-1-acetic acid | **No Match (Single paper)** | Selifononv et al, 1993[10](#_ENREF_10) |  |
| Oc1cccc(C([O-])=O)c1C([O-])=O  o-methoxy-o-phthalic acid | **Low Throughput Match**  Throughput: 0.0078 | Selifononv et al, 1993[10](#_ENREF_10) |  |
| Oc1ccccc1C([O-])=O  Salicylic Acid/ 2-hydroxybenzoic acid | **Low Throughput Match**  Throughput: 0.0027 | Ghosal et al, 2013[1](#_ENREF_1)  Mallick, 2019[3](#_ENREF_3)  Nayak et al, 2009[9](#_ENREF_9)  Poothrigpun et al, 2006[8](#_ENREF_8) |  |
| Gentisate  gentisic acid  O=C(O)c1cc(O)ccc1O | **Low Throughput Match**  Oc1ccc(O)c(c1)C([O-])=O  Throughput: 0.00039 | Poothrigpun et al, 2006[8](#_ENREF_8) |  |
| 3-Formyl Salicylic Acid  O=Cc1cccc(C(=O)O)c1O | **Low Throughput Match**  Oc1c(C=O)cccc1C([O-])=O  Throughput: 0.00022 | Mallick, 2019[3](#_ENREF_3) |  |
| Oc1ccc2ccccc2c1O  1,2-Dihydroxyacenapthylene | **Low Throughput Match**  Throughput: 0.000001 | Mallick, 2019[3](#_ENREF_3)  Nayak et al, 2009[9](#_ENREF_9) |  |
| O=C1OC(=O)c2cccc3cccc1c23  1,8-Naphthalic anhydride | **Partial Match**    Throughput: 0.015 | Mallick, 2019[3](#_ENREF_3)  Chapman et al, 1995[7](#_ENREF_7) |  |
| O=C(O)c1cccc2cccc(C(=O)O)c12  Napthalene-1,8-dicarboxylic acid dimethyl ester | **Low Throughput Match**  [O-]C(=O)c1cccc2cccc(C([O-])=O)c12  Throughput: 0.00017 | Mallick, 2019[3](#_ENREF_3)  Chapman et al, 1995[7](#_ENREF_7)  Nayak et al, 2009[9](#_ENREF_9)  Poothrigpun et al, 2006[8](#_ENREF_8)  Selifononv et al, 1993[10](#_ENREF_10) |  |
| Oc1ccccc1O  Catechol | **Low Throughput Match**  Throughput: 0.00055 | Mallick, 2019[3](#_ENREF_3), Mallick 2011[11](#_ENREF_11) |  |

**Table S1.3:** Acenaphthene Empirical Review Summary

|  |  |  |
| --- | --- | --- |
|  | Perfect Match | Partial Match |
| High Throughput | 2 | 4 |
| Low Throughput | 8 | 0 |

No Match (Single) 2

No Match (Multiple) 0

**References**

(1) Ghosal, D., et al., Characterization of the Metabolic Pathway Involved in Assimilation of Acenaphthene in Acinetobacter Sp. Strain Agat-W*.* *Res Microbiol*, 2013. 164(2): p. 155-63.

(2) Komatsu, T., T. Omori, and T. Kodama, Microbial-Degradation of the Polycyclic Aromatic-Hydrocarbons Acenaphthene and Acenaphthylene by a Pure Bacterial Culture*.* *Bioscience Biotechnology and Biochemistry*, 1993. 57(5): p. 864-865.

(3) Mallick, S., Biodegradation of Acenaphthene by Sphingobacterium Sp. Strain Rtsb Involving Trans-3-Carboxy-2-Hydroxybenzylidenepyruvic Acid as a Metabolite*.* *Chemosphere*, 2019. 219: p. 748-755.

(4) Schocken, M.J. and D.T. Gibson, Bacterial Oxidation of the Polycyclic Aromatic-Hydrocarbons Acenaphthene and Acenaphthylene*.* *Applied and Environmental Microbiology*, 1984. 48(1): p. 10-16.

(5) Birolli, W.G., et al., Biodegradation of Anthracene and Several Pahs by the Marine-Derived Fungus Cladosporium Sp Cbmai 1237*.* *Marine Pollution Bulletin*, 2018. 129(2): p. 525-533.

(6) Liu, L., R.D. Schmid, and V.B. Urlacher, Cloning, Expression, and Characterization of a Self-Sufficient Cytochrome P450 Monooxygenase from Rhodococcus Ruber Dsm 44319*.* *Applied Microbiology and Biotechnology*, 2006. 72(5): p. 876-882.

(7) Chapman, P.J., et al., Fossil Fuel Biodegradation: Laboratory Studies*.* *Environmental Health Perspectives*, 1995. 103: p. 79-83.

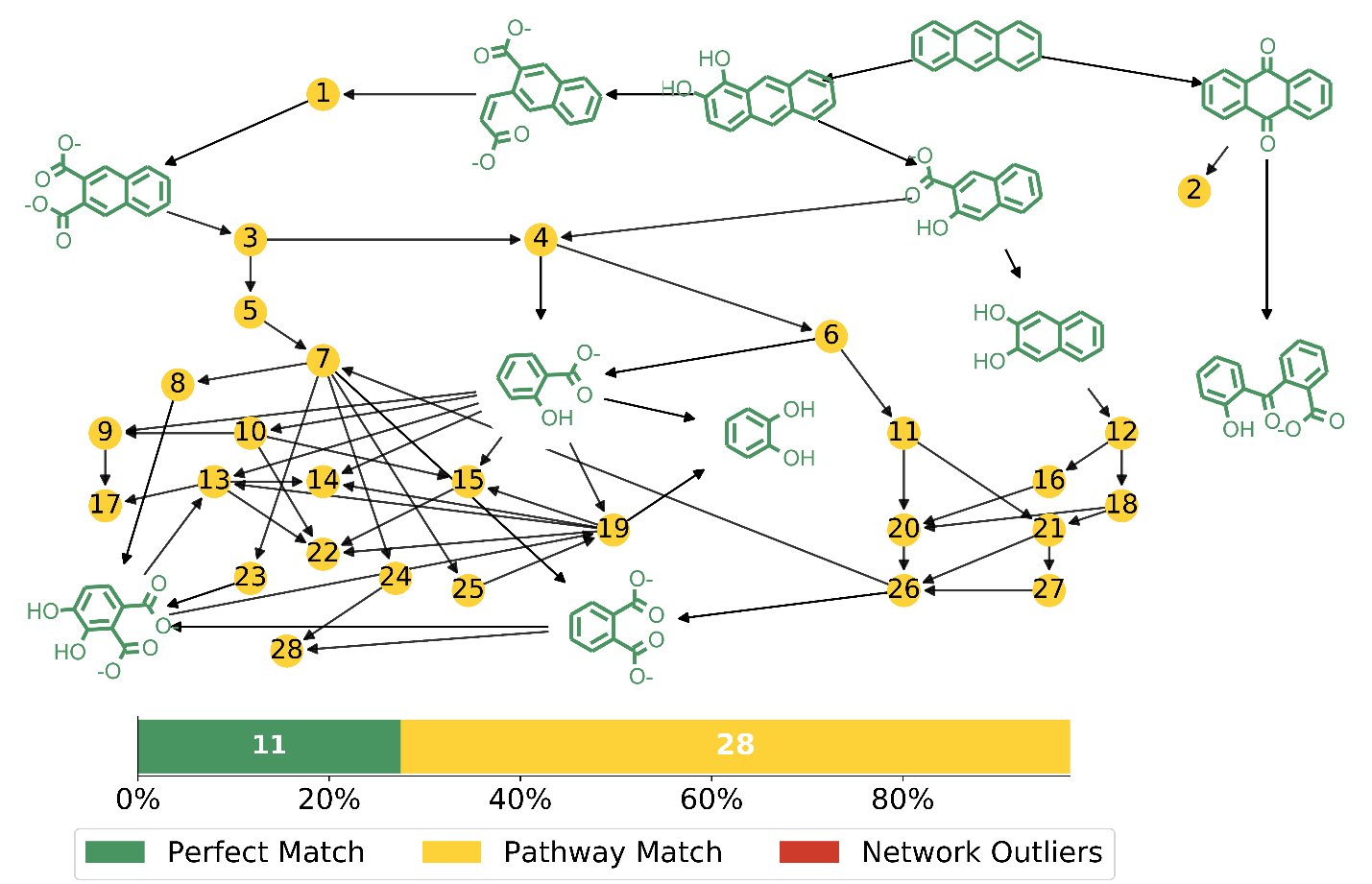
(8) Poonthrigpun, S., et al., Novel Intermediates of Acenaphthylene Degradation by Rhizobium Sp Strain Cu-A1: Evidence for Naphthalene-1,8-Dicarboxylic Acid Metabolism*.* *Applied and Environmental Microbiology*, 2006. 72(9): p. 6034-6039.

(9) Nayak, A.S., et al., Metabolism of Acenaphthylene Via 1,2-Dihydroxynaphthalene and Catechol by Stenotrophomonas Sp Rmsk*.* *Biodegradation*, 2009. 20(6): p. 837-843.

(10) Selifonov, S.A., et al., Acenaphthene Catabolism by Strains of Alcaligenes-Eutrophus and Alcaligenes-Paradoxus*.* *Microbiology*, 1993. 62(1): p. 85-92.

(11) Mallick, S., J. Chakraborty, and T.K. Dutta, Role of Oxygenases in Guiding Diverse Metabolic Pathways in the Bacterial Degradation of Low-Molecular-Weight Polycyclic Aromatic Hydrocarbons: A Review*.* *Crit Rev Microbiol*, 2011. 37(1): p. 64-90.

# Supporting Information Section 2: Anthracene



**Figure S2.1**: Anthracene High Throughput Plot with Pathway Match Metabolites. SMILES codes for the nodes are presented in Table S2.1.

Figure S2.1 shows a network constructed from only the High Node Throughput compounds. Reconstructing the network with only this subset of compounds yields a much smaller more manageable network.

Perfect Match compounds match a literature reference exactly. Pathway Match compounds are High Node Throughput compounds that are found on a degradation pathway to or from a Perfect Match compound. Network Outliers are High Node Throughput compounds that do not have a Perfect Match compound either above or below them in their degradation chains. Table S2.1 contains the SMILES codes for the Pathway Match compounds.

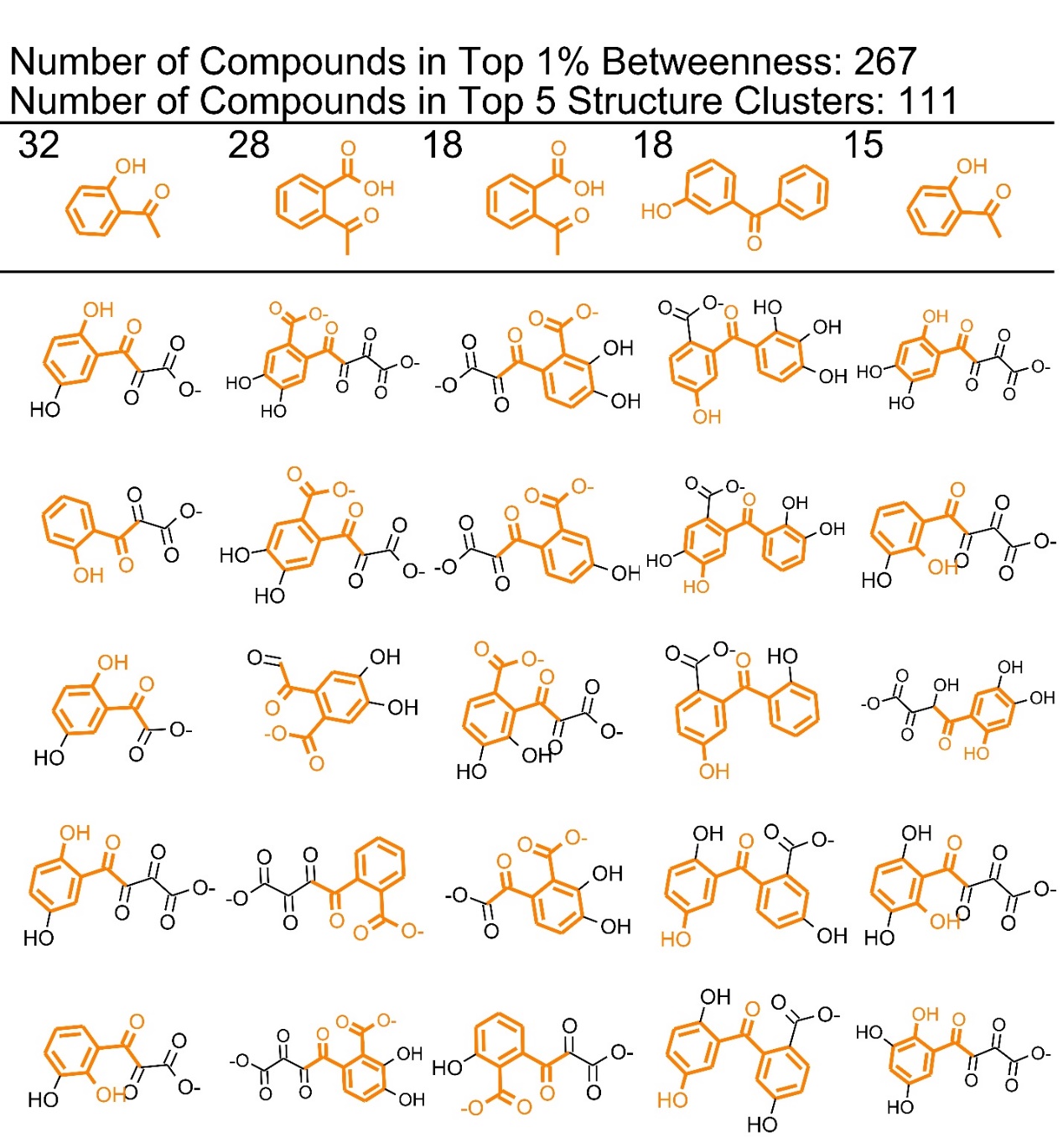
**Table S2.1:** Anthracene Yellow Node (Pathway Matches) Smiles Codes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | OC(CC([O-])=O)c1cc2ccccc2cc1C([O-])=O |  | 15 | Oc1ccc(O)c(C([O-])=O)c1O |
| 2 | O=c1oc2ccccc2c(=O)c2ccccc12 |  | 16 | [O-]C(=O)Cc1ccccc1C([O-])=O |
| 3 | Oc1c(C([O-])=O)c(cc2ccccc12)C([O-])=O |  | 17 | Oc1cc(O)c(O)c(O)c1O |
| 4 | Oc1c(O)c2ccccc2cc1C([O-])=O |  | 18 | OC(C(=O)C([O-])=O)c1ccccc1C([O-])=O |
| 5 | [O-]C(=O)C(=O)C(=C\c1ccccc1C([O-])=O)\C([O-])=O |  | 19 | Oc1cccc(C([O-])=O)c1O |
| 6 | Oc1cc2ccccc2c(O)c1O |  | 20 | OC(C([O-])=O)c1ccccc1C([O-])=O |
| 7 | [O-]C(=O)c1ccccc1C=O |  | 21 | [O-]C(=O)C(=O)C(=O)c1ccccc1C([O-])=O |
| 8 | Oc1ccc(C([O-])=O)c(C=O)c1O |  | 22 | Oc1ccc(O)c(O)c1O |
| 9 | Oc1cc(O)c(cc1O)C([O-])=O |  | 23 | Oc1ccc(C=O)c(C([O-])=O)c1O |
| 10 | Oc1ccc(O)c(c1)C([O-])=O |  | 24 | Oc1cc(C=O)c(cc1O)C([O-])=O |
| 11 | OC(C(=O)C(O)=O)c1ccccc1C(O)=O |  | 25 | Oc1cccc(C=O)c1O |
| 12 | OC(=O)C(=O)Cc1ccccc1C(O)=O |  | 26 | [O-]C(=O)C(=O)c1ccccc1C([O-])=O |
| 13 | Oc1ccc(C([O-])=O)c(O)c1O |  | 27 | [O-]C(=O)c1ccccc1C(=O)C=O |
| 14 | Oc1cccc(O)c1O |  | 28 | Oc1cc(C([O-])=O)c(cc1O)C([O-])=O |

**High Betweenness Maximum Common Substructures**:

The compounds with the highest betweenness centrality display similar substructures. Code was adapted from the RDKit Cookbook, referenced in the main text. The Tanimto Similarities between the Morgan Fingerprints were first calculated, and then clusters were developed with Butina Clustering using a cutoff of 0.4.

For full details and the actual python code to plot this figure, see the github linked in the main text.



**Figure S2.2:** Maximum Common Substructures (MCS) in the Top 1% of Betweenness: Anthracene. The first row shows the MCS in top 5 largest clusters. The numbers indicate how many compounds are contained in each cluster. The compounds in the vertical columns with the substructure highlighted in orange are presented as examples of which compounds are included in the cluster. In the case of anthracene, the same MCS occurs in several different clusters.

**Table S2.2**: Anthracene Empirical Literature Review

|  |  |  |
| --- | --- | --- |
| Structure: | Network Metrics | Empirical Studies finding the same thing |
| Anthracene |  |  |
| image1.png  O=C1c2ccccc2C(=O)c2ccccc12  9,10-Anthraquinone | **Perfect Match**  Node Throughput: 0.029 | Leneva et al, 2008 [1](#_ENREF_1), Moody et al, 2001 [2](#_ENREF_2), Ahmed et al, 2012 [3](#_ENREF_3), Hadibarata et al, 2012 [4](#_ENREF_4) Li et al, 2014 [5](#_ENREF_5) Ye et al, 2011 [6](#_ENREF_6)  Aranda et al, 2017[7](#_ENREF_7)  Birolli et al, 2018[8](#_ENREF_8)  Gunther et a, 1998[9](#_ENREF_9)  Hadibarata et al, 2013[10](#_ENREF_10) Hammel et al, 1991[11](#_ENREF_11), Baboshin et al, 2005[12](#_ENREF_12), Bezalel et al, 1996[13](#_ENREF_13), Bylik et al, 1999[14](#_ENREF_14)  Cajthaml et al, 2002[15](#_ENREF_15)  Chigu et al, 2010[16](#_ENREF_16)  Coutino-Gonzalez et al, 2010[17](#_ENREF_17)  Field et al, 1992[18](#_ENREF_18)  Guiraud et al, 2008[19](#_ENREF_19)  Kastner et al, 1999[20](#_ENREF_20)  Li et al, 2014[5](#_ENREF_5)  Lisowska et al, 1999[21](#_ENREF_21)  Pathak et al, 2008[22](#_ENREF_22)  Peng et al, 2018[23](#_ENREF_23)  Pizzul et al, 2006[24](#_ENREF_24)  Sutherland et al, 1992[25](#_ENREF_25)  Tarafdar et al, 2017[26](#_ENREF_26)  Traczewska et al, 2000[27](#_ENREF_27)  Weigand et al, 1999[28](#_ENREF_28)  Wu et al, 2008[29](#_ENREF_29)  Wu et al, 2010[30](#_ENREF_30) |
| 2-hydroxyanthraquinone  O=C1c2ccccc2C(=O)c2cc(O)ccc21 | **Partial Match**  image1.png  O=C1c2ccccc2C(=O)c2ccccc12  Throughput: 0.029 | Chigu et al, 2010[16](#_ENREF_16) |
| 1,4-Anthraquinone  O=C1C=CC(=O)C2C=c3ccccc3=CC12 | **No Match (Single paper)** | Godoy et al, 2016[31](#_ENREF_31) |
| 1-4 dihydroxyanthraquinone  O=C1c2ccccc2C(=O)c2c(O)ccc(O)c21 | **Partial Match**  image1.png  O=C1c2ccccc2C(=O)c2ccccc12  Throughput: 0.029 | Guiraud et al, 2008[19](#_ENREF_19) |
| 1,8 dihydroxyanthraquinone  O=C1c2cccc(O)c2C(=O)c2c(O)cccc21 | **Partial Match**  image1.png  O=C1c2ccccc2C(=O)c2ccccc12  Throughput: 0.029 | Guiraud et al, 2008[19](#_ENREF_19) |
| O=C1c2ccccc2Cc2ccccc12  9-Anthraquinone  Anthrone | **Partial Match**  image1.png  O=C1c2ccccc2C(=O)c2ccccc12  Throughput: 0.029 | Aranda et al, 2017[7](#_ENREF_7)  Birolli et al, 2018[8](#_ENREF_8)  Cajthaml et al, 2002[15](#_ENREF_15)  Chigu et al, 2010[16](#_ENREF_16)  Richnow et al, 1998[32](#_ENREF_32)  Wu et al, 2010[30](#_ENREF_30) |
| Oc1c2ccccc2c(O)c2ccccc12  9,10 dihydroxy anthracene | **Partial Match**  immediate parent and daughter contained in model | Bidi et al, 2018[33](#_ENREF_33)  Traczewska et al, 2000[27](#_ENREF_27) |
| image1.png  Oc1ccc2cc3ccccc3cc2c1O  1,2-dihydroxyanthracene | **Perfect Match**  Throughput: 0.97 | Chandrasekhar et al, 2010[34](#_ENREF_34)  Fukelar et al, 2017, [35](#_ENREF_35), Leneva et al, 2009 [1](#_ENREF_1) Moody et al, 2001 [2](#_ENREF_2), Ahmed et al, 2012 [3](#_ENREF_3) Hammel et al, 1991[11](#_ENREF_11), Bezalel et al, 1996[13](#_ENREF_13)  Brinda et al, 2011[36](#_ENREF_36)  da Silva et al, 2004[37](#_ENREF_37)  Pathak et al, 2008[22](#_ENREF_22)  Sutherland et al, 1992[25](#_ENREF_25)  van Herwijnen et al, 2003[38](#_ENREF_38) |
| 1-methylanthracene  Cc1cccc2cc3ccccc3cc12 | **No Match (Single paper)** | Tarafdar et al, 2017[26](#_ENREF_26) |
| 1-anthrol  Oc1cccc2cc3ccccc3cc12 | **Partial Match**  image1.png  Throughput: 0.97  Oc1ccc2cc3ccccc3cc2c1O | Sutherland et al, 1992[25](#_ENREF_25) |
| O=C1OC2=CC3=CC=CC=C3C=C2C=C1  6,7 benzocoumarin | **No Match (Multiple papers)**  however, this compound forms from an out-and back-pathway that does not contribute to the overall biodegradation chain. | Ahmed et al, 2012[3](#_ENREF_3) Leneva et al, 2009[1](#_ENREF_1)  Baboshin et al, 2005[12](#_ENREF_12)  Kim et al, 1997[39](#_ENREF_39)  Pathak et al, 2008[22](#_ENREF_22)  van Herwijnen et al, 2003[38](#_ENREF_38) |
| COC1=C(C=CC2=CC3=CC=CC=C3C=C21)O  1-methoxy-2-hydroxyanthracene | **No Match (Multiple papers)**  because EAWAG does not predict methylations. The immediate precursor 1,2-dihydroxyanthracene is predicted with a node throughput of 0.91  image1.png | Ahmed et al, 2012 [3](#_ENREF_3)  Pathak et al, 2008[22](#_ENREF_22) |
| Benzophenone  O=C(c1ccccc1)c1ccccc1 | **Low Throughput Partial Match**    Throughput: 0.0018  Oc1cccc(C(=O)c2ccccc2O)c1O | Wu et al, 2010[30](#_ENREF_30) |
| 2,2’-Methylenediphenol  Oc1ccccc1Cc1ccccc1O | **Low Throughput Partial Match**    Throughput: 0.0018  Oc1cccc(C(=O)c2ccccc2O)c1O | Tarafdar et al, 2017[26](#_ENREF_26) |
| 4,4’-(Hexane-3,4-diyl)diphenol  CCC(c1ccc(O)cc1)C(CC)c1ccc(O)cc1 | **No Match (Single paper)** | Tarafdar et al, 2017[26](#_ENREF_26) |
| 1-Methoxy-4-(p-tolylethynyl)benzene  COc1ccc(C#Cc2ccc(C)cc2)cc1 | **No Match (Single paper)** | Tarafdar et al, 2017[26](#_ENREF_26) |
| 2,2’-(1,4-Phenylene)dipropionic acid  CC(C(=O)O)c1ccc(C(C)C(=O)O)cc1 | **No Match (Single paper)** | Tarafdar et al, 2017[26](#_ENREF_26) |
| 9-fluorenol, 9-Hydroxyfluorene  OC1c2ccccc2-c2ccccc12 | **No Match (Single paper)** | Brinda et al, 2011[36](#_ENREF_36) |
| **2-Propanone, 1,1-diphenyl-**  CC(=O)C(c1ccccc1)c1ccccc1 | **No Match (Single paper)** | Wu et al, 2010[30](#_ENREF_30) |
| 6103image1.png  [O-]C(=O)\C=C/c1cc2ccccc2cc1C([O-])=O  3-[(Z)-2-carboxyvinyl]-2-naphthoate, 3-(2-carboxyvinyl)naphthalene-2-carboxylic | **Perfect Match**  Node Throughput: 0.32 | Moody et al, 2001 [2](#_ENREF_2)  Brinda et al, 2011[36](#_ENREF_36) |
| (E)-3-(2-hydroxynaphthalen-3-yl)acrylic acid  O=C(O)C=Cc2cc1ccccc1cc2O | **Partial Match**  Immediate Parent and Daughter predicted: Occupies the same place in the degradation chain as 3-[(Z)-2-carboxyvinyl]-2-naphthoate. | Ahmed et al, 2012 [3](#_ENREF_3) |
| *cis*-4-(2-Hydroxynaphth-3-yl)-2-oxobut-3-enoic acid  O=C(O)C(=O)/C=C\c1cc2ccccc2cc1O | **No Match (Single paper)** | van Herwijnen et al, 2003 [38](#_ENREF_38) |
| 4-(3-Hydroxy-2-naphthyl)butanoic acid  O=C(O)CCCc1cc2ccccc2cc1O | **No Match (Multiple papers)** | Weigand et al, 1999[28](#_ENREF_28)  Baboshin et al, 2005[12](#_ENREF_12) |
| 2-Hydroxy-3-naphthyl -propionic  Acid  O=C(O)CCc1cc2ccccc2cc1O | **No Match (Single paper)** | Baboshin et al, 2005[12](#_ENREF_12) |
| Naphthalene 2,3 dicarboxylic acid  [O-]C(=O)c1cc2ccccc2cc1C([O-])=O | **Perfect Match**  Throughput: 0.16 | van Herwijnen et al, 2003[40](#_ENREF_40) |
| 6103image1.png  Oc1cc2ccccc2cc1C([O-])=O  3-Hydroxy-2-naphthoate | **Perfect Match**  Throughput: 0.32 | Menn et al, 1993 [41](#_ENREF_41)  Weigand et al, 1999[28](#_ENREF_28)  Different protonation state:  Hadibarata et al, 2013[10](#_ENREF_10), Baboshin et al, 2005[12](#_ENREF_12)  Liu et al, 1992[42](#_ENREF_42)  Richnow et al, 1998[32](#_ENREF_32)  Story et al, 2001[43](#_ENREF_43)  Van Herwijnen et al, 2003[38](#_ENREF_38) |
| C1(C=2C(C(=O)O1)=CC=CC2)=O  Phthalic anhydride | **Partial Match**  Not predicted, but its immediate precursor and immediate daughter are | Ye et al, 2011[6](#_ENREF_6)  Bylik et al, 1999[14](#_ENREF_14)  Cajthaml et al, 2002[15](#_ENREF_15)  Traczewska et al, 2000[27](#_ENREF_27) |
| O=C1OCc2ccccc21  Phthalide | **No Match (Single paper)** | Cajthaml et al, 2002[15](#_ENREF_15) |
| 2-(2'-hydroxybenzoyl)-benzoic acid  Oc1ccccc1C(=O)c1ccccc1C([O-])=O | **Perfect Match**  Throughput: 0.015 | Cajthaml et al, 2002[15](#_ENREF_15) |
| dimethyl phthalic acid  COC(=O)c1ccccc1C(=O)OC | **No Match (Single paper)**  Note: EAWAG does not predict methylations | Cajthaml et al, 2002[15](#_ENREF_15) |
| c1ccc2ccccc2c1  Naphthalene | **Partial Match**    Throughput: 0.14  Oc1cc2ccccc2cc1O | Das et al, 2017 [44](#_ENREF_44) Godoy et al, 2016[31](#_ENREF_31)  Tarafdar et al, 2017[26](#_ENREF_26) |
| Cc1ccc2ccccc2c1  Napthalene-2-methyl  2-Methylnaphthalene | **No Match (Multiple papers)** | Das et al, 2017 [44](#_ENREF_44) Godoy et al, 2016[31](#_ENREF_31) |
| Cc1cccc2ccccc12  1-Methylnaphthalene | **No Match (Single paper)** | Godoy et al, 2016[31](#_ENREF_31) |
| Oc1cc2ccccc2cc1O  Dihydroxynaphthalene | **Perfect Match**  Throughput: 0.16 | Chandrasekhar et al, 2010[34](#_ENREF_34)  Hadibarata et al, 2013[10](#_ENREF_10) Baboshin et al, 2005[12](#_ENREF_12)  Hadibarata et al, 2013[45](#_ENREF_45)  Weigand et al, 1999[28](#_ENREF_28) |
| O=c1ccc2ccccc2o1  Coumarin | **No Match (Multiple papers)** | Hadibarata et al, 2013[10](#_ENREF_10)  Hadibarata et al, 2013[45](#_ENREF_45)  Tarafdar et al, 2017[26](#_ENREF_26) |
| [O-]C(=O)Cc1ccccc1  Phenylacetate/Benzeneacetic acid | **Low Throughput Partial Match**    Throughput 0.002296  Oc1ccccc1C(=O)C([O-])=O | Bidi et al, 2018, [33](#_ENREF_33) Das et al, 2017 [44](#_ENREF_44) |
| Ehyl (E)-3-(2-acetyl-4-methoxyphenyl)acrylate  CCOC(=O)/C=C/c1ccc(OC)cc1C(C)=O | **No Match (Single paper)** | Tarafdar et al, 2017[26](#_ENREF_26) |
| image1.png  Oc1ccc(cc1O)C([O-])=O  protocatechuic acid | **Low Throughput Match**  Node Throughput: 0.0011 | van Herwijnen et al, 2003[38](#_ENREF_38) |
| o-phthalic acid  [O-]C(=O)c1ccccc1C([O-])=O | **Perfect Match**  Node Throughput: 0.12 | Ahmed et al, 2012 [3](#_ENREF_3),  Das et al, 2017 [44](#_ENREF_44)  Chandrasekhar et al, 2010[34](#_ENREF_34)  Hadibarata et al, 2012 [4](#_ENREF_4)  Ye et al, 2011, [6](#_ENREF_6)  Richnow et al, 1998[32](#_ENREF_32)  Tarafdar et al, 2017[26](#_ENREF_26)  van Herwijnen et al, 2003[38](#_ENREF_38) |
| 3,4 dihydroxy phlalate  Oc1ccc(C([O-])=O)c(C([O-])=O)c1O | **Perfect Match**  Throughput 0.070 | Brinda et al, 2011[36](#_ENREF_36) |
| O=Cc1ccccc1  Benzaldehyde | **Partial Match**  its immediate precursor (phthalic acid) is predicted and a compound with the same carbon backbone is predicted    Throughput: 0.031  Oc1cccc(C=O)c1O | Das et al, 2017 [44](#_ENREF_44)  Tarafdar et al, 2017[26](#_ENREF_26) |
| Salicylaldehyde  Oc1ccccc1C=O | **Low Throughput Match**  Throughput: 0.000005 | Chandrasekhar et al, 2010[34](#_ENREF_34) |
| [O-]C(=O)c1ccccc1  Benzoic Acid | **Partial Match**    Throughput 0.11 | Hadibarata et al, 2012 [4](#_ENREF_4), Hadibarata et al, 2013[10](#_ENREF_10)  Tarafdar et al, 2017[26](#_ENREF_26) |
| Salicylate/Salicylic Acid  Oc1ccccc1C([O-])=O | **Perfect Match**  Throughput: 0.11 | Chandrasekhar et al, 2010[34](#_ENREF_34) |
| Catechol  Oc1ccccc1O | **Perfect Match**  Throughput: 0.015 | Chandrasekhar et al, 2010[34](#_ENREF_34)  Fulekar et al, 2017[35](#_ENREF_35)  Hadibarata et al, 2012 [4](#_ENREF_4)  Bidi et al, 2018, [33](#_ENREF_33) |
| Phenol  Oc1ccccc1 | **Low Throughput Match**  Throughput 6.766063e-08 | Chandrasekhar et al, 2010[34](#_ENREF_34)  Tarafdar et al, 2017[26](#_ENREF_26) |

**Table S2.3:** Anthracene Empirical Review Summary

|  |  |  |
| --- | --- | --- |
|  | Perfect Match | Partial Match |
| High Throughput | 11 | 11 |
| Low Throughput | 3 | 3 |

No Match (Single) 13

No Match (Multiple) 5

**No Match Explanations:**

Several of the No-Match Compounds are the result of methylation or other straight chain addition after the cleavage of another bound. EAWAG does not predict this type of transition, so these compounds would not be found. Certain anhydride, oxide, or coumarin structures are also not predicted, so these compounds are either classified as no Match or Partial Match if their immediate parent and daughter compounds are predicted.

**References**

(1) Leneva, N.A., et al., Phenanthrene and anthracene degradation by microorganisms of the genus Rhodococcus*.* *Applied Biochemistry and Microbiology*, 2009. **45**(2): p. 169-175.

(2) Moody, J.D., et al., Degradation of phenanthrene and anthracene by cell suspensions of Mycobacterium sp. strain PYR-1*.* *Appl Environ Microbiol*, 2001. **67**(4): p. 1476-83.

(3) Ahmed, A.T., et al., Degradation of Anthracene by Alkaliphilic Bacteria Bacillus badius*.* *Environment and Pollution*, 2012. **1**(2).

(4) Hadibarata, T., A.B. Khudhair, and M.R. Salim, Breakdown products in the metabolic pathway of anthracene degradation by a ligninolytic fungus Polyporus sp. S133*.* *Water, Air, & Soil Pollution*, 2012. **223**(5): p. 2201-2208.

(5) Li, X., et al., Peculiarities of metabolism of anthracene and pyrene by laccase-producing fungus Pycnoporus sanguineus H1*.* *Biotechnol Appl Biochem*, 2014. **61**(5): p. 549-54.

(6) Ye, J.S., et al., Biodegradation of anthracene by Aspergillus fumigatus*.* *J Hazard Mater*, 2011. **185**(1): p. 174-81.

(7) Aranda, E., et al., Isolation of of Ascomycota fungi with capability to transform PAHs: Insights into the biodegradation mechanisms of Penicillium oxalicum*.* *International Biodeterioration & Biodegradation*, 2017. **122**: p. 141-150.

(8) Birolli, W.G., et al., Biodegradation of anthracene and several PAHs by the marine-derived fungus Cladosporium sp CBMAI 1237*.* *Marine Pollution Bulletin*, 2018. **129**(2): p. 525-533.

(9) Gunther, T., et al., Oxidation of PAH and PAH-derivatives by fungal and plant oxidoreductases*.* *Journal of Basic Microbiology*, 1998. **38**(2): p. 113-122.

(10) Hadibarata, T., et al., Degradation and transformation of anthracene by white-rot fungus Armillaria sp F022*.* *Folia Microbiologica*, 2013. **58**(5): p. 385-391.

(11) Hammel, K.E., B. Green, and W.Z. Gai, Ring fission of anthracene by a eukaryote*.* *Proceedings of the National Academy of Sciences of the United States of America*, 1991. **88**(23): p. 10605-10608.

(12) Baboshin, M.A., et al., The microbial transformation of phenanthrene and anthracene*.* *Microbiology*, 2005. **74**(3): p. 303-309.

(13) Bezalel, L., et al., Initial oxidation products in the metabolism of pyrene, anthracene, fluorene, and dibenzothiophene by the white rot fungus Pleurotus ostreatus*.* *Applied and Environmental Microbiology*, 1996. **62**(7): p. 2554-2559.

(14) Bilyk, A. and T.M. Traczewska, Anthracene and phenanthrene degradation by bacteria immobilized on activated carbon and zeolites*.* *Environmental Protection Engineering*, 1999. **25**(3): p. 123-129.

(15) Cajthaml T, et al., Study of fungal degradation products of polycyclic aromatic hydrocarbons using gas chromatography with ion trap mass spectrometry detection*.* *Journal of Chromatography*, 2002. **974**: p. 213–222.

(16) Chigu, N.L., et al., Cytochrome P450 monooxygenases involved in anthracene metabolism by the white-rot basidiomycete Phanerochaete chrysosporium*.* *Applied Microbiology and Biotechnology*, 2010. **87**(5): p. 1907-1916.

(17) Coutino-Gonzalez, E., et al., The earthworm Eisenia fetida accelerates the removal of anthracene and 9, 10-anthraquinone, the most abundant degradation product, in soil*.* *International Biodeterioration & Biodegradation*, 2010. **64**(6): p. 525-529.

(18) Field, J.A., et al., BIODEGRADATION OF POLYCYCLIC AROMATIC-HYDROCARBONS BY NEW ISOLATES OF WHITE ROT FUNGI*.* *Applied and Environmental Microbiology*, 1992. **58**(7): p. 2219-2226.

(19) Guiraud, P., et al., Involvement of Tetrahymena pyriformis and selected fungi in the elimination of anthracene, and toxicity assessment of the biotransformation products*.* *Ecotoxicology and Environmental Safety*, 2008. **69**(2): p. 296-305.

(20) Kastner, M., et al., Formation of bound residues during microbial degradation of C-14 anthracene in soil*.* *Applied and Environmental Microbiology*, 1999. **65**(5): p. 1834-1842.

(21) Lisowska, K. and J. Długoński, Removal of anthracene and phenanthrene by filamentous fungi capable of cortexolone 11-hydroxylation*.* *Journal of Basic Microbiology*, 1999. **39**(2): p. 117-125.

(22) Pathak, H., et al., Technical Note: Degradation of Phenanthrene and Anthracene byPseudomonasStrain, Isolated From Coastal Area*.* *Bioremediation Journal*, 2008. **12**(2): p. 111-116.

(23) Peng, T., et al., Identification of A Ring-Hydroxylating Dioxygenases Capable of Anthracene and Benz a anthracene Oxidization from Rhodococcus sp. P14*.* *Journal of Molecular Microbiology and Biotechnology*, 2018. **28**(4): p. 183-189.

(24) Pizzul, L., M.D. Castillo, and J. Stenstrom, Characterization of selected actinomycetes degrading polyaromatic hydrocarbons in liquid culture and spiked soil*.* *World Journal of Microbiology & Biotechnology*, 2006. **22**(7): p. 745-752.

(25) Sutherland, J.B., et al., IDENTIFICATION OF XYLOSIDE CONJUGATES FORMED FROM ANTHRACENE BY RHIZOCTONIA-SOLANI*.* *Mycological Research*, 1992. **96**: p. 509-517.

(26) Tarafdar, A., A. Sinha, and R.E. Masto, Biodegradation of anthracene by a newly isolated bacterial strain, Bacillus thuringiensis AT.ISM.1, isolated from a fly ash deposition site*.* *Letters in Applied Microbiology*, 2017. **65**(4): p. 327-334.

(27) Traczewska, T.M., Changes of toxicological properties of biodegradation products of anthracene and phenanthrene*.* *Water Science and Technology*, 2000. **41**(12): p. 31-38.

(28) Weigand, H., et al., Desorption controlled mobility and intrinsic biodegradation of anthracene in unsaturated soil*.* *Physics and Chemistry of the Earth Part B-Hydrology Oceans and Atmosphere*, 1999. **24**(6): p. 549-555.

(29) Wu, M.L., et al., Analysis of phenanthrene biodegradation by using FTIR, UV and GC-MS*.* *Spectrochim Acta A Mol Biomol Spectrosc*, 2010. **75**(3): p. 1047-50.

(30) Wu, Y.R., Z.H. Luo, and L.L. Vrijmoed, Biodegradation of anthracene and benz[a]anthracene by two Fusarium solani strains isolated from mangrove sediments*.* *Bioresour Technol*, 2010. **101**(24): p. 9666-72.

(31) Godoy, P., et al., Exploring the potential of fungi isolated from PAH-polluted soil as a source of xenobiotics-degrading fungi*.* *Environmental Science and Pollution Research*, 2016. **23**(20): p. 20985-20996.

(32) Richnow, H.H., et al., The use of 13C-labelled polycyclic aromatic hydrocarbons for the analysis of their transformation in soil*.* *Chemosphere*, 1998. **36**(10): p. 2211-2224.

(33) Bibi, N., et al., Anthracene biodegradation capacity of newly isolated rhizospheric bacteria Bacillus cereus S13*.* *PLoS One*, 2018. **13**(8): p. e0201620.

(34) Chandrasekhar, N. and C.S. Karigar, Bacteral degradation of anthracene by Pseudomonas fluorescens KCP2*.* *Asian Journal of Microbiology, Biotechnology and Environmental Sciences*. **12**(3): p. 591-597.

(35) Fulekar, M.H., Microbial degradation of petrochemical waste-polycyclic aromatic hydrocarbons*.* *Bioresour Bioprocess*, 2017. **4**(1): p. 28.

(36) Brinda, L.M. and M. Velan, Biodegradation of Anthracene Influence Of selected physiochemical parameters and metabolism*.* *Proceedings of the International Conference on Green Technology and Environmental Conservation, GTEC-2011*, 2011: p. 167-172.

(37) da Silva, M., et al., Metabolism of aromatic hydrocarbons by the filamentous fungus Cyclothyrium sp*.* *Chemosphere*, 2004. **57**(8): p. 943-952.

(38) van Herwijnen, R., et al., Degradation of anthracene by Mycobacterium sp strain LB501T proceeds via a novel pathway, through o-phthalic acid*.* *Applied and Environmental Microbiology*, 2003. **69**(1): p. 186-190.

(39) Kim, E.B., et al., Evidence for the role of 2-hydroxychromene-2-carboxylate isomerase in the degradation of anthracene by Sphingomonas yanoikuyae B1*.* *Fems Microbiology Letters*, 1997. **153**(2): p. 479-484.

(40) van Herwijnen, R., et al., Elucidation of the metabolic pathway of fluorene and cometabolic pathways of phenanthrene, fluoranthene, anthracene and dibenzothiophene by Sphingomonas sp LB126*.* *Research in Microbiology*, 2003. **154**(3): p. 199-206.

(41) Menn, F.-M., B.M. Applegate, and G.S. Sayler, NAH Plasmid-Mediated Catabolism of Anthracene and Phenanthrene to Naphthoic Acids*.* *Applied and Environmental Microbiology*, 1993. **59**(6): p. 1938-1942.

(42) Liu, D., et al., Microbial degradation of polycyclic aromatic hydrocarbons and polycyclic aromatic nitrogen heterocyclics*.* *Environmental Toxicology and Water Quality*, 1992. **7**(4): p. 355-372.

(43) Story, S.P., et al., Convergent and divergent points in catabolic pathways involved in utilization of fluoranthene, naphthalene, anthracene, and phenanthrene by Sphingomonas paucimobilis var. EPA505*.* *Journal of Industrial Microbiology & Biotechnology*, 2001. **26**(6): p. 369-382.

(44) Das, M., et al., Enhanced Biodegradation of Anthracene by Bacillus Cereus Strain JMG-01 Isolated from Hydrocarbon Contaminated Soils*.* *Soil and Sediment Contamination: An International Journal*, 2017. **26**(5): p. 510-525.

(45) Hadibarata, T., et al., Microbial transformation and sorption of anthracene in liquid culture*.* *Bioprocess and Biosystems Engineering*, 2013. **36**(9): p. 1229-1233.

# Section 3: Fluorene



**Figure S3.1**: Fluorene High Throughput Plot with Pathway Match Metabolites. SMILES codes for the nodes are presented in Table S3.1.

Figure S3.1 shows a network constructed from only the High Node Throughput compounds. Reconstructing the network with only this subset of compounds yields a much smaller more manageable network.

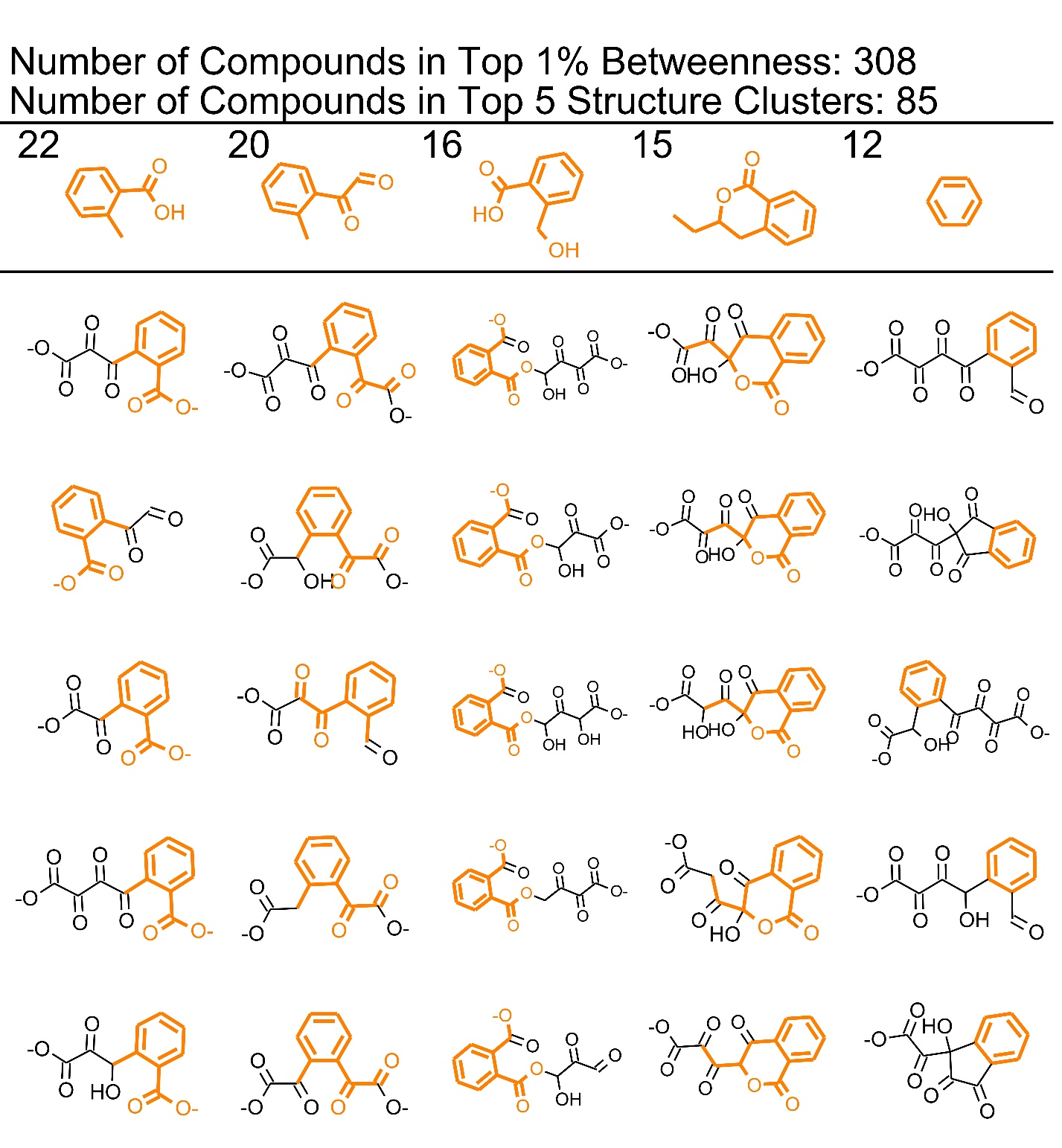
Perfect Match compounds match a literature reference exactly. Pathway Match compounds are High Node Throughput compounds that are found on a degradation pathway to or from a Perfect Match compound. Network Outliers are High Node Throughput compounds that do not have a Perfect Match compound either above or below them in their degradation chains. Table S3.1 contains the SMILES codes for the Pathway Match compounds.

**Table S3.1:** Fluorene Yellow Node (Pathway Matches) Smiles Codes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | [O-]C(=O)C(=O)Cc1ccccc1C([O-])=O |  | 44 | [O-]C(=O)\C=C/C1=CC(=O)c2ccccc12 |
| 2 | [O-]C(=O)Cc1ccccc1C([O-])=O |  | 45 | Oc1ccc2oc(=O)\c(=C\C([O-])=O)\c(=C/C([O-])=O)\c2c1O |
| 3 | Oc1cc2Cc3ccccc3-c2cc1O |  | 46 | Oc1ccc2c(oc(=O)\c(=C\C([O-])=O)\c\2=C/C([O-])=O)c1O |
| 4 | OC(C(=O)C([O-])=O)c1ccccc1C([O-])=O |  | 47 | OC1C(=Cc2ccccc12)C([O-])=O |
| 5 | Oc1ccc2Cc3ccccc3-c2c1O |  | 48 | OC(CC([O-])=O)C1=Cc2ccccc2C1=O |
| 6 | Oc1ccc2-c3ccccc3Cc2c1O |  | 49 | OC1C(\C=C/C([O-])=O)C(=O)c2ccccc12 |
| 7 | OC(C([O-])=O)c1ccccc1C([O-])=O |  | 50 | [O-]C(=O)\C=C/c1cc2ccccc2oc1=O |
| 8 | [O-]C(=O)C(=O)C(=O)c1ccccc1C([O-])=O |  | 51 | OC1(CC(=O)c2ccccc12)\C=C/C([O-])=O |
| 9 | [O-]C(=O)\C=C1\Cc2ccccc2\C\1=C\C([O-])=O |  | 52 | OC(CC([O-])=O)C1=CC(=O)c2ccccc12 |
| 10 | O\C(=C\C=C1\Cc2ccccc2C1=O)C([O-])=O |  | 53 | [O-]C(=O)\C=C/c1cc(=O)oc2ccccc12 |
| 11 | [O-]C(=O)C(=O)c1ccccc1C([O-])=O |  | 54 | [O-]C(=O)\C=C/c1oc(=O)\c(=C\C([O-])=O)\c(=C/C([O-])=O)\c1C([O-])=O |
| 12 | [O-]C(=O)\C=C/C1=C(C([O-])=O)c2ccccc2C1 |  | 55 | [O-]C(=O)\C=C/c1c(oc(=O)\c(=C\C([O-])=O)\c\1=C/C([O-])=O)C([O-])=O |
| 13 | [O-]C(=O)C1Cc2ccccc2C1=O |  | 56 | [O-]C(=O)\C=C/c1c(oc(=O)c2ccccc12)C([O-])=O |
| 14 | O\C(=C\C=C1/C(=O)Cc2ccccc12)C([O-])=O |  | 57 | [O-]C(=O)\C=C/c1c\c(=C\C([O-])=O)\c(=C/C([O-])=O)\c(=O)o1 |
| 15 | [O-]C(=O)\C=C/C1=C(Cc2ccccc12)C([O-])=O |  | 58 | OC(CC([O-])=O)c1oc(=O)\c(=C\C([O-])=O)\c(=C/C([O-])=O)\c1C([O-])=O |
| 16 | OC1\C(=C\C([O-])=O)\C(=C/C([O-])=O)\c2ccccc12 |  | 59 | OC(CC([O-])=O)c1c(oc(=O)\c(=C\C([O-])=O)\c\1=C/C([O-])=O)C([O-])=O |
| 17 | [O-]C(=O)C(=O)\C=C/C1Cc2ccccc2C1=O |  | 60 | [O-]C(=O)\C=C/c1coc(=O)\c(=C\C([O-])=O)\c\1=C/C([O-])=O |
| 18 | OC(CC1Cc2ccccc2C1=O)C(=O)C([O-])=O |  | 61 | [O-]C(=O)C1=Cc2ccccc2C1=O |
| 19 | [O-]C(=O)C(=O)\C=C/C1Cc2ccccc2C(=O)O1 |  | 62 | OC1C(C([O-])=O)C(=O)c2ccccc12 |
| 20 | OC(CC(=O)C([O-])=O)C1Cc2ccccc2C1=O |  | 63 | OC1(C([O-])=O)C(=O)Cc2ccccc12 |
| 21 | [O-]C(=O)\C=C/C1=Cc2ccccc2C1 |  | 64 | Oc1ccc2c(\C=C/C([O-])=O)c(oc(=O)c2c1O)C([O-])=O |
| 22 | OC1C(\C=C/C([O-])=O)=C(C([O-])=O)c2ccccc12 |  | 65 | Oc1ccc2c(c1O)c(\C=C/C([O-])=O)c(oc2=O)C([O-])=O |
| 23 | [O-]C(=O)C(=O)\C=C/C1C(=O)Cc2ccccc12 |  | 66 | O\C(C([O-])=O)=C(\C=C/C([O-])=O)/c1ccccc1C([O-])=O |
| 24 | OC1C(C([O-])=O)=C(\C=C/C([O-])=O)c2ccccc12 |  | 67 | OC(CC([O-])=O)c1c\c(=C\C([O-])=O)\c(=C/C([O-])=O)\c(=O)o1 |
| 25 | OC1C(\C=C/C(=O)C([O-])=O)C(=O)c2ccccc12 |  | 68 | [O-]C(=O)\C=c1\c(C([O-])=O)c(oc(=O)\c\1=C\C([O-])=O)C([O-])=O |
| 26 | OC1(Cc2ccccc2C1=O)\C=C/C(=O)C([O-])=O |  | 69 | OC(CC([O-])=O)c1coc(=O)\c(=C\C([O-])=O)\c\1=C/C([O-])=O |
| 27 | OC1(\C=C/C(=O)C([O-])=O)C(=O)Cc2ccccc12 |  | 70 | Oc1cccc(CCC([O-])=O)c1O |
| 28 | OC(CC1C(=O)Cc2ccccc12)C(=O)C([O-])=O |  | 71 | O=C1Cc2ccccc2CO1 |
| 29 | [O-]C(=O)\C=C1\C(=O)c2ccccc2\C\1=C\C([O-])=O |  | 72 | OC1C(=O)Cc2ccccc12 |
| 30 | OC(CC([O-])=O)C1=Cc2ccccc2C1 |  | 73 | OC1(OC(=O)Cc2ccccc12)C([O-])=O |
| 31 | OC1C(\C=C/C([O-])=O)=Cc2ccccc12 |  | 74 | [O-]C(=O)\C=c1\cc(oc(=O)\c\1=C\C([O-])=O)C([O-])=O |
| 32 | [O-]C(=O)\C=C/C1=C(C([O-])=O)c2ccccc2C1=O |  | 75 | [O-]C(=O)\C=c1\c(coc(=O)\c\1=C\C([O-])=O)C([O-])=O |
| 33 | OC(Cc1ccccc1C([O-])=O)\C=C/C(=O)C([O-])=O |  | 76 | Oc1cccc(C([O-])=O)c1O |
| 34 | [O-]C(=O)C(=O)\C=C/C1OC(=O)Cc2ccccc12 |  | 77 | O=C1Cc2ccccc2C1=O |
| 35 | OC(CC(=O)C([O-])=O)C1C(=O)Cc2ccccc12 |  | 78 | OC1OC(=O)Cc2ccccc12 |
| 36 | [O-]C(=O)\C=C/C1=C(C([O-])=O)C(=O)c2ccccc12 |  | 79 | [O-]C(=O)\C=c1\ccoc(=O)\c\1=C\C([O-])=O |
| 37 | [O-]C(=O)\C=c1\c2ccccc2oc(=O)\c\1=C\C([O-])=O |  | 80 | Oc1cc(C([O-])=O)c(cc1O)C([O-])=O |
| 38 | OC(CC([O-])=O)C1=Cc2ccccc2C1O |  | 81 | Oc1ccc(C([O-])=O)c(C([O-])=O)c1O |
| 39 | [O-]C(=O)\C=C/C1=Cc2ccccc2C1=O |  | 82 | Oc1ccc(C([O-])=O)c(O)c1O |
| 40 | [O-]C(=O)\C=C/c1c(C([O-])=O)c2ccccc2oc1=O |  | 83 | Oc1cc2COC(=O)Cc2cc1O |
| 41 | OC(CC1(O)C(=O)Cc2ccccc12)C(=O)C([O-])=O |  | 84 | Oc1ccc2CC(=O)OCc2c1O |
| 42 | OC(\C=C/C(=O)C([O-])=O)c1ccccc1CC([O-])=O |  | 85 | OCc1ccccc1CC([O-])=O |
| 43 | [O-]C(=O)C1C(=O)Cc2ccccc12 |  | 86 | Oc1ccc2COC(=O)Cc2c1O |
|  |  |  | 87 | O=C1Cc2ccccc2C(=O)O1 |

**High Betweenness Maximum Common Substructures:**

The compounds with the highest betweenness centrality display similar substructures. Code was adapted from the RDKit Cookbook, referenced in the main text. The Tanimto Similarities between the Morgan Fingerprints were first calculated, and then clusters were developed with Butina Clustering using a cutoff of 0.4. For full details and the actual python code, see the github linked in the main text.



**Figure S3.2:** Maximum Common Substructures (MCS) in the Top 1% of Betweenness: Fluorene. The first row shows the MCS in top 5 largest clusters. The numbers indicate how many compounds are contained in each cluster. The compounds in the vertical columns with the substructure highlighted in orange are presented as examples of which compounds are included in the cluster. In the case of fluorene, the same MCS occurs in several different clusters.

**Table S3.2**: Fluorene Empirical Literature Review

|  |  |  |
| --- | --- | --- |
| Structure: | Network Metrics | Empirical Studies finding the same thing |
| Fluorene |  |  |
| Fluorene, 1,4-dihydro-  C1=CCC2=C(C1)Cc1ccccc12 | **Partial Match** due to parent/daughter products | Bharti et al, 2019[1](#_ENREF_1) |
| Dibenzofuran  C1=CC2OC3C=CC=CC3C2C=C1 | **No Match (Multiple papers)** | Yamazoe et al, 2004[2](#_ENREF_2)  Cajthaml et al, 2002[3](#_ENREF_3)  Yu et al, 2014[4](#_ENREF_4) |
| 2-hydroxydibenzofuran  OC1=CC2C(C=C1)OC1C=CC=CC12 | **No Match (Single paper)** | Yamazoe et al, 2004[2](#_ENREF_2) |
| 1,2-dihydroxy-,1,2-dihydrodibenzofuran  OC1C=Cc2oc3ccccc3c2C1O | **No Match (Single paper)** | Yamazoe et al, 2004[2](#_ENREF_2) |
| 9-fluorenol, 9-Hydroxyfluorene  OC1c2ccccc2-c2ccccc12 | **Perfect Match**  Throughput: 0.01 | Zhang et al[5](#_ENREF_5)  Yamazoe et al, 2004[6](#_ENREF_6) Chupungars et al, 2009[7](#_ENREF_7)  Juckpeck et al, 2012[8](#_ENREF_8) Lazim et al, 2016[9](#_ENREF_9)  Liu et al, 2006[10](#_ENREF_10)  Luan et al, 2006[11](#_ENREF_11)  Pozdnyakova et al, 2016[12](#_ENREF_12)  Sokolovska et al, 2002[13](#_ENREF_13)  Finkelstein et al, 2003[14](#_ENREF_14)  Aranda et al, 2010[15](#_ENREF_15)  Bezalel et al, 1996[16](#_ENREF_16)  Birolli et al, 2018[17](#_ENREF_17)  Chapman et al, 1995[18](#_ENREF_18)  Grifoll et al, 1992[19](#_ENREF_19)  Grifoll et al, 1994[20](#_ENREF_20)  Kolomytseva et al[21](#_ENREF_21)  Pothuluri et al, 1993[22](#_ENREF_22)  Reddy et al, 2017[23](#_ENREF_23)  Torres-Farradá et al, 2019[24](#_ENREF_24)  Van Herwijnen et al, 2003[25](#_ENREF_25)  Wischmann et al, 1996[26](#_ENREF_26) |
| 9-Fluorenone  O=C1c2ccccc2-c2ccccc12 | **Perfect Match**  Throughput: 0.01 | Zhang et al[5](#_ENREF_5)  Yamazoe et al, 2004[6](#_ENREF_6) Chupungars et al, 2009[7](#_ENREF_7) Hadibarata et al, 2014[27](#_ENREF_27) Shao et al, 2015,[28](#_ENREF_28)  Luan et al, 2006[11](#_ENREF_11)  Nam et al, 2015[29](#_ENREF_29) Pozdnyakova et al, 2016[12](#_ENREF_12)  Sokolovska et al, 2002[13](#_ENREF_13)  Finkelstein et al, 2003[14](#_ENREF_14)  Aranda et al, 2010[15](#_ENREF_15),  Bezalel et al, 1996[16](#_ENREF_16)  Bressler et al, 2000[30](#_ENREF_30)  Chapman et al, 1995[18](#_ENREF_18)  Grifoll et al, 1992[19](#_ENREF_19)  Grifoll et al, 1994[20](#_ENREF_20)  Pothuluri et al, 1993[22](#_ENREF_22)  Reddy et al, 2017[23](#_ENREF_23)  Shao et al, 2015[28](#_ENREF_28)  Torres-Farradá et al, 2019[24](#_ENREF_24)  Van Herwijnen et al, 2003[25](#_ENREF_25)  Wischmann et al, 1996[26](#_ENREF_26) |
| 2-Hydroxy-9-fluorenone  O=c2c1ccccc1c3ccc(O)cc23 | **Partial Match**    Throughput: 0.01  O=C1c2ccccc2-c2ccccc12 | Finkelstein et al, 2003[14](#_ENREF_14)  Kolomytseva et al[21](#_ENREF_21)  Pothuluri et al, 1993[22](#_ENREF_22) |
| 1,1a-dihydroxy-hydro-fluorenone  O=C1c2ccccc2C2=CC=CC(O)C12O | **Partial Match**    Throughput:0.01  O=C1c2ccccc2-c2ccccc12 | Grifoll et al, 1994[20](#_ENREF_20) |
| 1,2 Dihydroxy-9 fluorenone  Oc1ccc2-c3ccccc3C(=O)c2c1O | **Low Throughput Match**  Throughput: 0.000019  Oc1ccc2-c3ccccc3C(=O)c2c1O | Reddy et al, 2017[23](#_ENREF_23) |
| 8-hydroxy-3,4-benzocoumarin  O=c1oc2c(O)cccc2c2ccccc12 | **No Match (Multiple papers)** | Grifoll et al, 1994[20](#_ENREF_20)  van Herwijnen et al, 2003[25](#_ENREF_25) |
| 2,7-dihydroxyfluorene  OC1=CC2CC3C=C(O)C=CC3C2C=C1 | **Partial Match**    Throughput: 0.33  Oc1ccc2-c3ccccc3Cc2c1O | Finkelstein et al, 2003[14](#_ENREF_14) |
| Oc1cccc2c1Cc1ccccc1-2  1-hydroxyfluorene | **Partial Match**    Throughput: 0.33  Oc1ccc2-c3ccccc3Cc2c1O | Luan et al, 2006[11](#_ENREF_11) |
| Oc1ccc2c(c1)Cc1ccccc1-2  2-hydroxyfluorene | **Partial Match**    Throughput: 0.33  Oc1ccc2-c3ccccc3Cc2c1O | Luan et al, 2006[11](#_ENREF_11)  Finkelstein et al, 2003[14](#_ENREF_14)  Aranda et al, 2010[15](#_ENREF_15)  Birolli et al, 2018[17](#_ENREF_17) |
| Oc1ccc2c(c1)-c1ccccc1C2  3-hydroxyfluorene | **Partial Match**    Throughput: 0.33  Oc1ccc2Cc3ccccc3-c2c1O | Luan et al, 2006[11](#_ENREF_11) |
| O=C1C2C=CC=CC2C2C(O)=CC=CC12  4-hydroxy-9-fluorenone | **Partial Match**    Throughput: 0.01 | Nam et al, 2015[29](#_ENREF_29)  Birolli et al, 2018[17](#_ENREF_17)  Casellas et al[32](#_ENREF_32),  Yamazoe et al, 2004,[6](#_ENREF_6) |
| O=C1OC2C(O)=CC=CC2C2C=CC=CC12  8-hydroxy-3,4-benzocoumarin | **No Match (Multiple papers)** | Nam et al, 2015[29](#_ENREF_29)  Trenz et al, 1994[31](#_ENREF_31) |
| Benzophenone  O=C(c1ccccc1)c1ccccc1 | **No Match (Single paper)** | Birolli et al, 2018[17](#_ENREF_17) |
| 2-carboxy-2',3'-dihydroxy-biphenyl  O=C(O)c1ccccc1-c1cccc(O)c1O | **Perfect Match**  (different protonation state)  Throughput: 0.01  Oc1cccc(c1O)-c1ccccc1C([O-])=O | van Herwijnen et al, 2003[25](#_ENREF_25) |
| Oc1ccccc1-c1cccc(O)c1O  2,3,2'-trihydroxybiphenyl | **Low Throughput Partial Match**    Throughput: 0.0032  Oc1cccc(c1O)-c1cccc(O)c1O | Yamazoe et al, 2004[2](#_ENREF_2) |
| O\C(=C\C=C\C(=O)c1ccccc1O)C([O-])=O  2-Hydroxy-6-(2-hydroxyphenyl)-6-oxo-2,4-hexadienoic acid | **Low Throughput Match**  Throughput: 3.364349e-10  O\C(=C\C=C\C(=O)c1ccccc1O)C([O-])=O  (different protonation state) | Yamazoe et al, 2004[2](#_ENREF_2) |
| O=C(O)C(=O)C=CC=C(O)c1ccccc1O  2-oxo-6-(2’-hydroxyphenyl)-6-hydroxy-3,5-hexadienoic acid | **Low Throughput Partial Match**    Throughput: 3.364349e-10  O\C(=C\C=C\C(=O)c1ccccc1O)C([O-])=O | Yamazoe et al, 2004[2](#_ENREF_2) |
| O=C(O)C(O)=CC=c2oc1ccccc1c2=O  2-hydroxy-4-(3'-oxo-3'H-benzofuran-2'-yliden)but-2-enoic acid | **No Match (Single paper)** | Yamazoe et al, 2004[2](#_ENREF_2) |
| O=C(O)C(=O)C=CC2Oc1ccccc1C2=O  2-oxo-4-(3’-hdyroxy-benzofuran-2’-yl)-but-3-enoic acid | **No Match (Single paper)** | Yamazoe et al, 2004[2](#_ENREF_2) |
| 3-(1-oxo-2,3-dihydro-1H-inden-2-yl) propanoic acid  O=C(O)CCC1Cc2ccccc2C1=O | **Low Throughput Partial Match**    OC(CC1Cc2ccccc2C1=O)C([O-])=O  Throughput: 0.0024 | Zhang et al,[5](#_ENREF_5) |
| β-Indanone-β-hydroxyacetic acid  O=C(O)C(O)=C1Cc2ccccc2C1=O | **Low Throughput Partial Match**    (different protonation state)  [O-]C(=O)\C=C1\Cc2ccccc2C1=O  Throughput: 7.421633e-07 | Finkelstein et al, 2003[14](#_ENREF_14) |
| 2-(1-oxo-2,3-dihydro-1H-inden-2-yl) acetic acid  O=C(O)CC1Cc2ccccc2C1=O | **Low Throughput Match**  (different protonation state)  [O-]C(=O)CC1Cc2ccccc2C1=O  Throughput:  0.00025 | Zhang et al,[5](#_ENREF_5) |
| 1-Indanone  O=C1CCc2ccccc12 | **Perfect Match**  Throughput: 0.093 | Casellas et al[32](#_ENREF_32), Zhang et al,[5](#_ENREF_5) Kristanti et al, 2015,[33](#_ENREF_33) Finkelstein et al, 2003[14](#_ENREF_14) |
| 2,3-Dihydroxy inandone  OC1C(O)c2ccccc2C1=O | **Low Throughput Match**  Throughput: 0.0074 | Reddy et al, 2017[23](#_ENREF_23) |
| O=C1Cc2ccccc2C1  2-Indanone | **Perfect Match**  Throughput: 0.048 | Casellas et al[32](#_ENREF_32), |
| Formyl-indanone  O=CC1CC2C=CC=CC2C1=O | **No Match (Multiple papers)** | Casellas et al[32](#_ENREF_32),  Finkelstein et al, 2003[14](#_ENREF_14) |
| 4-hydroxy-1-tetralone  O=C1CCC(O)c2ccccc21 | **No Match (Single paper)** | Ressler et al, 1999,[34](#_ENREF_34) |
| O=C1CCOc2ccccc21  Chromone | **Partial Match**    Throughput: 0.093  O=C1CCc2ccccc2O1 | Yamazoe et al, 2004[2](#_ENREF_2) |
| O=C1CCc2ccccc2O1  3,4-Dihydroxycumarine | **Perfect Match**  Throughput: 0.093 | Finkelstein et al, 2003[14](#_ENREF_14)  Grifoll et al, 1992[19](#_ENREF_19) |
| 3-(chroman-4-on-2-yl) pyruvate  O=C(O)C(O)=CC1CC(=O)c2ccccc2O1 | **No Match (Single paper)** | Yamazoe et al, 2004[2](#_ENREF_2) |
| 3(2-Hydroxyphenyl)propionate  Oc1ccccc1CCC([O-])=O | **Perfect Match**  Throughput**:** 0.093 | Casellas et al[32](#_ENREF_32), |
| Phthalic Acid  [O-]C(=O)c1ccccc1C([O-])=O | **Perfect Match**  Throughput**:** 0.039 | Yamazoe et al, 2004[6](#_ENREF_6)  Luan et al, 2006[11](#_ENREF_11)  Pozdnyakova et al, 2016[12](#_ENREF_12)  Sokolovska et al, 2002[13](#_ENREF_13)  Grifoll et al, 1994[20](#_ENREF_20)  Reddy et al, 2017[23](#_ENREF_23)  Torres-Farradá et al, 2019[24](#_ENREF_24)  Trenz et al, 1994[31](#_ENREF_31)  van Herwijnen et al, 2003[25](#_ENREF_25)  Yu et al, 2014[4](#_ENREF_4) |
| 2-Carboxybenzaldehyde  O=Cc1ccccc1C(=O)O | **Perfect Match**  [O-]C(=O)c1ccccc1C=O  Throughput  0.027 | Reddy et al, 2017[23](#_ENREF_23) |
| 4,5-dihydroxyphthalate  O=C(O)c1cc(O)c(O)cc1C(=O)O | **Low Throughput Match**  OC(=O)c1cc(O)c(O)cc1C([O-])=O  Throughput  5.596969e-08 | Grifoll et al, 1994[20](#_ENREF_20) |
| Phthalic acid anhydride  O=C1OC(=O)c2ccccc21 | **No Match (Single paper)** | Yamazoe et al, 2004,[6](#_ENREF_6) |
| O=C(O)c1ccc(O)c(O)c1  Protocatechuic acid  3,4-Dihydroxybenzoic acid | **Partial Match**    OC(=O)c1cccc(O)c1O  Throughput: 1.87e-10 | Lazim et al, 2016[9](#_ENREF_9)  Sokolovska et al, 2002[13](#_ENREF_13)  van Herwijnen et al, 2003[25](#_ENREF_25)  Wattiau et al, 2001[35](#_ENREF_35)  Different protonation  Grifoll et al, 1994[20](#_ENREF_20) |
| Catechol  Oc1ccccc1O | **Low Throughput Match**  Throughput: 0.0045 | Hadibarata et al, 2014,[27](#_ENREF_27)  Yamazoe et al, 2004[2](#_ENREF_2)  Reddy et al, 2017[23](#_ENREF_23) |
| Benzoic Acid  O=C(O)c1ccccc1 | **Partial Match**    Throughput: 0.031 | Reddy et al, 2017[23](#_ENREF_23) |
| 4-Hydroxybenzoic acid  O=C(O)c1ccc(O)cc1 | **Low Throughput Partial Match**    Throughput: 0.00035  Oc1ccc(cc1O)C([O-])=O | Reddy et al, 2017[23](#_ENREF_23) |
| 3,4-Dihydroxybenzoate  Protocatechuic acid  Oc1ccc(cc1O)C([O-])=O | **Low Throughput Match**  Throughput: 0.00035 | Torres-Farradá et al, 2019[24](#_ENREF_24) |
| Salicylic Acid  Salicylate  Oc1ccccc1C([O-])=O | **Perfect Match**  Throughput:  0.031 | Shao et al, 2015,[28](#_ENREF_28)  Yamazoe et al, 2004[2](#_ENREF_2) Casellas et al[32](#_ENREF_32), Kristanti et al, 2015,[33](#_ENREF_33) Hadibarata et al, 2014,[27](#_ENREF_27)  Shao et al, 2015[28](#_ENREF_28) |
| 1-hydroxybenzene  Phenol  Oc1ccccc1 | **Low Throughput Match**  Throughput: 2.199903e-08 | Yu et al, 2014[4](#_ENREF_4) |

**Table S3.3** Fluorene Empirical Review Summary

|  |  |  |
| --- | --- | --- |
|  | Perfect Match | Partial Match |
| High Throughput | 10 | 11 |
| Low Throughput | 8 | 5 |

No Match (Single) 8

No Match (Multiple) 4

**No Match Explanations:**

Several of the No-Match Compounds are the result of methylation or other straight chain addition after the cleavage of another bound. EAWAG does not predict this type of transition, so these compounds would not be found. Certain anhydride, oxide, or coumarin structures are also not predicted, so these compounds are either classified as no Match or Partial Match if their immediate parent and daughter compounds are predicted.

**References**

(1) Bharti, V., B. Gupta, and J. Kaur, Novel Bacterial Strains Pseudomonas sp. and Bacillus sp. Isolated from Petroleum Oil Contaminated Soils for Degradation of Flourene and Phenanthrene*.* *Pollution*, 2019. **5**(3): p. 657-669.

(2) Yamazoe, A., O. Yagi, and H. Oyaizu, Degradation of polycyclic aromatic hydrocarbons by a newly isolated dibenzofuran-utilizing Janibacter sp. strain YY-1*.* *Appl Microbiol Biotechnol*, 2004. **65**(2): p. 211-8.

(3) Cajthaml T, et al., Study of fungal degradation products of polycyclic aromatic hydrocarbons using gas chromatography with ion trap mass spectrometry detection*.* *Journal of Chromatography*, 2002. **974**: p. 213–222.

(4) Yu, C., J. Yao, and J. Jin, Characteristics and metabolic pathways of fluorene (FLU)degradation by strain Rhodococcussp. USTB-C isolated from crude oil*.* *Journal of Chemical and Pharmaceutical Research*, 2014. **6**(3): p. 560-565.

(5) Zhang, Z.Z., et al., Degradation of n-alkanes and polycyclic aromatic hydrocarbons in petroleum by a newly isolated Pseudomonas aeruginosa DQ8*.* *Bioresource Technology*, 2011. **102**(5): p. 4111-4116.

(6) Yamazoe, A., O. Yagi, and H. Oyaizu, Biotransformation of fluorene, diphenyl ether, dibenzo-p-dioxin and carbazole by Janibacter sp*.* *Biotechnology Letters*, 2004. **26**(6): p. 479-486.

(7) Chupungars, K., P. Rerngsamran, and S. Thaniyavarn, Polycyclic aromatic hydrocarbons degradation by Agrocybe sp. CU-43 and its fluorene transformation*.* *International Biodeterioration & Biodegradation*, 2009. **63**(1): p. 93-99.

(8) Juckpech, K., O. Pinyakong, and P. Rerngsamran, Degradation of polycyclic aromatic hydrocarbons by newly isolated Curvularia sp F18, Lentinus sp S5, and Phanerochaete sp T20*.* *Scienceasia*, 2012. **38**(2): p. 147-156.

(9) Lazim, Z.M. and T. Hadibarata, Ligninolytic fungus Polyporus sp S133 mediated metabolic degradation of fluorene*.* *Brazilian Journal of Microbiology*, 2016. **47**(3): p. 610-616.

(10) Liu, L., R.D. Schmid, and V.B. Urlacher, Cloning, expression, and characterization of a self-sufficient cytochrome P450 monooxygenase from Rhodococcus ruber DSM 44319*.* *Applied Microbiology and Biotechnology*, 2006. **72**(5): p. 876-882.

(11) Luan, T.G., et al., Study of metabolites from the degradation of polycyclic aromatic hydrocarbons (PAHs) by bacterial consortium enriched from mangrove sediments*.* *Chemosphere*, 2006. **65**(11): p. 2289-96.

(12) Pozdnyakova, N.N., et al., Degradation of Fluorene and Fluoranthene by the Basidiomycete Pleurotus ostreatus*.* *Applied Biochemistry and Microbiology*, 2016. **52**(6): p. 621-628.

(13) Sokolovska, I., et al., Biodegradation of fluorene at low temperature by a psychrotrophic Sphingomonas sp L-138*.* *Chemical Papers-Chemicke Zvesti*, 2002. **56**(1): p. 36-40.

(14) Finkelstein, Z.I., et al., Fluorene transformation by bacteria of the genus Rhodococcus*.* *Microbiology*, 2003. **72**(6): p. 660-665.

(15) Aranda, E., R. Ullrich, and M. Hofrichter, Conversion of polycyclic aromatic hydrocarbons, methyl naphthalenes and dibenzofuran by two fungal peroxygenases*.* *Biodegradation*, 2010. **21**(2): p. 267-281.

(16) Bezalel, L., et al., Initial oxidation products in the metabolism of pyrene, anthracene, fluorene, and dibenzothiophene by the white rot fungus Pleurotus ostreatus*.* *Applied and Environmental Microbiology*, 1996. **62**(7): p. 2554-2559.

(17) Birolli, W.G., et al., Biodegradation of anthracene and several PAHs by the marine-derived fungus Cladosporium sp CBMAI 1237*.* *Marine Pollution Bulletin*, 2018. **129**(2): p. 525-533.

(18) Chapman, P.J., et al., Fossil fuel biodegradation: Laboratory studies*.* *Environmental Health Perspectives*, 1995. **103**: p. 79-83.

(19) Grifoll, M., et al., ISOLATION AND CHARACTERIZATION OF A FLUORENE-DEGRADING BACTERIUM - IDENTIFICATION OF RING OXIDATION AND RING FISSION-PRODUCTS*.* *Applied and Environmental Microbiology*, 1992. **58**(9): p. 2910-2917.

(20) Grifoll, M., S.A. Selifonov, and P.J. Chapman, EVIDENCE FOR A NOVEL PATHWAY IN THE DEGRADATION OF FLUORENE BY PSEUDOMONAS SP STRAIN F274*.* *Applied and Environmental Microbiology*, 1994. **60**(7): p. 2438-2449.

(21) Kolomytseva, M.P., et al., Role of surfactants in optimizing fluorene assimilation and intermediate formation by Rhodococcus rhodochrous VKM B-2469*.* *Bioresource Technology*, 2009. **100**(2): p. 839-844.

(22) Pothuluri, J.V., et al., BIOTRANSFORMATION OF FLUORENE BY THE FUNGUS CUNNINGHAMELLA-ELEGANS*.* *Applied and Environmental Microbiology*, 1993. **59**(6): p. 1977-1980.

(23) Reddy, P.V., et al., Catabolism of fluorene through 2,3-dihydroxy indanone in Paenibacillus sp PRNK-6*.* *International Biodeterioration & Biodegradation*, 2017. **123**: p. 156-163.

(24) Torres-Farradá, G., et al., Biodegradation of polycyclic aromatic hydrocarbons by native Ganoderma sp. strains: identification of metabolites and proposed degradation pathways*.* *Applied Microbiology and Biotechnology*, 2019.

(25) van Herwijnen, R., et al., Elucidation of the metabolic pathway of fluorene and cometabolic pathways of phenanthrene, fluoranthene, anthracene and dibenzothiophene by Sphingomonas sp LB126*.* *Research in Microbiology*, 2003. **154**(3): p. 199-206.

(26) Wischmann, H., et al., Degradation of selected PAHs in soil/compost and identification of intermediates*.* *International Journal of Environmental Analytical Chemistry*, 1996. **64**(4): p. 247-255.

(27) Hadibarata, T. and R.A. Kristanti, Fluorene biodegradation and identification of transformation products by white-rot fungus Armillaria sp F022*.* *Biodegradation*, 2014. **25**(3): p. 373-382.

(28) Shao, Y.X., et al., Biodegradation of PAHs by Acinetobacter isolated from karst groundwater in a coal-mining area*.* *Environmental Earth Sciences*, 2015. **73**(11): p. 7479-7488.

(29) Nam, I.-H., et al., Effects of Heavy Metals on Biodegradation of Fluorene by a Sphingobacterium sp. Strain (KM-02) Isolated from Polycyclic Aromatic Hydrocarbon-Contaminated Mine Soil*.* *Environmental Engineering Science*, 2015. **32**(10): p. 891-898.

(30) Bressler, D.C., P.M. Fedorak, and M.A. Pickard, Oxidation of carbazole, N-ethylcarbazole, fluorene, and dibenzothiophene by the laccase of Coriolopsis gallica*.* *Biotechnology Letters*, 2000. **22**(14): p. 1119-1125.

(31) Trenz, S.P., et al., DEGRADATION OF FLUORENE BY BREVIBACTERIUM SP STRAIN DPO-1361 - A NOVEL C-C BOND-CLEAVAGE MECHANISM VIA 1,10-DIHYDRO-1,10-DIHYDROXYFLUOREN-9-ONE*.* *Journal of Bacteriology*, 1994. **176**(3): p. 789-795.

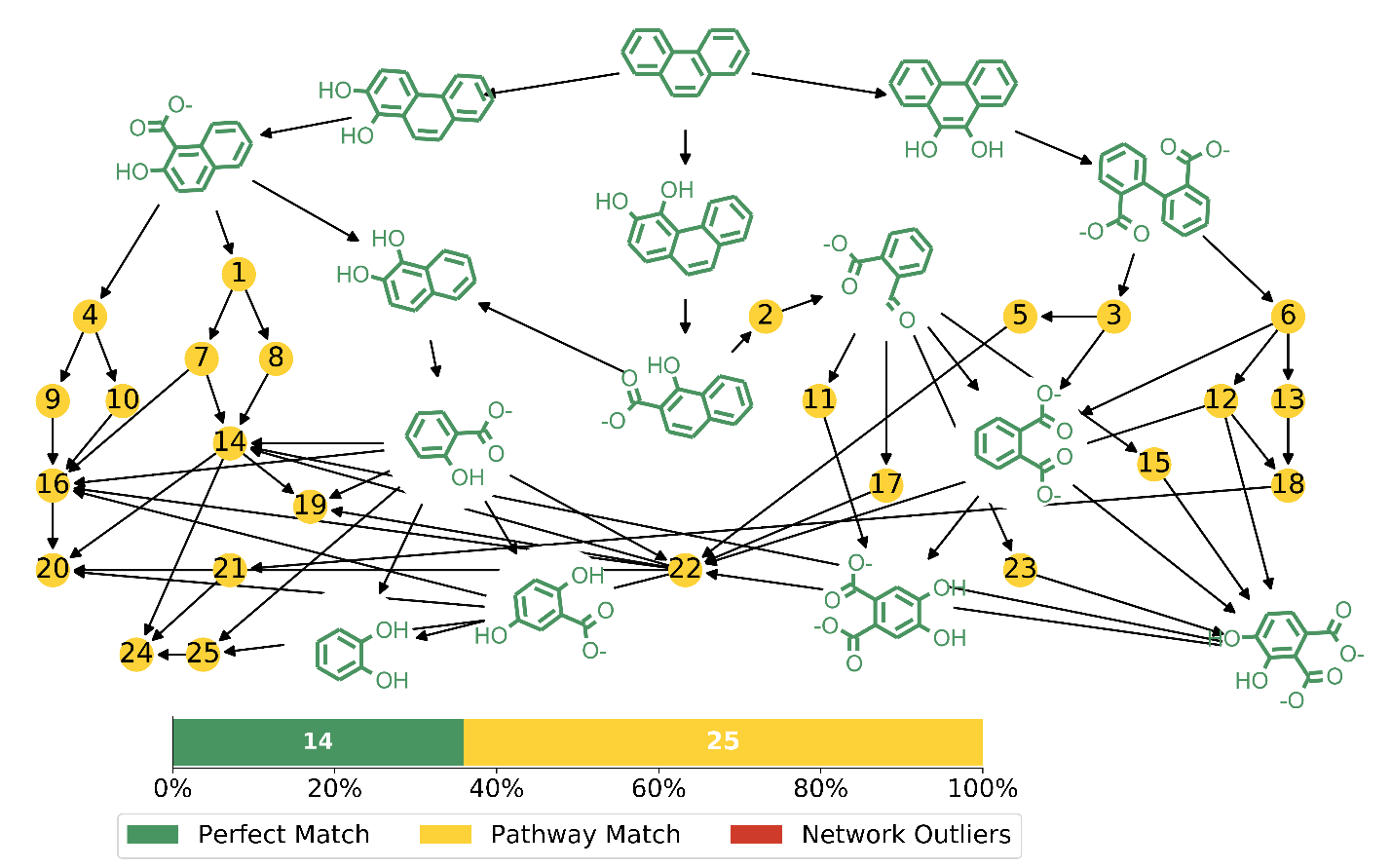
(32) Casellas, M., et al., New metabolites in the degradation of fluorene by Arthrobacter sp strain F101*.* *Applied and Environmental Microbiology*, 1997. **63**(3): p. 819-826.

(33) Kristanti, R.A. and T. Hadibarata, Biodegradation and Identification of Transformation Products of Fluorene by Ascomycete Fungi*.* *Water, Air, & Soil Pollution*, 2015. **226**(12).

(34) Ressler, B.P., H. Kneifel, and J. Winter, Bioavailability of polycyclic aromatic hydrocarbons and formation of humic acid-like residues during bacterial PAH degradation*.* *Applied Microbiology and Biotechnology*, 1999. **53**(1): p. 85-91.

(35) Wattiau, P., et al., Fluorene degradation by Sphingomonas sp LB126 proceeds through protocatechuic acid: a genetic analysis*.* *Research in Microbiology*, 2001. **152**(10): p. 861-872.

# Section 4: Phenanthrene

****

**Figure S4.1:** Phenanthrene High Throughput Plot with Pathway Match Metabolite. SMILES codes for the nodes are presented in Table S1.1.

Figure S4.1 shows a network constructed from only the High Node Throughput compounds. Reconstructing the network with only this subset of compounds yields a much smaller more manageable network.

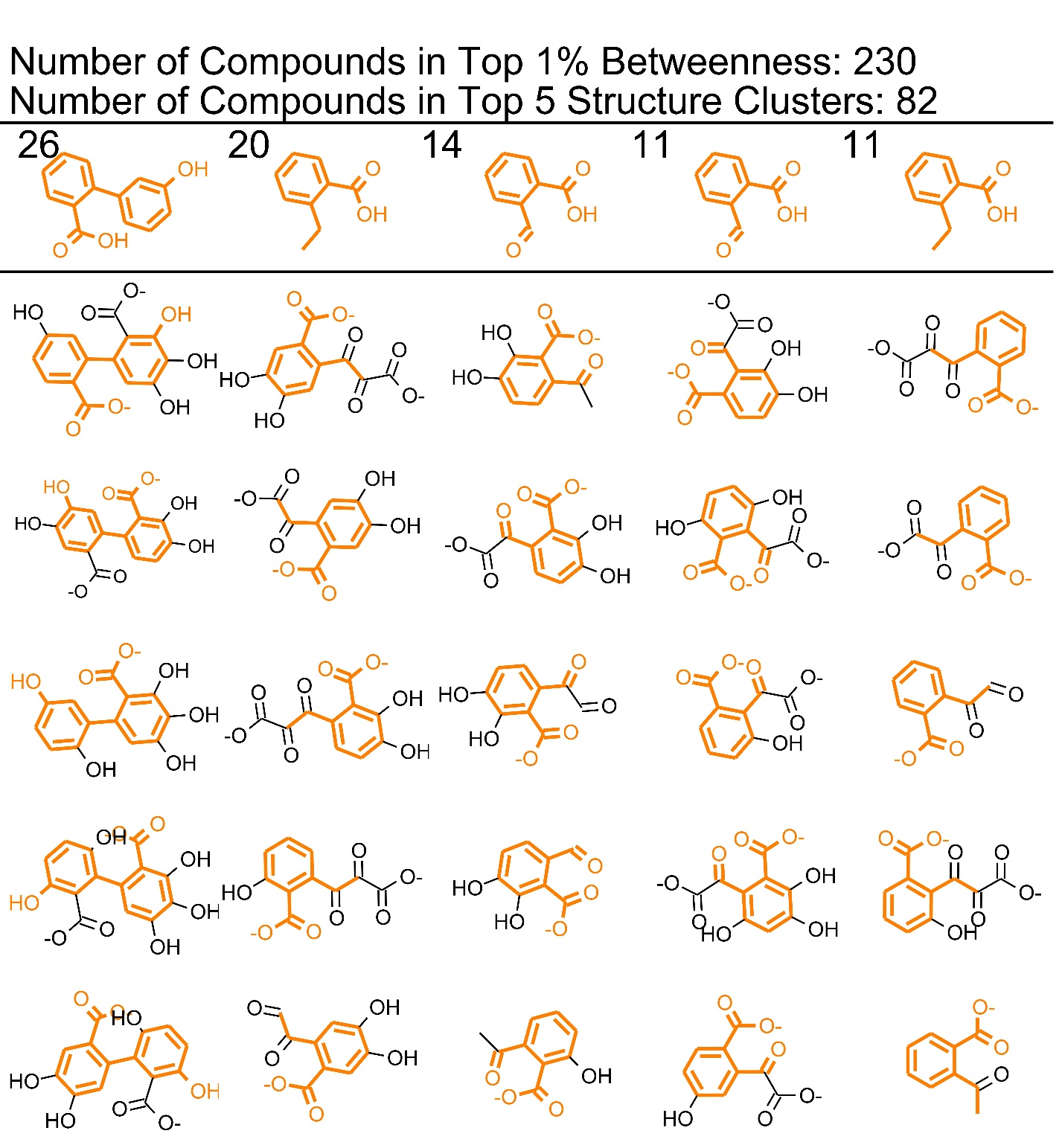
Perfect Match compounds match a literature reference exactly. Pathway Match compounds are High Node Throughput compounds that are found on a degradation pathway to or from a Perfect Match compound. Network Outliers are High Node Throughput compounds that do not have a Perfect Match compound either above or below them in their degradation chains. Table S4.1 contains the SMILES codes for the Pathway Match compounds.

**Table S4.1:** Phenanthrene Yellow Node (Pathway Matches) Smiles Codes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | Oc1ccc2ccc(O)c(C([O-])=O)c2c1O |  | 13 | Oc1cc(O)c(O)c(c1O)-c1ccccc1C([O-])=O |
| 2 | [O-]C(=O)C(=O)\C=C\c1ccccc1C([O-])=O |  | 14 | Oc1ccc(C([O-])=O)c(O)c1O |
| 3 | Oc1cccc(c1O)-c1ccccc1C([O-])=O |  | 15 | Oc1ccc(C([O-])=O)c(C=O)c1O |
| 4 | Oc1ccc2c(C([O-])=O)c(O)ccc2c1O |  | 16 | Oc1ccc(O)c(C([O-])=O)c1O |
| 5 | Oc1cccc(c1O)-c1cccc(O)c1O |  | 17 | Oc1cccc(C=O)c1O |
| 6 | Oc1ccc(C([O-])=O)c(c1O)-c1ccccc1C([O-])=O |  | 18 | Oc1cccc(c1O)-c1c(O)c(O)cc(O)c1O |
| 7 | Oc1ccc(C([O-])=O)c(O)c1C([O-])=O |  | 19 | Oc1cccc(O)c1O |
| 8 | Oc1ccc2ccc(O)c(O)c2c1O |  | 20 | Oc1ccc(O)c(O)c1O |
| 9 | Oc1ccc2c(O)c(O)ccc2c1O |  | 21 | Oc1cc(O)c(O)c(C([O-])=O)c1O |
| 10 | Oc1ccc(O)c(C([O-])=O)c1C([O-])=O |  | 22 | Oc1cccc(C([O-])=O)c1O |
| 11 | Oc1cc(C=O)c(cc1O)C([O-])=O |  | 23 | Oc1ccc(C=O)c(C([O-])=O)c1O |
| 12 | Oc1cccc(c1O)-c1c(O)c(O)ccc1C([O-])=O |  | 24 | Oc1cc(O)c(O)c(O)c1O |
|  |  |  | 25 | Oc1cc(O)c(cc1O)C([O-])=O |

**High Betweenness Maximum Common Substructures:**

The compounds with the highest betweenness centrality display similar substructures. Code was adapted from the RDKit Cookbook, referenced in the main text. The Tanimto Similarities between the Morgan Fingerprints were first calculated, and then clusters were developed with Butina Clustering using a cutoff of 0.4. For full details and the actual python code, see the github linked in the main text.



**Figure S4.2:** Maximum Common Substructures (MCS) in the Top 1% of Betweenness: Phenanthrene. The first row shows the MCS in top 5 largest clusters. The numbers indicate how many compounds are contained in each cluster. The compounds in the vertical columns with the substructure highlighted in orange are presented as examples of which compounds are included in the cluster. In the case of phenanthrene, the same MCS occurs in several different clusters.

**Table S4.2**: Phenanthrene Empirical Literature Review

|  |  |  |
| --- | --- | --- |
| Structure: | Network Metrics | Empirical Studies finding the same thing |
| Phenanthrene |  |  |
| 1,2-Dihydroxyphenanthrene, cis-phenanthrene dihydrodiol  Oc1ccc2c(ccc3ccccc23)c1O | **Perfect Match**  Throughput 0.33 | Gao et al, 2013, 1  Ghosal et al, 20102  Guerin et al, 19983  Lee et al, 2009,4  Lin et al, 2014,5  Lu et al, 2013,6  Keum et al, 20067  Kim et al, 20058  Kweon et al, 20109  Lisowska et al, 200610  Mallick et al, 2007,11  Roy et al, 2012,12  Sack et al, 199713  Seo et al, 200714  Seo et al, 2009,15  Seo et al, 2012,16  van Herwijnen et al, 200317 |
| 1,2 Phenanthrene Oxide  C1=CC2OC2c2ccc3ccccc3c21 | **Partial Match**  also immediate parent/daughter predicted    Throughput 0.33  Oc1ccc2c(ccc3ccccc23)c1O | Nie et al, 2016,18 |
| 1-Hydroxyphenanthrene,  1-phenanthrol;  Oc1cccc2c1ccc1ccccc12 | **Partial Match**  also immediate parent/daughter predicted    Oc1ccc2c(ccc3ccccc23)c1O  Throughput: 0.33 | Aranda et al, 201019  Bourguignon et al, 201920  Casillas et al, 199621  da Silva et al, 200422  Hesham et al, 2014,23  Liu et al, 1992,24  Liu et al, 199224  Sack et al, 199725  Sack et al, 199713  Schrlau et al, 201726  Zhong et al, 201127, |
| 2-phenanthrol  Oc1ccc2c(ccc3ccccc32)c1 | **Partial Match**  also immediate parent/daughter predicted    Oc1ccc2c(ccc3ccccc23)c1O  Throughput: 0.33 | Bourguignon et al, 201920  Casillas et al, 199621  Sack et al, 199725  Sack et al, 199713 |
| 3,4-dihydroxyphenanthrene  Oc1ccc2ccc3ccccc3c2c1O | **Perfect Match**  Throughput:0.33 | Allen et al, 199928  Casillas et al, 199621  da Silva et al, 200422  Gao et al, 2013,1  Keum et al, 20067  Kweon et al, 20109  Lisowska et al, 200610  Pathak et al, 200829  Rehmann et al, 199630  Roy et al, 2012,12  Sack et al, 199713  Seo et al, 200714  Seo et al, 2009,15  Seo et al, 2012,16  Stingley et al, 200431  Story et al, 200132  Sun et al, 2019,33  Sutherland et al, 199134  Umar et al, 201835  Umar et al, 2017,36  Zeinali et al, 200837  Zhang et al, 201038 |
| 4-phenanthrol  Oc1cccc2ccc3ccccc3c12 | **Partial Match**    Throughput:0.33  Oc1ccc2ccc3ccccc3c2c1O | Aranda et al, 201019  Casillas et al, 199621  da Silva et al, 200422  Sack et al, 199713  Schrlau et al, 2017, 26  Sutherland et al, 199134  Zhong et al, 201127 |
| 4-methoxy phenanthrene  COc1cccc2ccc3ccccc3c12 | **No Match (Single paper)** | Bourguignon et al, 201920 |
| 1-methoxy phenanthrene  COc1cccc2c1ccc1ccccc12 | **No Match (Single paper)** | Sack et al, 199725 |
| 3H-benzo[f]chromene-2-carboxylic acid  O=C(O)C1=Cc2c(ccc3ccccc23)OC1 | **No Match (Single paper)** | Safoneva et al, 200539 |
| 3-hydroxyphenanthrene;  3-phenanthrol  Oc1ccc2ccc3ccccc3c2c1 | **Partial Match**  Oc1ccc2ccc3ccccc3c2c1O  Throughput: 0.33 | Aranda et al, 201019 Baboshin et al, 200540 Bourguignon et al, 201920  Casillas et al, 199621  da Silva et al, 200422  Leneva et al 2009,41  Rehmann et al, 199630  Sack et al, 199713  Schrlau et al, 2017, 26  Sutherland et al, 199134  Zhong et al, 201127, |
| 3,4 phenanthrene oxide  C1=CC2OC2c2c1ccc1ccccc21 | **Partial Match**    Oc1ccc2ccc3ccccc3c2c1O  Throughput: 0.33 | Nie et al, 2016,18 |
| 9,10 phenanthrene oxide  c1ccc2c(c1)-c1ccccc1C1OC21 | **Partial Match**  Oc1c(O)c2ccccc2c2ccccc12  Throughput: 0.33 | Bezalel et al, 199642 |
| 9,10-Dihydroxyphenanthrene  cis-9,10-Dihydrophenanthrene-9,10-diol  phenanthrene 9,10-dihydrodiol  Oc1c(O)c2ccccc2c2ccccc12 | **Perfect Match**  Throughput: 0.33 | Bezalel et al, 199643, Bezalel et al, 199642  Bourguignon et al, 201920  Cajthaml et al, 200244  Casillas et al, 199621  Gao et al, 2013, 1  Fu et al, 201845  Hennessee et al,46  Hidayat et al, 2018, 47  Kim et al, 20058  Kweon et al, 20109  Lisowska et al, 200610  Luan et al 2006,48  Muratova et al, 2014, 49,  Mishra et al, 201950  Ning et al, 201051  Pathak et al, 200829  Rehmann et al, 199630  Sack et al, 199713  Sack et al, 199752  Safoneva et al, 200539  Schrlau et al, 2017,26  Seo et al, 200653  Seo et al, 2012,16 Sun et al, 2019,33  Sutherland et al, 199134  Sutherland et al, 199054  Tatarko et al, 199355  van Herwijnen et al, 2003,17 Yang et al, 201456 Zeinali et al, 200837, |
| 9-phenanthrol  Oc1cc2ccccc2c2ccccc12 | **Partial Match**    Oc1c(O)c2ccccc2c2ccccc12  Throughput: 0.33 | Bezalel et al, 199642  Bourguignon et al, 201920  Capotorti et al, 2005,57  Casillas et al, 199621  Fu et al, 201845  Jin et al, 201658,  Schrlau et al, 2017, 26  Sutherland et al, 199054  Sutherland et al, 199134  Sun et al, 2019,33  Szewczyk et al, 200359  Tatarko et al, 199355  van Herwijnen et al, 200317  Yang et al, 201456  Zhang et al, 201038  Zhong et al, 201127 |
| 9-methoxyphenanthrene  COc1cc2ccccc2c2ccccc12 | **No Match (Single paper)** | Cajthaml et al, 200244 |
| 1-phenanthrene carboxylic acid  O=C(O)c1cccc2c1ccc1ccccc12 | **No Match (Single paper)** | Mishra et al, 201950 |
| 1-Methoxy-2-hydroxyphenanthrene  COc1c(O)ccc2c1ccc1ccccc12 | **No Match (Single paper)** | Kim et al, 20058 |
| 2-hydroxy-7-methoxy-phenanthrene  COc1ccc2c(ccc3cc(O)ccc32)c1 | **No Match (Single paper)** | da Silva et al, 200422 |
| 9,10-phenanthrenequinone  O=C1C(=O)c2ccccc2-c2ccccc12 | **Partial Match**    Oc1c(O)c2ccccc2c2ccccc12  Throughput:0.33 | Fu et al, 201845  Hadibarata et al, 201060  Hadibarata et al, 201161  Li et al, 201862  Muratova et al, 2014, 49  van Herwijnen et al, 200317 |
| 5,6-Benzocoumarin  O=c1ccc2c(ccc3ccccc32)o1 | **No Match (Multiple papers)** | Ghosal et al, 20102  Hennessee et al,46  Keum et al, 20067  Mallick et al, 2007,11  Nie et al, 2016,18  Pinyakong et al, 2000,63  Roy et al, 2012,12  Seo et al, 200653  Seo et al, 200714  Seo et al, 2012,16 |
| 7,8-Benzocoumarin  O=c1ccc2ccc3ccccc3c2o1 | **No Match** **(Multiple papers)** | Baboshin et al, 200540  Hennessee et al,46  Keum et al, 20067  Leneva et al, 2009, 41  Muratova et al, 2014, 49  Nie et al, 2016,18  Pinyakong et al, 2000,63  Roy et al, 2012,12  Seo et al, 200653  Seo et al, 200714  Seo et al, 2012,16  Sun et al, 2019,33, |
| 3,4-benzocoumarin  6H-Benzo[c]chromen-6-one  O=c1oc2ccccc2c2ccccc12 | **No Match (Multiple papers)** | Seo et al, 2012,16  Cajthaml et al, 200244  Wang et al, 201664  \*Cajthaml suggested that this structure forms from another due to dehydration in the GM/MS process |
| 8-Hydroxy-3,4-benzocoumarin  O=c1oc2c(O)cccc2c2ccccc12 | **No Match (Single paper)** | Seo et al, 2012,16 |
| 2-dibenzofuranol  Oc1ccc2oc3ccccc3c2c1 | **No Match (Single paper)** | Li et al, 201862 |
| 2-(2-Carboxy-vinyl)-1-naphthoic acid  O=C(O)C=Cc1ccc2ccccc2c1C(=O)O | **No Match (Multiple papers)** | Hennessee et al,46 Seo et al, 2012,16 |
| 1-[(E)-2-carboxyvinyl]-2-naphthoic acid  O=C(O)/C=C/c1c(C(=O)O)ccc2ccccc12 | **No Match (Single paper)** | Seo et al, 2012,16 |
| 4-(1-Hydroxynaphthalen-2-yl)-2-oxobut-3-enoic acid  O=C(O)C(=O)C=Cc1ccc2ccccc2c1O | **No Match (Multiple papers)** | Adebusuyi et al, 201265  Keum et al, 200866  Seo et al, 2012,16  Zeinali et al, 200837 |
| Naphthalene-1,2-dicarboxylic acid,  1,2-Dicarboxynaphthalene  O=C(O)c1ccc2ccccc2c1C(=O)O | **No Match (Multiple papers)** | Hennessee et al,46  Keum et al, 20067  Kim et al, 2005,8  Mallick et al, 2007,11  Seo et al, 200653  Seo et al, 2012,16 |
| Naphthalene-1,2-dicarboxylic acid anhydride  2,3-naphthalic anhydride  O=C1OC(=O)c2c1ccc1ccccc21 | **No Match (Multiple papers)** | Cajthaml et al, 200244  Hennessee et al,46 |
| 2-hydroxy-2,3-dihydronaphtho[ 2,1-b]furan  OC1Cc2c(ccc3ccccc23)O1 | **No Match (Single paper)** | Nie et al, 2016,18 |
| Benzoic acid, 2-benzoyl-, methyl ester  COC(=O)c1ccccc1C(=O)c1ccccc1 | **No Match (Single paper)** | Luan et al 2006,48 |
| 2,2′-Diphenic Acid  [O-]C(=O)c1ccccc1-c1ccccc1C([O-])=O | **Perfect Match**  Throughput: 0.33 | Arias et al 200867, Bezalel et al, 199643.  Fu et al, 201845  Hidayat et al, 2018, 47,  Kim et al, 2005,8  Hadibarata et al, 201060  Hadibarata et al, 201161  Mishra et al, 201950  Moody et al, 2001,68  Muratova et al, 2014, 49,  Rehmann et al, 199630  Seo et al, 200653  Seo et al, 2012,16  Sun et al, 2019,33  Zeinali et al, 200837, |
| 2,2'-biphenyldimethanol  OCc1ccccc1-c1ccccc1CO | **Partial Match**    Throughput: 0.33  [O-]C(=O)c1ccccc1-c1ccccc1C([O-])=O | Hammel et al, 1992, 69 |
| 2,2’ Dihydroxybenzophenone  O=C(c1ccccc1O)c1ccccc1O | **No Match (Single paper)** | Wang et al, 201664 |
| Biphenyl-2-carboxylic acid  O=C(O)c1ccccc1-c1ccccc1 | **Partial Match**    Throughput: 0.16  Oc1cccc(c1O)-c1ccccc1C([O-])=O | Seo et al, 2012,16 |
| 2'-hydroxy-2-carboxy biphenyl  O=C(O)c1ccccc1-c1ccccc1O | **Partial Match**    Throughput: 0.16  Oc1cccc(c1O)-c1ccccc1C([O-])=O | Cajthaml et al, 200244 |
| Dimethyl-2-6-napthalenedicarboxylate  COC(=O)c1ccc2cc(C(=O)OC)ccc2c1 | **No Match (Single paper)** | Wang et al, 201664 |
| 2-hydroxy-1-napthoic acid methyl ester,  Methyl 2-hydroxy-1-naphthoate  COC(=O)c1c(O)ccc2ccccc12 | **No Match (Multiple papers)** | Lu et al, 2013,6  Mallick et al, 2007,11 Roy et al, 2012,12 |
| 3-naphthyl allyl alcohol  OCC=Cc1cccc2ccccc12 | **No Match (Multiple papers)** | Nie et al, 2016,18  Nzila et al, 201870 |
| 1-[(E)-2-Carboxyvinyl]-2-naphthoic acid  O=C(O)/C=C/c1c(C(=O)O)ccc2ccccc12 | **No Match (Single paper)** | Seo et al, 200653 |
| 2-[(E)-2-Carboxyvinyl]-1-naphthoic acid  O=C(O)/C=C/c1ccc2ccccc2c1C(=O)O | **No Match (Single paper)** | Seo et al, 200653 |
| 1-hydroxy-2-naphthaldehyde, 1-hydroxynaphthalene-  2-carbaldehyde  Oc1c(C=O)ccc2ccccc12 | **Partial Match**    Throughput: 0.17  Oc1c(ccc2ccccc12)C([O-])=O | Keum et al, 20067  Leneva et al, 2009, 41  Nie et al, 2016,18  Zhang et al, 201038 |
| 2-hydroxynaphthalene-1-carbaldehyde  O=Cc1c(O)ccc2ccccc12 | **Partial Match**    Throughput: 0.17  Oc1ccc2ccccc2c1C([O-])=O | Nie et al, 2016,18 |
| 1,2-dihydroxynaphthalene; naphthalene-1,2-diol  Oc1ccc2ccccc2c1O | **Perfect Match**  Throughput 0.14 | Keum et al, 20067  Prabhu et al, 2003,71  Prakash and Lal, 2013,72  Seo et al, 200653  Seo et al, 200714  Seo et al, 2012,16  Story et al, 200132 |
| 2-hydroxy-1-(1- hydroxy-2- naphthalenyl)- ethanone  O=C(CO)c1ccc2ccccc2c1O | **No Match (Single paper)** | Nie et al, 2016,18 |
| Oc1c(ccc2ccccc12)C([O-])=O  1-hydroxy-2-naphthoic acid | **Perfect Match**  Throughput: 0.17 | Hennessee et al,46  Leneva et al, 2009, 41  Menn et al, 199373  Baboshin et al, 200540 Balashova et al, 199974  Coppotelli et al, 2010, 75, Doddamani et al, 2000, 76,  Festa et al, 201777  Froehner et al, 200978  Guerin et al, 19983  Noll et al, 199679  Hadibarata et al, 200780  Huang et al, 2016,81  Keum et al, 20067  Kiyohara et al, 197682  Li et al, 201483  Liu et al, 199224  Machate et al, 199784  Mallick et al, 2008,85  Muratova et al, 2014, 49  Ovchinnikova et al, 200986  Pinyakong et al, 2000,63 Prabhu et al, 200371 Prakash and Lal, 2013,72  , Lin et al, 2014,5, Liu et al, 1992,24 Lu et al, 2013,6  Puntus et al, 200887  Rehmann et al, 199630  Sanseverino et al, 1993,88  Samanta et al, 199989  Seo et al, 200714  Seo et al, 200653  Seo et al, 2012,16  Story et al, 200132  Sun et al, 2019,33  Tao et al, 200790  Tittabutr et al, 2011,91 Wang et al, 201692 Wang et al, 200893  Wischmann et al, 199694  Zhang et al, 200595  Zeinali et al, 200837 Zhong et al, 201127  Different protonation:  Keum et al, 200866 |
| 2-hydroxy-1-naphthoic acid  Oc1ccc2ccccc2c1C([O-])=O | **Perfect Match**  Throughput: 0.17 | Balashova et al, 199974  Ghosal et al, 20102  Keum et al, 20067  Puntus et al, 200887  Seo et al, 200714  Seo et al, 200653  Tittabutr et al, 2011,91 |
| 1-hydroxy-2-naphthoic acid methyl ester  COC(=O)c1ccc2ccccc2c1O | **No Match (Single paper)** | Roy et al, 2012,12 |
| 5,8-Dihydroxy-1,4-naphthoquinone  O=C1C=CC(=O)c2c(O)ccc(O)c21 | **No Match (Single paper)** | Wang et al, 201664 |
| Oc1cccc2ccccc12  1-napthol, α-naphthol | **Partial Match**    Throughput: 0.14  Oc1ccc2ccccc2c1O | Lu et al, 2012, 96 Huang et al, 2016,81, Kim et al, 2005,8  Li et al, 201483  Lu et al, 2013,6 Nie et al, 2016,18 Prabhu et al, 200371, Ressler et al, 1999,97  Samanta et al, 199989  Tao et al, 200790  Wang et al, 200893  Zhang et al, 200595  Zhong et al, 201127 |
| Oc1ccc2ccccc2c1  2-naphthol | **Partial Match**    Throughput: 0.14  Oc1ccc2ccccc2c1O | Balashova et al, 199974  Ghosal et al, 20102  Mallick et al, 2007,11 Mallick et al, 2008,85  Nie et al, 2016,18  Puntus et al, 200887 |
| 1-Naphthalenemethanol  1-naphthyl-methylalcohol  OCc1cccc2ccccc12 | **Partial Match**    Throughput: 0.17  Oc1ccc2ccccc2c1C([O-])=O | Nie et al, 2016,18 |
| 2-Naphthalenemethanol  OCc1ccc2ccccc2c1 | **Partial Match**    Throughput: 0.17  Oc1c(ccc2ccccc12)C([O-])=O | Nie et al, 2016,18 |
| [O-]C(=O)c1cccc2ccccc12  Napthoic Acid (1-napthoic acid shown, but the position was not specified by the paper) | **Partial Match**    Throughput: 0.17  Oc1ccc2ccccc2c1C([O-])=O | Dictor et al, 2006, 98 |
| 1,5-dihydroxy-2-naphthoic acid  O=C(O)c1ccc2c(O)cccc2c1O | **Low Throughput Partial Match**  Throughput: 0.002451  Oc1cc(C([O-])=O)c(O)c2ccccc12 | Pinyakong et al, 2000,63 |
| 4-Hydroxycoumarin  O=c1cc(O)c2ccccc2o1 | **No Match (Single paper)** | Wang et al, 201664 |
| O=c1occc2ccccc12  Isocoumarin | **No Match (Single paper)** | Nie et al, 2016,18 |
| Coumarin  O=c1ccc2ccccc2o1 | **No Match (Multiple papers)** | Nie et al, 2016,18  Pinyakong et al, 2000,63  Seo et al, 200714  Seo et al, 2012,16  Wang et al, 200893 |
| O=C1OC(=O)c2ccccc21  Phthalic anhydride | **No Match (Single paper)** | Huang et al, 201681 |
| COc1ccc(C=CC(=O)O)c(O)c1  2-Hydroxy-4-methoxy cinnamate | **No Match (Single paper)** | Hesham et al, 2014, 23 |
| O=C(O)C(=O)/C=C\c1ccc2ccccc2c1O  Cis-4-(1-hydroxynaphth-2-yl)-2-oxobut-3-enoic acid | **No Match (Multiple papers)** | Kim et al, 20058  Lin et al, 2014,5 |
| O=C(O)C(=O)c1ccc2ccccc2c1O  1-hydroxy-α-oxo-2- naphthaleneacetic acid | **No Match (Single paper)** | Nie et al, 2016, 18 |
| Diethyl phthalate  CCOC(=O)C1=C(C=CC=C1)C(=O)OCC | **No Match (Multiple paper)**  However, this is most likely formed from o-phthalate, a matched compound. | Moscoso et al, 2012,99  Masakorala et al, 2013100  Nzila et al, 201870 |
| 2-Acetoxybenzoic acid; aspirin  CC(=O)Oc1ccccc1C(=O)O | **No Match (Multiple paper)** | Lu et al, 2012,96 Lu et al, 2013,6 |
| 2-Carboxy-cinnamic acid  trans-2-carboxycinnamic acid  [O-]C(=O)\C=C\c1ccccc1C([O-])=O | **Low Throughput Match**  Throughput: 5.07e-06  [O-]C(=O)\C=C\c1ccccc1C([O-])=O | Hennessee et al,46  Keum et al, 20067  Seo et al, 200714  Seo et al, 200653  Seo et al, 2012,16 |
| 2-Hydroxybenzalpyruvic acid  O=C(O)C(=O)C=Cc1ccccc1O | **Low Throughput Partial Match**    Throughput: 2.74e-13  Oc1cccc(\C=C\C(=O)C([O-])=O)c1O | Seo et al, 200714 |
| 4-hydroxyphenylacetic acid  O=C(O)Cc1ccc(O)cc1 | **Low Throughput Partial Match**    Throughput: 2.12e-08  OC(=O)Cc1cc(O)ccc1O | Zeinali et al, 200837 |
| o-hydroxyphenylacetic acid  O=C(O)Cc1ccccc1O | **Low Throughput Partial Match**    Throughput: 0.000001  OC(=O)Cc1cccc(O)c1O | Zeinali et al, 200837 |
| phthalic acid,  1,2-benzenedicarboxylic acid  [O-]C(=O)c1ccccc1C([O-])=O | **Perfect Match**  Throughput: 0.10 | Brinda et al, 2013,101  Hennessee et al,46  Hidayat et al, 2018, 47,  Kim et al, 2005, 8  Luan et al, 2006,48  Moody et al, 2001,68,  Rehmann et al, 199630  Samanta et al, 199989  Seo et al, 2012,16  Sun et al, 2019,33  Torres-Farradá et al, 2019102  Umar et al, 2017,36  Umar et al, 201835  Wang et al, 201664  Wang et al, 201692  Wang et al, 200893  Zhang et al, 200595  Zeinali et al, 200837  Zhang et al, 201038  Different Protonation:  Bourguignon et al, 201920  Chebbi et all, 2017103  Guerin et al, 19983  Hadibarata et al, 201060  Keum et al, 20067  Li et al, 201483  Mishra et al, 201950  Seo et al, 200714  Seo et al, 200653  Thomas et al, 2016104 |
| 3,4-Dihydroxyphthalate  Oc1ccc(C([O-])=O)c(C([O-])=O)c1O | **Perfect Match**  Throughput: 0.067 | Seo et al, 200653  Seo et al, 2012,16 |
| 4,5-Dihydroxyphthalate  Oc1cc(C([O-])=O)c(cc1O)C([O-])=O | **Perfect Match**  Throughput: 0.057 | Seo et al, 2012,16  Keum et al, 20067 |
| 2-hydroxy-1-naphthaldehyde  O=Cc1c(O)ccc2ccccc12 | **Partial Match**    Throughput:0.17  Oc1ccc2ccccc2c1C([O-])=O | Keum et al, 20067 |
| Oc1ccccc1C([O-])=O  Salicylic Acid/ 2-hydroxybenzoic acid | **Perfect Match**  Throughput:0.068 | Coppotelli et al, 2010, 75,  Chen et al, 1999105  Ghosal et al, 20102  Guerin et al, 19983  Hadibarata et al, 201161  Hadibarata et al, 200780  Janbandhu et al, 2011106  Leneva et al, 2009, 41  Li et al, 201483  Lin et al, 2014,5  Lu et al, 2013,6  Mallick et al, 2007, 11  Mallick et al, 2008,85  Muratova et al, 2014, 49  Nie et al, 2016,18  Prabhu et al, 2003,71 Prakash and Lal, 2013,72  Samanta et al, 199989  Seo et al, 200714  Seo et al, 200653  Tao et al, 200790  Roy et al, 2012,12 Seo et al, 2012,16 Sun et al, 2019,33 Wang et al, 200893  Different protonation  Shao et al, 2015107 |
| 3-methylsalicylic acid  Cc1cccc(C(=O)O)c1O | **No Match (Single paper)** | Mishra et al, 201950 |
| Oc1ccccc1C=O  Salicylaldehyde | **Partial Match**    Throughput:0.015  Oc1cccc(C=O)c1O | Ghosal et al, 20102  Guerin et al, 19983  Hadibarata et al, 200780  Janbandhu et al, 2011106  Mallick et al, 2007,11 Muratova et al, 2014, 49 Roy et al, 2012,12 |
| Catechol  Oc1ccccc1O | **Perfect Match**  Throughput:0.010 | Ghosal et al, 20102  Guerin et al, 19983  Hadibarata et al, 201161  Hadibarata et al, 200780  Janbandhu et al, 2011106  Leneva et al, 2009,41  Li et al, 201483  Mallick et al, 2007,11 Mallick et al, 2008,85  Nie et al, 2016,18  Roy et al, 2012,12  Seo et al, 2012,16  Wang et al, 200893  Torres-Farradá et al, 2019102  Zhang et al, 200595 |
| 3,4-Dihydroxybenzoate  Protocatechuic acid  Oc1ccc(cc1O)C([O-])=O | **Low Throughput Match**  Throughput: 0.0011 | Guerin et al, 19983  Hadibarata et al, 201060  Keum et al, 20067  Keum et al, 200866  Li et al, 201483  Rehmann et al, 199630  Samanta et al, 199989  Seo et al, 200653  Seo et al, 2012,16  Sun et al, 2019,33  Torres-Farradá et al, 2019102 |
| Gentisate  Oc1ccc(O)c(c1)C([O-])=O | **Perfect Match**  Throughput: 0.011 | Keum et al, 200866, Nzila et al, 201870  Different protonation:  Seo et al, 200714 |
| 4-Hydroxybenzoic acid  O=C(O)c1ccc(O)cc1 | **Partial Match**    Throughput: 0.068  Oc1ccccc1C([O-])=O | Hidayat et al, 2018, 47 |
| 2-(Hydroxymethyl)benzoic acid  O=C(O)c1ccccc1CO | **Partial Match**    Throughput: 0.079  [O-]C(=O)c1ccccc1C=O | Seo et al, 2012,16 |
| 2-carboxybenzaldehyde  2-Formylbenzoic acid  [O-]C(=O)c1ccccc1C=O | **Perfect Match**  Throughput: 0.079 | Keum et al, 20067  Rehmann et al, 199630  Seo et al, 200653  Zhang et al, 201038 |
| 2-methylphenol  Cc1ccccc1O | **Partial Match**    Throughput: 0.015  Oc1cccc(C=O)c1O | Li et al, 201862 |
| Benzoic Acid  O=C(O)c1ccccc1 | **Partial Match**    Throughput: 0.068  Oc1ccccc1C([O-])=O | Hidayat et al, 2018, 47  Li et al, 201483  Zeinali et al, 200837  Different Protonation:  Bourguignon et al, 201920 |
| Phenol  Oc1ccccc1 | **Low Throughput Match**  Throughput: 1.28e-07 | Nie et al, 2016,18 |
| Diosbutyl phthalate  CCCOC(=O)c1ccccc1C(=O)OCCC | No Match  However this study used alkane chains mixed in with the PAH.  These compounds are most likely derived from o-phthalate.  These compounds are not counted in the accuracy scoring. | Wang et al, 2007108 |
| Dibutyl phthalate  CCCCOC(=O)c1ccccc1C(=O)OCCCC |  | Wang et al, 2007108 |
| mono(2-ethylhexyl) phthalate  CCCCC(CC)COC(=O)C1=C(C=CC=C1)C(O)=O |  | Wang et al, 2007108 |
| 1,2-benzenedicarboxylic acid, diisooctyl ester  CC(C)CCCCCOC(=O)C1=CC=CC=C1C(=O)OCCCCCC(C)C |  | Mishra et al, 201950 |

**Table S4.3:** PhenanthreneEmpirical Review Summary

|  |  |  |
| --- | --- | --- |
|  | Perfect Match | Partial Match |
| High Throughput | 14 | 25 |
| Low Throughput | 3 | 4 |

No Match (Single) 25

No Match (Multiple) 13

**No Match Explanations:**

Several of the No-Match Compounds are the result of methylation or other straight chain addition after the cleavage of another bound. EAWAG does not predict this type of transition, so these compounds would not be found. Certain anhydride, oxide, or coumarin structures are also not predicted, so these compounds are either classified as no Match or Partial Match if their immediate parent and daughter compounds are predicted.

**References**

(1) Gao, S., et al., Multiple Degradation Pathways of Phenanthrene by Stenotrophomonas Maltophilia C6*.* *Int Biodeterior Biodegradation*, 2013. **79**: p. 98-104.

(2) Ghosal, D., et al., Degradation of Phenanthrene Via Meta-Cleavage of 2-Hydroxy-1-Naphthoic Acid by Ochrobactrum Sp. Strain Pwtjd*.* *FEMS Microbiol Lett*, 2010. **313**(2): p. 103-10.

(3) Guerin, W.F. and G.E. Jones, Mineralization of Phenanthrene by a Mycobacterium Sp*.* *Applied and Environmental Microbiology*, 1988. **54**(4): p. 937-944.

(4) Lee, S., H. Ryu, and K. Nam, Phenanthrene Metabolites Bound to Soil Organic Matter by Birnessite Following Partial Biodegradation*.* *Environmental Toxicology and Chemistry*, 2009. **28**(5): p. 946-952.

(5) Lin, M., et al., Biodegradation of Phenanthrene by Pseudomonas Sp. Bz-3, Isolated from Crude Oil Contaminated Soil*.* *International Biodeterioration & Biodegradation*, 2014. **94**: p. 176-181.

(6) Lu, J., et al., A Fusant of Sphingomonas Sp. Gy2b and Pseudomonas Sp. Gp3a with High Capacity of Degrading Phenanthrene*.* *World J Microbiol Biotechnol*, 2013. **29**(9): p. 1685-94.

(7) Keum, Y.S., et al., Degradation Pathways of Phenanthrene by Sinorhizobium Sp. C4*.* *Appl Microbiol Biotechnol*, 2006. **71**(6): p. 935-41.

(8) Kim, Y.H., et al., Effects of Ph on the Degradation of Phenanthrene and Pyrene by Mycobacterium Vanbaalenii Pyr-1*.* *Appl Microbiol Biotechnol*, 2005. **67**(2): p. 275-85.

(9) Kweon, O., et al., Substrate Specificity and Structural Characteristics of the Novel Rieske Nonheme Iron Aromatic Ring-Hydroxylating Oxygenases Nidab and Nida3b3 from Mycobacterium Vanbaalenii Pyr-1*.* *MBio*, 2010. **1**(2).

(10) Lisowska, K., et al., The Effect of the Corticosteroid Hormone Cortexolone on the Metabolites Produced During Phenanthrene Biotransformation in Cunninghamella Elegans*.* *Chemosphere*, 2006. **64**(9): p. 1499-1506.

(11) Mallick, S., S. Chatterjee, and T.K. Dutta, A Novel Degradation Pathway in the Assimilation of Phenanthrene by Staphylococcus Sp. Strain Pn/Y Via Meta-Cleavage of 2-Hydroxy-1-Naphthoic Acid: Formation of Trans-2,3-Dioxo-5-(2'-Hydroxyphenyl)-Pent-4-Enoic Acid*.* *Microbiology*, 2007. **153**(Pt 7): p. 2104-15.

(12) Roy, M., P. Khara, and T.K. Dutta, Meta-Cleavage of Hydroxynaphthoic Acids in the Degradation of Phenanthrene by Sphingobium Sp. Strain Pnb*.* *Microbiology*, 2012. **158**(Pt 3): p. 685-95.

(13) Sack, U., et al., Comparison of Phenanthrene and Pyrene Degradation by Different Wood-Decaying Fungi*.* *Applied and Environmental Microbiology*, 1997. **63**(10): p. 3919-3925.

(14) Seo, J.S., et al., Degradation of Phenanthrene by Burkholderia Sp C3: Initial 1,2- and 3,4-Dioxygenation and Meta- and Ortho-Cleavage of Naphthalene-1,2-Diol*.* *Biodegradation*, 2007. **18**(1): p. 123-131.

(15) Seo, J.S., Y.S. Keum, and Q.X. Li, Bacterial Degradation of Aromatic Compounds*.* *Int J Environ Res Public Health*, 2009. **6**(1): p. 278-309.

(16) Seo, J.-S., Y.-S. Keum, and Q.X. Lic, Mycobacterium Aromativorans Js19b1t Degrades Phenanthrene through C-1,2, C-3,4 and C-9,10 Dioxygenation Pathways*.* *70*, 2012. **Int Biodeterior Biodegradation**: p. 96–103.

(17) van Herwijnen, R., et al., Elucidation of the Metabolic Pathway of Fluorene and Cometabolic Pathways of Phenanthrene, Fluoranthene, Anthracene and Dibenzothiophene by Sphingomonas Sp Lb126*.* *Research in Microbiology*, 2003. **154**(3): p. 199-206.

(18) Nie, M.Q., et al., Phenanthrene Metabolites from a New Polycyclic Aromatic Hydrocarbon-Degrading Bacterium Aeromonas Salmonicida Subsp. Achromogenes Strain Ny4*.* *Polycyclic Aromatic Compounds*, 2016. **36**(2): p. 132-151.

(19) Aranda, E., R. Ullrich, and M. Hofrichter, Conversion of Polycyclic Aromatic Hydrocarbons, Methyl Naphthalenes and Dibenzofuran by Two Fungal Peroxygenases*.* *Biodegradation*, 2010. **21**(2): p. 267-281.

(20) Bourguignon, N., et al., Identification of Proteins Induced by Polycyclic Aromatic Hydrocarbon and Proposal of the Phenanthrene Catabolic Pathway in Amycolatopsis Tucumanensis Dsm 45259*.* *Ecotoxicology and Environmental Safety*, 2019. **175**: p. 19-28.

(21) Casillas, R.P., et al., Initial Oxidative and Subsequent Conjugative Metabolites Produced During the Metabolism of Phenanthrene by Fungi*.* *Journal of Industrial Microbiology*, 1996. **16**(4): p. 205-215.

(22) da Silva, M., et al., Metabolism of Aromatic Hydrocarbons by the Filamentous Fungus Cyclothyrium Sp*.* *Chemosphere*, 2004. **57**(8): p. 943-952.

(23) Hesham Ael, L., et al., Biodegradation Ability and Catabolic Genes of Petroleum-Degrading Sphingomonas Koreensis Strain Asu-06 Isolated from Egyptian Oily Soil*.* *Biomed Res Int*, 2014. **2014**: p. 127674.

(24) Liu, D., et al., Microbial Degradation of Polycyclic Aromatic Hydrocarbons and Polycyclic Aromatic Nitrogen Heterocyclics*.* *Environmental Toxicology and Water Quality*, 1992. **7**(4): p. 355-372.

(25) Sack, U., et al., Novel Metabolites in Phenanthrene and Pyrene Transformation by Aspergillus Niger*.* *Applied and Environmental Microbiology*, 1997. **63**(7): p. 2906-2909.

(26) Schrlau, J.E., et al., Formation of Developmentally Toxic Phenanthrene Metabolite Mixtures by Mycobacterium Sp. Elw1*.* *Environ Sci Technol*, 2017. **51**(15): p. 8569-8578.

(27) Zhong, Y., et al., Production of Metabolites in the Biodegradation of Phenanthrene, Fluoranthene and Pyrene by the Mixed Culture of Mycobacterium Sp. And Sphingomonas Sp*.* *Bioresour Technol*, 2011. **102**(3): p. 2965-72.

(28) Allen, C.C.R., et al., Contrasting Effects of a Nonionic Surfactant on the Biotransformation of Polycyclic Aromatic Hydrocarbons to Cis-Dihydrodiols by Soil Bacteria*.* *Applied and Environmental Microbiology*, 1999. **65**(3): p. 1335-1339.

(29) Pathak, H., et al., Technical Note: Degradation of Phenanthrene and Anthracene Bypseudomonasstrain, Isolated from Coastal Area*.* *Bioremediation Journal*, 2008. **12**(2): p. 111-116.

(30) Rehmann, K., C. Steinberg, and A. Kettrup, Branched Metabolic Pathway for Phenanthrene Degradation in a Pyrene-Degrading Bacterium*.* *Polycyclic Aromatic Compounds*, 1996. **11**(1-4): p. 125-130.

(31) Stingley, R.L., A.A. Khan, and C.E. Cerniglia, Molecular Characterization of a Phenanthrene Degradation Pathway in Mycobacterium Vanbaalenii Pyr-1*.* *Biochem Biophys Res Commun*, 2004. **322**(1): p. 133-46.

(32) Story, S.P., et al., Convergent and Divergent Points in Catabolic Pathways Involved in Utilization of Fluoranthene, Naphthalene, Anthracene, and Phenanthrene by Sphingomonas Paucimobilis Var. Epa505*.* *Journal of Industrial Microbiology & Biotechnology*, 2001. **26**(6): p. 369-382.

(33) Sun, S.S., et al., Salicylate and Phthalate Pathways Contributed Differently on Phenanthrene and Pyrene Degradations in Mycobacterium Sp. Wy10*.* *Journal of Hazardous Materials*, 2019. **364**: p. 509-518.

(34) Sutherland, J.B., et al., Metabolism of Phenanthrene by Phanerochaete-Chrysosporium*.* *Applied and Environmental Microbiology*, 1991. **57**(11): p. 3310-3316.

(35) Umar, Z.D., et al., Effective Phenanthrene and Pyrene Biodegradation Using Enterobacter Sp. Mm087 (Kt933254) Isolated from Used Engine Oil Contaminated Soil*.* *Egyptian Journal of Petroleum*, 2018. **27**(3): p. 349-359.

(36) Umar, Z.D., et al., Rapid Biodegradation of Polycyclic Aromatic Hydrocarbons (Pahs) Using Effective Cronobacter Sakazakii Mm045 (Kt933253)*.* *MethodsX*, 2017. **4**: p. 104-117.

(37) Zeinali, M., M. Vossoughi, and S.K. Ardestani, Degradation of Phenanthrene and Anthracene by Nocardia Otitidiscaviarum Strain Tsh1, a Moderately Thermophilic Bacterium*.* *J Appl Microbiol*, 2008. **105**(2): p. 398-406.

(38) Zhang, Z., C. Inoue, and G. Li, Coordination in Phenanthrene Biodegradation: Pyruvate as Microbial Demarcation*.* *Bull Environ Contam Toxicol*, 2010. **85**(6): p. 581-4.

(39) Safonova, E., et al., Biodegradation of Phenanthrene by the Green Alga Scenedesmus Obliquus Es-55*.* *Engineering in Life Sciences*, 2005. **5**(3): p. 234-239.

(40) Baboshin, M.A., et al., The Microbial Transformation of Phenanthrene and Anthracene*.* *Microbiology*, 2005. **74**(3): p. 303-309.

(41) Leneva, N.A., et al., Phenanthrene and Anthracene Degradation by Microorganisms of the Genus Rhodococcus*.* *Applied Biochemistry and Microbiology*, 2009. **45**(2): p. 169-175.

(42) Bezalel, L., Y. Hadar, and C.E. Cerniglia, Enzymatic Mechanisms Involved in Phenanthrene Degradation by the White Rot Fungus Pleurotus Ostreatus*.* *Applied and Environmental Microbiology*, 1997. **63**(7): p. 2495-2501.

(43) Bezalel, L., et al., Metabolism of Phenanthrene by the White Rot Fungus Pleurotus Ostreatus*.* *Applied and Environmental Microbiology*, 1996. **62**(7): p. 2547-2553.

(44) Cajthaml T, et al., Study of Fungal Degradation Products of Polycyclic Aromatic Hydrocarbons Using Gas Chromatography with Ion Trap Mass Spectrometry Detection*.* *Journal of Chromatography*, 2002. **974**: p. 213–222.

(45) Fu, W.Q., et al., Biodegradation of Phenanthrene by Endophytic Fungus Phomopsis Liquidambari in Vitro and in Vivo*.* *Chemosphere*, 2018. **203**: p. 160-169.

(46) Hennessee, C.T. and Q.X. Li, Effects of Polycyclic Aromatic Hydrocarbon Mixtures on Degradation, Gene Expression, and Metabolite Production in Four Mycobacterium Species*.* *Appl Environ Microbiol*, 2016. **82**(11): p. 3357-3369.

(47) Hidayat, A. and D.H.Y. Yanto, Biodegradation and Metabolic Pathway of Phenanthrene by a New Tropical Fungus, Trametes Hirsuta D7*.* *Journal of Environmental Chemical Engineering*, 2018. **6**(2): p. 2454-2460.

(48) Luan, T.G., et al., Study of Metabolites from the Degradation of Polycyclic Aromatic Hydrocarbons (Pahs) by Bacterial Consortium Enriched from Mangrove Sediments*.* *Chemosphere*, 2006. **65**(11): p. 2289-96.

(49) Muratova, A., et al., Degradation of Phenanthrene by the Rhizobacterium Ensifer Meliloti*.* *Biodegradation*, 2014. **25**(6): p. 787-95.

(50) Mishra, A., et al., Degradation and Detoxification of Phenanthrene by Actinobacterium Zhihengliuella Sp. Istpl4*.* *Environmental Science and Pollution Research*, 2019.

(51) Ning, D., et al., Novel Evidence of Cytochrome P450-Catalyzed Oxidation of Phenanthrene in Phanerochaete Chrysosporium under Ligninolytic Conditions*.* *Biodegradation*, 2010. **21**(6): p. 889-901.

(52) Sack, U., M. Hofrichter, and W. Fritsche, Degradation of Phenanthrene and Pyrene by Nematoloma Frowardii*.* *Journal of Basic Microbiology*, 1997. **37**(4): p. 287-293.

(53) Seo, J.S., et al., Phenanthrene Degradation in Arthrobacter Sp. P1-1: Initial 1,2-, 3,4- and 9,10-Dioxygenation, and Meta- and Ortho-Cleavages of Naphthalene-1,2-Diol after Its Formation from Naphthalene-1,2-Dicarboxylic Acid and Hydroxyl Naphthoic Acids*.* *Chemosphere*, 2006. **65**(11): p. 2388-2394.

(54) Sutherland, J.B., et al., Stereoselective Formation of a K-Region Dihydrodiol from Phenanthrene by Streptomyces Flavovirens*.* *Archives of Microbiology*, 1990. **154**(3): p. 260-266.

(55) Tatarko, M. and J.A. Bumpus, Biodegradation of Phenanthrene by Phanerochaete-Chrysosporium - on the Role of Lignin Peroxidase*.* *Letters in Applied Microbiology*, 1993. **17**(1): p. 20-24.

(56) Yang, H.Y., et al., Degradation of Recalcitrant Aliphatic and Aromatic Hydrocarbons by a Dioxin-Degrader Rhodococcus Sp. Strain P52*.* *Environ Sci Pollut Res Int*, 2014. **21**(18): p. 11086-93.

(57) Capotorti, G., et al., Formation of Sulfate Conjugates Metabolites in the Degradation of Phenanthrene, Anthracene, Pyrene and Benzo a Pyrene by the Ascomycete Aspergillus Terreus*.* *Polycyclic Aromatic Compounds*, 2005. **25**(3): p. 197-213.

(58) Jin, J., J. Yao, and Q. Zhang, Biodegradation of Phenanthrene by Pseudomonas Sp. Jpn2 and Structure-Based Degrading Mechanism Study*.* *Bull Environ Contam Toxicol*, 2016. **97**(5): p. 689-694.

(59) Szewczyk, R., et al., Application of Microscopic Fungi Isolated from Polluted Industrial Areas for Polycyclic Aromatic Hydrocarbons and Pentachlorophenol Reduction*.* *Biodegradation*, 2003. **14**(1): p. 1-8.

(60) Hadibarata, T. and S. Tachibana, Characterization of Phenanthrene Degradation by Strain Polyporus Sp. S133*.* *Journal of Environmental Sciences*, 2010. **22**(1): p. 142-149.

(61) Hadibarata, T., S. Tachibana, and M. Askari, Identification of Metabolites from Phenanthrene Oxidation by Phenoloxidases and Dioxygenases of Polyporus Sp S133*.* *Journal of Microbiology and Biotechnology*, 2011. **21**(3): p. 299-304.

(62) Li, X., et al., Diversity of Phenanthrene and Benz[a]Anthracene Metabolic Pathways in White Rot Fungus Pycnoporus Sanguineus 14*.* *International Biodeterioration & Biodegradation*, 2018. **134**: p. 25-30.

(63) Pinyakong, O., et al., Identification of Novel Metabolites in the Degradation of Phenanthrene by Sphingomonas Sp Strain P2*.* *Fems Microbiology Letters*, 2000. **191**(1): p. 115-121.

(64) Wang, F.K., et al., Characterization of a Phenanthrene-Degrading Microbial Consortium Enriched from Petrochemical Contaminated Environment*.* *International Biodeterioration & Biodegradation*, 2016. **115**: p. 286-292.

(65) Adebusuyi, A.A., et al., The Emhabc Efflux Pump Decreases the Efficiency of Phenanthrene Biodegradation by Pseudomonas Fluorescens Strain Lp6a*.* *Applied Microbiology and Biotechnology*, 2012. **95**(3): p. 757-766.

(66) Keum, Y.S., et al., Comparative Metabolomic Analysis of Sinorhizobium Sp. C4 During the Degradation of Phenanthrene*.* *Appl Microbiol Biotechnol*, 2008. **80**(5): p. 863-72.

(67) Arias, L., et al., A Microcosm System and an Analytical Protocol to Assess Pah Degradation and Metabolite Formation in Soils*.* *Biodegradation*, 2008. **19**(3): p. 425-34.

(68) Moody, J.D., et al., Degradation of Phenanthrene and Anthracene by Cell Suspensions of Mycobacterium Sp. Strain Pyr-1*.* *Appl Environ Microbiol*, 2001. **67**(4): p. 1476-83.

(69) Hammel, K.E., et al., Oxidative-Degradation O Phenanthrene by the Ligninolytic Fungus Phanerochaete-Chrysosporium*.* *Applied and Environmental Microbiology*, 1992. **58**(6): p. 1832-1838.

(70) Nzila, A., et al., Isolation and Characterisation of Bacteria Degrading Polycyclic Aromatic Hydrocarbons: Phenanthrene and Anthracene*.* *Archives of Environmental Protection*, 2018. **44**(2): p. 43-54.

(71) Prabhu, Y. and P.S. Phale, Biodegradation of Phenanthrene by Pseudomonas Sp. Strain Pp2: Novel Metabolic Pathway, Role of Biosurfactant and Cell Surface Hydrophobicity in Hydrocarbon Assimilation*.* *Appl Microbiol Biotechnol*, 2003. **61**(4): p. 342-51.

(72) Prakash, O., Role of Unstable Phenanthrene-Degrading Pseudomonas Species in Natural Attenuation of Phenanthrene-Contaminated Site*.* *Korean Journal of Microbiology and Biotechnology*, 2013. **41**(1): p. 79-87.

(73) Menn, F.-M., B.M. Applegate, and G.S. Sayler, Nah Plasmid-Mediated Catabolism of Anthracene and Phenanthrene to Naphthoic Acids*.* *Applied and Environmental Microbiology*, 1993. **59**(6): p. 1938-1942.

(74) Balashova, N.V., et al., Phenanthrene Metabolism by Pseudomonas and Burkholderia Strains*.* *Process Biochemistry*, 1999. **35**(3-4): p. 291-296.

(75) Coppotelli, B.M., et al., Study of the Degradation Activity and the Strategies to Promote the Bioavailability of Phenanthrene by Sphingomonas Paucimobilis Strain 20006fa*.* *Microb Ecol*, 2010. **59**(2): p. 266-76.

(76) Doddamani, H.P. and H.Z. Ninnekar, Biodegradation of Phenanthrene by a Bacillus Species*.* *Current Microbiology*, 2000. **41**(1): p. 11-14.

(77) Festa, S., et al., Assigning Ecological Roles to the Populations Belonging to a Phenanthrene-Degrading Bacterial Consortium Using Omic Approaches*.* *Plos One*, 2017. **12**(9): p. 21.

(78) Froehner, S., E.C. da Luz, and M. Maceno, Enhanced Biodegradation of Naphthalene and Anthracene by Modified Vermiculite Mixed with Soil*.* *Water Air and Soil Pollution*, 2009. **202**(1-4): p. 169-177.

(79) Noll, H., T. Machate, and A. Kettrup, Microbial Degradation of Phenanthrene in a Constructed Wetland*.* *Polycyclic Aromatic Compounds*, 1996. **11**(1-4): p. 137-144.

(80) Hadibarata, T., S. Tachibana, and K. Itoh, Biodegradation of Phenanthrene by Fungi Screened from Nature*.* *Pakistan Journal of Biological Sciences*, 2007. **10**(15): p. 2535-2543.

(81) Huang, X., et al., Biodegradation of Phenanthrene by Rhizobium Petrolearium Sl-1*.* *J Appl Microbiol*, 2016. **121**(6): p. 1616-1626.

(82) Kiyohara, H., K. Nagao, and R. Nomi, Degradation of Phenanthrene through O-Phthalate by an Aeromonas Sp*.* *Agricultural and Biological Chemistry*, 1976. **40**(6): p. 1075-1082.

(83) Li, F., L. Zhu, and D. Zhang, Effect of Surfactant on Phenanthrene Metabolic Kinetics by Citrobacter Sp. Sa01*.* *Journal of Environmental Sciences (China)*, 2014. **26**(11): p. 2298-2306.

(84) Machate, T., et al., Degradation of Phenanthrene and Hydraulic Characteristics in a Constructed Wetland*.* *Water Research*, 1997. **31**(3): p. 554-560.

(85) Mallick, S. and T.K. Dutta, Kinetics of Phenanthrene Degradation by Staphylococcus Sp. Strain Pn/Y Involving 2-Hydroxy-1-Naphthoic Acid in a Novel Metabolic Pathway*.* *Process Biochemistry*, 2008. **43**(9): p. 1004-1008.

(86) Ovchinnikova, A.A., et al., Phenanthrene Biodegradation and the Interaction of Pseudomonas Putida Bs3701 and Burkholderia Sp Bs3702 in Plant Rhizosphere*.* *Microbiology*, 2009. **78**(4): p. 433-439.

(87) Puntus, I.F., et al., Phenanthrene Degradation by Bacteria of the Genera Pseudomonas and Burkholderia in Model Soil Systems*.* *Microbiology*, 2008. **77**(1): p. 7-15.

(88) Sanseverino, J., et al., Plasmid-Mediated Mineralization of Naphthalene, Phenanthrene, and Anthracene*.* *Applied and Environmental Microbiology*, 1993. **59**(6): p. 1931-1937.

(89) Samanta, S.K., A.K. Chakraborti, and R.K. Jain, Degradation of Phenanthrene by Different Bacteria: Evidence for Novel Transformation Sequences Involving the Formation of 1-Naphthol*.* *Applied Microbiology and Biotechnology*, 1999. **53**(1): p. 98-107.

(90) Tao, X.Q., et al., A Phenanthrene-Degrading Strain Sphingomonas Sp Gy2b Isolated from Contaminated Soils*.* *Process Biochemistry*, 2007. **42**(3): p. 401-408.

(91) Tittabutr, P., I.K. Cho, and Q.X. Li, Phn and Nag-Like Dioxygenases Metabolize Polycyclic Aromatic Hydrocarbons in Burkholderia Sp. C3*.* *Biodegradation*, 2011. **22**(6): p. 1119-33.

(92) Wang, H., et al., Efficient Biodegradation of Phenanthrene by a Novel Strain Massilia Sp. Wf1 Isolated from a Pah-Contaminated Soil*.* *Environ Sci Pollut Res Int*, 2016. **23**(13): p. 13378-88.

(93) Wang, J., et al., Kinetics and Characteristics of Phenanthrene Degradation by a Microbial Consortium*.* *Petroleum Science*, 2008. **5**(1): p. 73-78.

(94) Wischmann, H., et al., Degradation of Selected Pahs in Soil/Compost and Identification of Intermediates*.* *International Journal of Environmental Analytical Chemistry*, 1996. **64**(4): p. 247-255.

(95) Zhang, L., et al., Phenanthrene-Degrading Pathway of Agrobacterium Sp Phx1*.* *Science in China Series D-Earth Sciences*, 2005. **48**: p. 269-275.

(96) Lu, J., et al., Biodegradation Kinetics of Phenanthrene by a Fusant Strain*.* *Curr Microbiol*, 2012. **65**(3): p. 225-30.

(97) Ressler, B.P., H. Kneifel, and J. Winter, Bioavailability of Polycyclic Aromatic Hydrocarbons and Formation of Humic Acid-Like Residues During Bacterial Pah Degradation*.* *Applied Microbiology and Biotechnology*, 1999. **53**(1): p. 85-91.

(98) Dictor, M.C., et al., Influence of Ageing of Polluted Soils on Bioavailability of Phenanthrene*.* *Oil & Gas Science and Technology*, 2006. **58**(4): p. 481-488.

(99) Moscoso, F., et al., Technoeconomic Assessment of Phenanthrene Degradation by Pseudomonas Stutzeri Cect 930 in a Batch Bioreactor*.* *Bioresour Technol*, 2012. **104**: p. 81-9.

(100) Masakorala, K., et al., Isolation and Characterization of a Novel Phenanthrene (Phe) Degrading Strain Psuedomonas Sp. Ustb-Ru from Petroleum Contaminated Soil*.* *Journal of Hazardous Materials*, 2013. **263**: p. 493-500.

(101) Brinda, L.M., K. Muthukumar, and M. Velan, Optimization of Minimal Salt Medium for Efficient Phenanthrene Biodegradation Bymycoplanasp. Mvmb2 Isolated from Petroleum Contaminated Soil Using Factorial Design Experiments*.* *CLEAN - Soil, Air, Water*, 2013. **41**(1): p. 51-59.

(102) Torres-Farradá, G., et al., Biodegradation of Polycyclic Aromatic Hydrocarbons by Native Ganoderma Sp. Strains: Identification of Metabolites and Proposed Degradation Pathways*.* *Applied Microbiology and Biotechnology*, 2019.

(103) Chebbi, A., et al., Polycyclic Aromatic Hydrocarbon Degradation and Biosurfactant Production by a Newly Isolated Pseudomonas Sp. Strain from Used Motor Oil-Contaminated Soil*.* *International Biodeterioration & Biodegradation*, 2017. **122**: p. 128-140.

(104) Thomas, F., et al., Isolation and Substrate Screening of Polycyclic Aromatic Hydrocarbon Degrading Bacteria from Soil with Long History of Contamination*.* *International Biodeterioration & Biodegradation*, 2016. **107**: p. 1-9.

(105) Chen, S.H. and M.D. Aitken, Salicylate Stimulates the Degradation of High Molecular Weight Polycyclic Aromatic Hydrocarbons by Pseudomonas Saccharophila P15*.* *Environmental Science & Technology*, 1999. **33**(3): p. 435-439.

(106) Janbandhu, A. and M.H. Fulekar, Biodegradation of Phenanthrene Using Adapted Microbial Consortium Isolated from Petrochemical Contaminated Environment*.* *Journal of Hazardous Materials*, 2011. **187**(1-3): p. 333-340.

(107) Shao, Y.X., et al., Biodegradation of Pahs by Acinetobacter Isolated from Karst Groundwater in a Coal-Mining Area*.* *Environmental Earth Sciences*, 2015. **73**(11): p. 7479-7488.

(108) Wang, J., H.K. Xu, and S.H. Guo, Isolation and Characteristics of a Microbial Consortium for Effectively Degrading Phenanthrene*.* *Petroleum Science*, 2007. **4**(3): p. 68-75.

# Section 5: Analysis and Plotting Code

A full analysis of the PAH degradation network will generate the following files. Examples are provided at <https://github.com/ngLabGroup/networksPlotting>

Raw Data Files:

RawExampleData.xlsx – Edgelist obtained from EAWAG or similar pathway predictor

BATCH.OUT files from EPI Suite™. These contain the Log Kow (LogP) value, there may be several depending on the size of the network.

RuleData.xlsx – Contains the aerobic weightings as defined by EAWAG. This is an input for the nodeThroughput.py script

Generated Data Files:

ExampleEpiLogKow.xlsx – Log Kow Estimation from EpiSuite. Similar files could be generated if parameters were desired than cannot be computed by RDKit. Other files can merge them in based on the SMILES codes.

ExamplePositions.xlsx – Optional file for improving waterfall plots. Allows nodes to be manually moved for viewing

ExampleQFNodes.xlsx – Example of the output from the nodeThroughput.py

LiteratureMatches\_Highthroughput.xlsx – Reference file for compounds that match literature and that pass whatever threshold is selected as high throughput. Can be tabbed out to serve multiple compounds. The high throughput compounds can be easily copied from the node throughput file. It is usually desirable to be able to reference these compounds easily, which is why they are stored in their own excel file rather than being pulled from the node throughput file

ExamplePathWeight.xlsx The Edgeless after Relevant Pathway filtering has been applied

Supporting Code Scripts:

- CustomFunctions\_PAHteam.py contains several functions that are needed for the other scripts

Analysis Code Scripts:

Run the scripts in the order that they appear on the page

- addEpiSuiteLogKow.py This script is used to package the outputs from the *Kowwin* module of the EpiSuite BATCH.OUT files into an excel file

- nodeThroughput.py This script applies the nodeThroughput algorithm and write out the ExampleNT\_BTW\_Nodes.xlsx file

- relevantPathwayFiltering.py This script allies the Relevant Pathway filter and writes out the ExamplePathWeight.xlsx

- network\_Molecular\_correlations.py This script computes the correlations between network data and chemistry data

- substructureFinder.py This script is used to locating Partial Matches. This script works best if it is manipulated deliberately.

Plotting Code Scripts:

- waterfallPlot\_LitMatchs.py – This script generates Figure 5 in the main test as well as the plots in the SI for each PAH.

- pathwayMatchLegend.py – This script generates the percentage bar in Figure 5. These two plots can be combined with an image program.

- lit\_comparisonBarPlots.py – This script generates Figure 6 in the main text

- betweennessMCS\_plotter.py – This script will not work with the small example data set as the MCS code requires many more compounds, but is provided as reference.

Example Data Sets:

We cannot provide the full data sets, however small partial datasets are provided to demonstrate the workings of the code.

# Section 6. Sensitivity Analysis

**Weighting Sensitivity Analysis**

The compounds in the High Node Throughput category can be classified based on how they connect to each other, especially by how they connect to the empirical literature matches (Perfect Matches). This is one of the major strengths of applying a networks approach to this problem. As the figures of the individual network plots in the earlier SI sections illustrate, High Node Throughput Compounds can be classified as either “Perfect Match”, where compounds match an empirical literature value exactly, “Pathway Match” where they are on a degradation pathway to or from a “Perfect Match” compound, or as a “Network Outlier” where the node does not match an empirical literature study exactly and there is no direct pathway among other High Node Throughput compounds to or from a literature match.

EAWAG-PPS categorizes the biodegradation reactions (network edges) by aerobic likelihood. In this paper, the categories of “Neutral”, “Likely”, and “Very Likely” were considered. The categories of “Unlikely” and “Very Unlikely” were assumed to be negligible. In order to determine the proper numerical weighting (Edge Aerobic Likelihood Score) to assign to the different aerobic likelihood categories, a sensitivity analysis was conducted. Figure S6.1 shows the results of this analysis for the High Node Throughput compounds, (nodes with a greater than 0.01 Node Throughput value).

The ideal Edge Aerobic Likelihood Score would assign scores (weights) to the biodegradation edges such that after applying the Node Throughput algorithm, as many of the Perfect Match literature metabolites as possible fall into the “High Node Throughput” category after the Node Throughput algorithm was applied. In order to determine the ideal weighting scheme an initial scoring of 1,2,3 (spread of 1) was assigned to the edge categories of “Neutral”, “Likely”, and “Very Likely”. The spread was then incrementally increased by integer values; a spread of 2 would be a scheme of 1,3,5, and a spread of 3 would be a scheme of 1,4,7, etc, on up to a spread of 99. The number of predicted metabolites above the High Node Threshold of 0.01 was then considered to determine the best Edge Aerobic Likelihood scoring scheme. The 0.01 threshold as the definition for High Node Throughput was estimated initially as a good threshold for selecting the most relevant metabolites and was re-verified below.

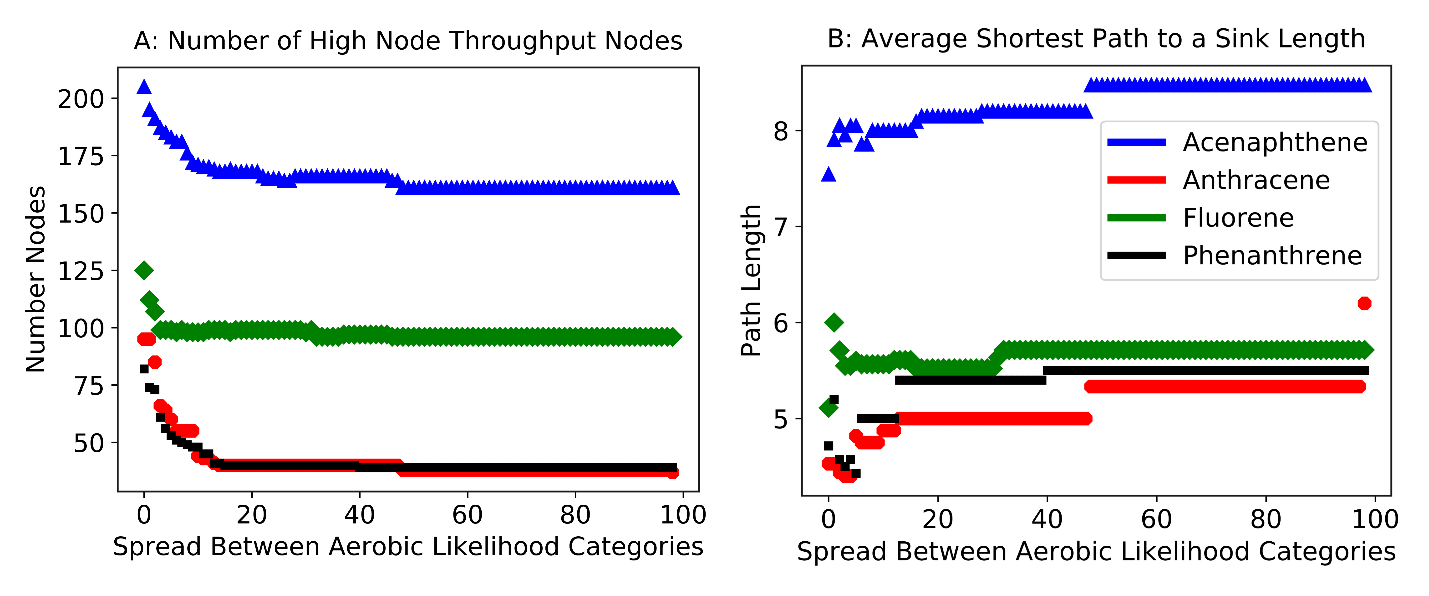
As the Edge Aerobic Likelihood Scores are increased, the “Neutral” category is help constant at 1 and the weightings of the “Likely” and “Very Likely” categories are increased. The Aerobic Likelihood Scores are part of the Edge Throughput Value Equation (1) in the main text.

(1)

and each node’s “Throughput” value is assigned by:

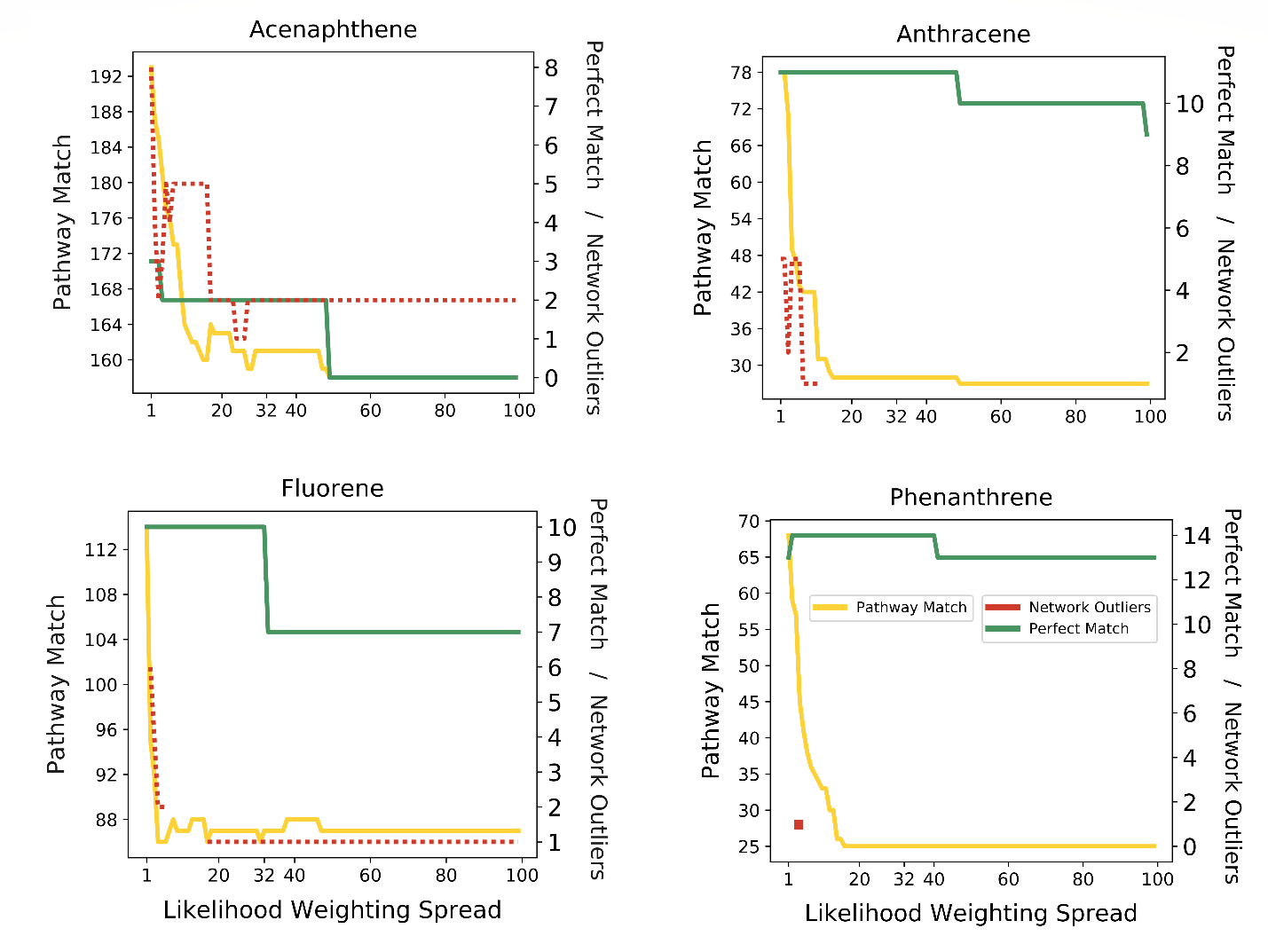
(2)

The greater the spread of the Aerobic Likelihood categories, the greater the final Edge Throughput Value assigned to the “Likely” and “Very Likely” edges will be, and the lower the final Edge Throughput Value assigned to the “Neutral” edges will be. A greater spread between the categories will result in assigning more Edge Throughput to the “Likely” and “Very Likely” edges and less to the “Neutral” Edges.



**Figure S6.1:** Network Metrics at Different Aerobic Likelihood Spreads

The increase in spread will result in an overall lower number of nodes because branches that are started by a “Neutral” edge will have lower Node Throughput assigned to them and their daughter products consequently will fall below the cutoff for High Throughput earlier along the degradation pathway. Additionally, the average shortest path length to a sink compound increases as more weight is assigned to the “Likely” and “Very Likely” categories, as the degradation chains will go farther before the final sink has less than a 0.01 Node Throughput value. These two networks measurements are summarized for each PAH in Figure S6.2.

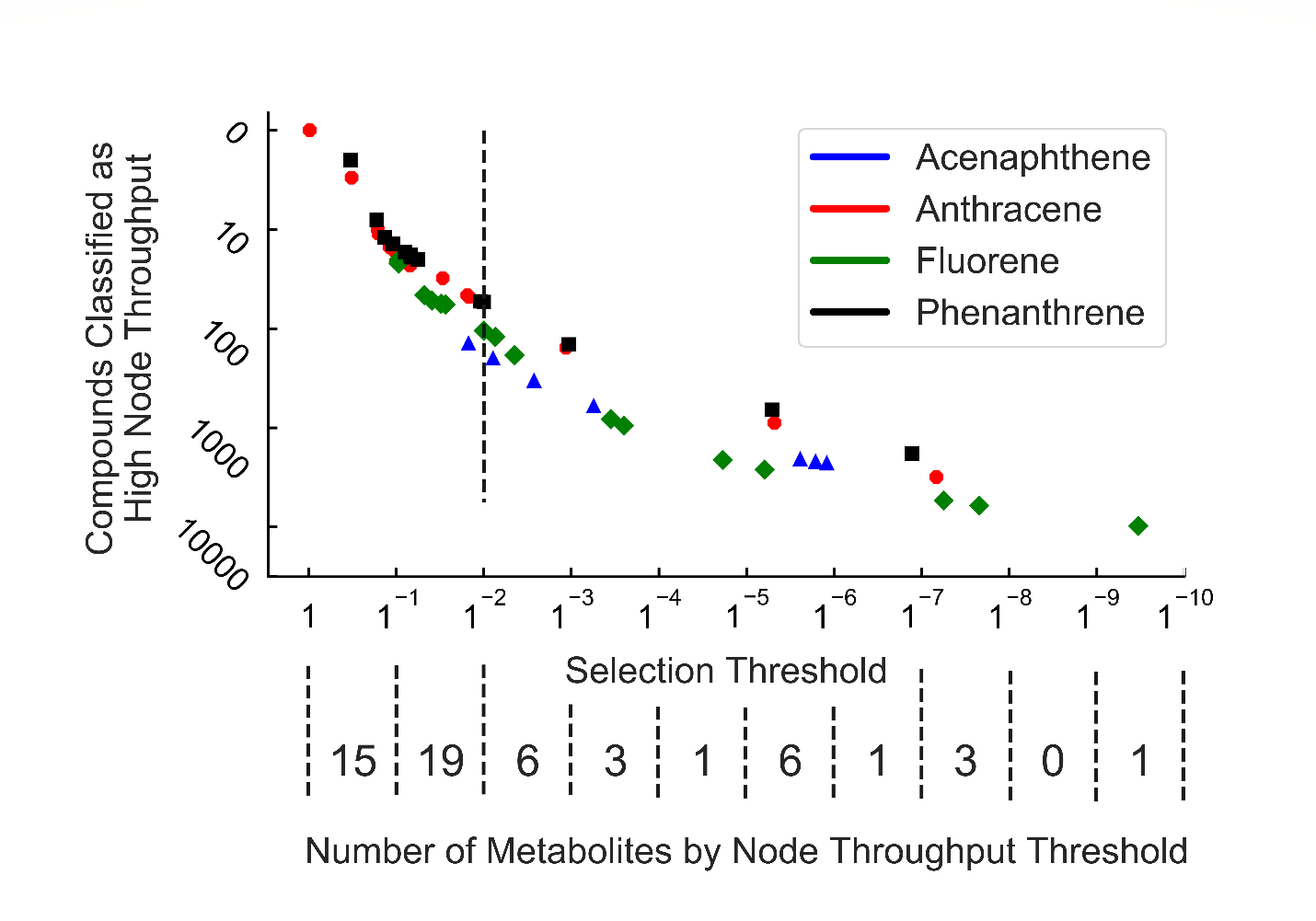
****

**Figure S6.2:** Sensitivity Analysis of Different Weighting Schemes. The aerobic likelihood weighting of the aerobic likelihood categories of “Neutral”, “Likely”, and “Very Likely” was incrementally increased from 1,2,3, to 1, 100, 199. The categorization of the High Node Throughput compounds based on the Edge Aerobic Likelihood spread is shown for each PAH.

The spread was increased as much as possible while minimizing Network Outliers and maximizing Perfect Match Compounds. The optimal Edge Aerobic Likelihood scheme was found to be a spread of 32, resulting in a numerical scoring scheme of 1,33,65 for the 3 different categories. A spread of 33 would have caused 3 of the Fluorene metabolites to fall below the High Node Throughput cutoff of 0.01. This process maximized the number of Perfect Match metabolites that showed a Node Throughput of greater than 0.01.

**High Node Throughput Threshold Selection**

The threshold of 0.01 was used as an initial screening threshold for optimizing the edge weighting. After the ideal edge weighting was selected, the threshold was re-verified by considering the number of literature compounds that fell above possible thresholds vs the total number of compounds selected. 0.01 proves to be a good threshold for the definition of High Node Throughput, identifying 34 of the possible 55 (~62%) of the Perfect Match category metabolites while selecting a reasonable number of other predicted metabolites to consider. This analysis is summarized in Figure S6.3.



**Figure S6.3**: Selection of the Threshold for High Node Throughput. Both axes are log10 transformed. The heatmap shows how many literature metabolites are found above each order of magnitude of Node Throughput. The dashed line shows the threshold for High Node Throughput of 0.01 used in the main text. Each literature compound is plotted with its Node Throughput value on the vertical axis and the total number of High Node Throughput Compounds that would be selected if that value was used as the threshold on the horizontal axis (the number of High Node Throughput Compounds with a Node Throughput value greater than Throughput value of the Perfect Match value plotted)