

芳香族化合物による土壌攪乱における微生物遺伝子プールの変動

Fluctuation of Gene Pools of Soil Microbiota Spiked by Aromatic Hydrocarbons

加藤 広海, 大坪 嘉行, 永田 裕二, 津田 雅孝*

HIROMI KATO, YOSHIYUKI OHTUBO, YUJI NAGATA, and MASATAKA TSUDA

東北大学・大学院生命科学研究所 〒980-8577 仙台市青葉区片平 2-1-1

* FAX: 022-217-5699

* E-mail: mtsuda@ige.tohoku.ac.jp

Graduate School of Life Sciences, Tohoku University, 2-1-1 Katahira, Sendai 980-8577, Japan

キーワード: 土壌環境, メタゲノム, 環境汚染物質分解, 多様性

Key words: soil environment, metagenome, biodegradation of environmental pollutants, biodiversity

(原稿受付 2010年11月22日/原稿受理 2010年12月1日)

1. はじめに

2004年に Venter らが海洋細菌群集メタゲノムの大規模塩基配列解析¹⁸⁾を行って以来、鉱山排水¹⁷⁾やヒト腸内^{3,7)}など様々な環境に棲息する細菌群集メタゲノムの大規模塩基配列解析が行われ、各環境中の生物多様性やそれら生態系が示す機能ポテンシャルの類推がなされてきた。一方、土壌環境は、Tringe らのミネソタ農場土壌棲息細菌群集メタゲノムの大規模塩基配列解析¹³⁾や、Roesch らの土壌微生物の多様性解析⁹⁾により、最も生物多様性に富んだ環境のひとつであることがわかってきたが、この意味で、目の前にある一握りの土が人類に残された最後の秘境とみなせる。著者らの研究室では、このような土壌の未知遺伝子資源の探索を目的として、多環芳香族炭化水素化合物 (polycyclic aromatic hydrocarbon, PAH) や高度塩素化有機農薬である γ -HCH (γ -hexachlorocyclohexane) などの環境汚染物質の分解能力を宿主細菌に付与する土壌メタゲノムライブラリーのスクリーニングを実施し^{5,15)}、芳香族化合物酸化酵素群をコードする DNA 断片の土壌メタゲノムからの取得を行ってきた。この研究過程で、芳香族化合物分解特異的ジオキソゲナーゼ遺伝子群は、芳香族化合物汚染経験のない土壌では効率的スクリーニングに十分なほどにはメタゲノム内で濃縮しておらず、土壌に芳香族化合物を添加して濃縮することで、遺伝子取得が容易になることを経験した。では、このような分解遺伝子群の濃縮前後の背景で、どのような遺伝子プールの変動が起きているのだろうか? 様々な土壌環境におけるメタゲノムの大規模塩基配列解析はこの 1-2 年で大変盛んになってきており、土壌細菌群集のトランスクリプトーム等のオミックス研究も行われるようになってきた¹¹⁾。しかしこれら研究は土壌環境のある一時点のみをとらえたメタゲノムのスナップショットであり、トランスクリプトーム研究の

根底となるメタゲノムがどの程度安定 / 不安定なのか、どのように変動しているかについての知見は謎のままである。

これらの研究を踏まえた上で我々は、PAH 汚染で変動する土壌微生物群集の遺伝子プールの変動を包括的に理解することを目的とした研究を進めている (図 1)。実際には、土壌の PAH 汚染後に経時的に調製したメタゲノムを大規模塩基配列解析に供するとともに、当該土壌微生物群集からの分解細菌 / コミュニティの分離やメタゲノムからの機能性遺伝子のスクリーニングとその後の解析を実施している。自然環境を対象にした研究では、得た結果の妥当性の検証や更なる補足的データ取得をあとから行うことが一般的に困難であるが、我々は、時系列に伴うサンプルを適宜保存することで、大規模塩基配列解析のような “dry” な研究と細菌株分離や遺伝子スクリーニングのような “wet” な研究を互いにフィードバックできる体制を組んだ。これら包括的研究の一端として、本稿では汚染土壌メタゲノムの大規模塩基配列解析による土壌微生物群集の遺伝子プールの変動について紹介する。

2. 経時的なメタゲノム解析

我々は、土壌に接種した単一細菌株の生きざまをゲノムレベルで解明することを目的とした別研究において、芳香族化合物による汚染歴の無い土壌 (愛媛県農林水産研究所) を用いてきた^{4,14,16)} が、今回紹介する研究においても、この土壌を用いることにした。実際に起きる環境汚染においては、例えば原油汚染を想定した場合、単一芳香族化合物よりも、易分解性および難分解性の複数の芳香族化合物によって汚染される。そこで本研究では、易分解性芳香族化合物として 3-クロロ安息香酸 (3CB)、難分解性芳香族化合物としてフェナントレン、

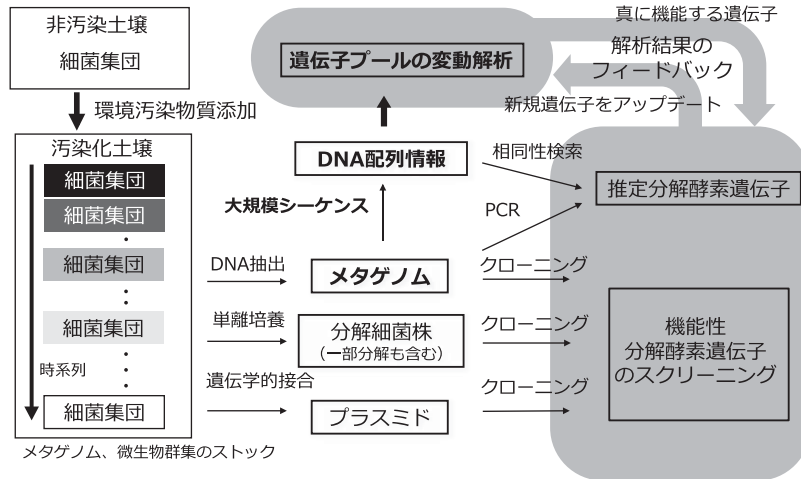
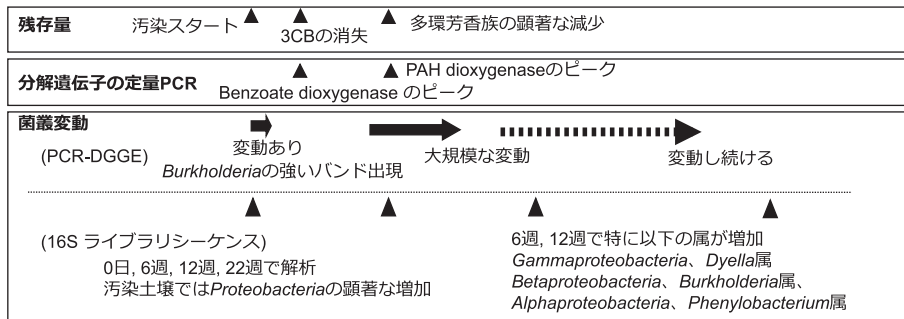
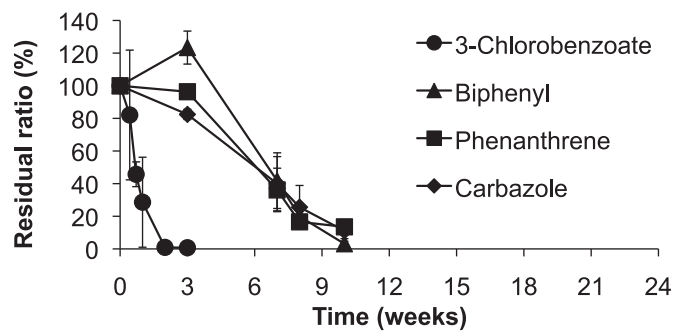


図1. 土壌遺伝子プールの包括的理解を目指した研究の全体像



メタゲノム解析のタイムポイントの選定

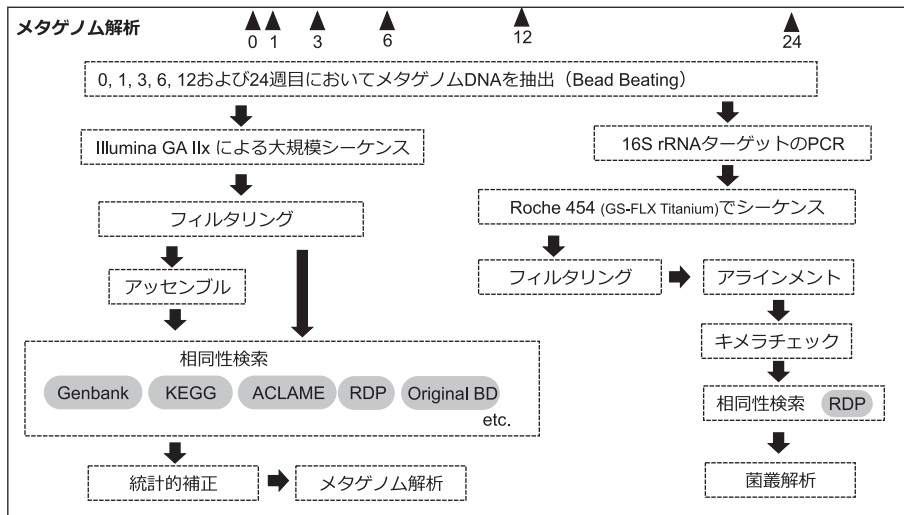


図2. 経時的に調製したメタゲノム DNA の解析フロー

ピフェニルおよびカルバゾールの計4種類化合物による複合汚染を設定した。これら物質をメタノールに溶かしたあとにセライトに吸着させ、終濃度約125 ppmで土壌と混合することで汚染土壌とした。対照区としては芳香族化合物を吸着させていないセライトを混合した土壌(コントロール土壌)を用意した。土壌中での残存量を経時的に定量した結果、3CBは1週間から2週間後までに分解され、他3種化合物は6週目前後を境に分解され始め、10週目から12週目後には検出されなくなった(図2)。

さて、どの時点でメタゲノムの大規模塩基配列解析を行うべきかを判断するのは、遺伝子プールの変動を調べる上で最も重要かつ難しい問題である。そこで我々は、汚染開始後の異なる時期から調製した土壌メタゲノムDNAについて、予めPCR-DGGE法や16S rRNAクローンライブラリー解析、そして既知の各種芳香族化合物酸化酵素遺伝子のreal-time PCRを行うことで、遺伝子プールの変動をある程度把握することにした。これらの情報を基に、汚染直前(0週目)、1、3、6、12、および24週目をメタゲノムの大規模塩基配列解析の時点とすることで、汚染土壌での遺伝子プールの変動がコントロール土壌でのそれとどのように異なるかを調べることにした。実際の解析フローは図2に示した。細菌種によるDNA抽出効率のばらつきを少なくするようビーズビティングによって抽出したメタゲノムDNAは、Illumina社のGenomic Analyzer Iixを用いて大規模塩基配列決定(75塩基をペアドエンドで解読)した。得られたリードに対してクオリティ等によるフィルタリングを行った結果、各サンプルについて約1千万リード得られ、合計で10 Gb相当の土壌メタゲノム塩基配列情報を得た(表1)。実はこの時点でシークエンサーから出力される生リードデータの1/8にまで減少しているわけだが、どの程度のクオリティフィルターを採用するかは、研究者の裁量に依る部分も多く、今後もセンシティブな問題であろう。最終的に得られた高クオリティリードを、RDP(Ribosomal Database Project)やKEGG(Kyoto Encyclopedia

of Genes and Genomes)、ACLAME(A Classification of Mobile Genetic Elements)、GenBank-nr等の各種データベースへのBLAST解析へ供した。また、種々の芳香族化合物代謝に関わる酵素遺伝子に関してはマニュアルでデータベースを構築した。これらのデータベースにヒットしたリードを用いることで、遺伝子プールの菌叢解析や代謝パスウェイ解析を行った。

3. 微生物群集構造の変動

我々のメタゲノム解析では、細菌群集構造の解析を最終目的としているわけではなく、むしろ、“メタゲノム”としてどのような代謝機能ポテンシャルを有するのかに主眼を置いている。ただ、メタゲノム内の遺伝子プールの変動が個々の細菌細胞の増加/減少によって引き起こされることから、遺伝子プール変動を理解するうえでは、菌叢変動の把握が已然として重要である。RDPでのBLAST結果より16S rRNA遺伝子に帰属されたIlluminaリードは、各サンプルで5000–15000リードであった(表1)。これらリードを使って、芳香族化合物による攪乱が菌叢に与えた影響を評価したところ、コントロール土壌では小規模な、汚染土壌では極めて大規模かつ長期的な変動が起きていた(図3)。コントロール土壌ではMethylotrophの一時的な増加が見られたが、これは、セライトに吸着させたメタノールが残存していた影響と考えられる。一方4種の芳香族化合物による汚染土壌では、3CBが分解される1週目にBurkholderiaやPseudomonasなどのProteobacteriaが優占し、その他3種多環芳香族化合物が顕著に分解される6週目においてMycobacteriumなどのグラム陽性菌の増加が観察された。これら細菌群は芳香族化合物分解細菌としても従来からよく知られているが、本土壌においても汚染物質の分解への直接あるいは間接的に関与し、細胞数を増加させることで、メタゲノム内における16S rRNA遺伝子の割合を増加させたと解釈できる。Illuminaリードを16S rRNA遺伝子に帰属させる方法のみでは、(1)属決定に

表1. 各土壌メタゲノムDNAの解析に用いたリード数および各種データベースにおけるBLAST結果

sample ^a	Total (Read)	KEGG ^b		16S rRNA ^c		PAH dioxygenase ^d	
		(Read)	(%)	(Read)	(%)	(Read)	(%)
0	12,561,547	2,537,108	20	5,834	0.05	324	0.003
1C	13,367,184	3,265,978	24	11,614	0.09	409	0.003
3C	8,584,326	2,066,327	24	4,671	0.05	219	0.003
6C	11,912,537	2,538,976	21	5,790	0.05	308	0.003
12C	11,022,660	2,385,055	22	4,715	0.04	338	0.003
24C	12,825,888	2,590,254	20	7,491	0.06	353	0.003
1M	11,562,982	3,608,931	31	14,281	0.12	1,407	0.012
3M	10,255,058	2,978,664	29	10,072	0.10	783	0.008
6M	11,331,760	2,993,158	26	9,028	0.08	629	0.006
12M	7,273,428	1,825,235	25	4,115	0.06	236	0.003
24M	14,358,223	3,375,639	24	9,737	0.07	422	0.003

^a 0–24は汚染物質が添加されてからの時間(週)を表す。C: Control soil. M: Mix contaminated soil.

^b KEGGデータベースに対してBLASTX(identity ≥ 70%, score ≥ 40)でヒットしたリード数。

^c RDPデータベースに対してBLASTN(identity ≥ 85%, coverage ≥ 80%)でヒットしたリード数。

^d PAH dioxygenaseのオリジナルデータベースに対してBLASTX(identity ≥ 70%, coverage ≥ 80%)でヒットしたリード数。

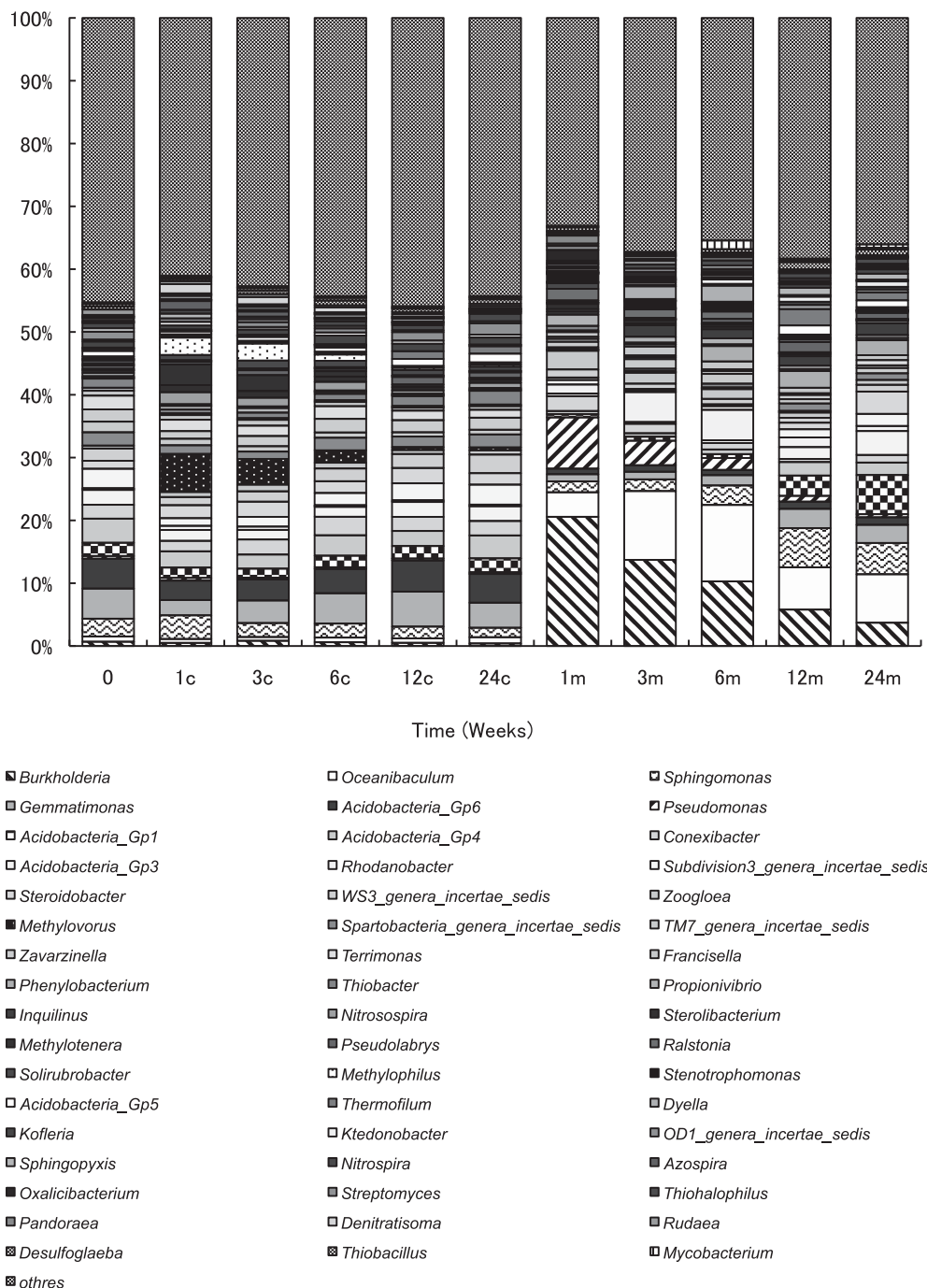


図3. 16S rRNA 遺伝子に帰属された Illumina リードによる汚染土壌菌叢解析。サンプル名の 0~24 は汚染物質が添加されてからの時間 (週) を表す。C, Control soil; M, Mix contaminated soil.

必ずしも有効でないリードが数多くあること、そして、(2) 未知細菌属の 16S rRNA 遺伝子を検出できないこと、の 2 点を克服できない。そこで、16S rRNA 遺伝子内の 450 塩基ほどを特異的増幅且つ属判別が可能でプライマー対を共同研究先の東工大・黒川グループが設計し、これを用いて、上記各時点のメタゲノム DNA から 16S rRNA 遺伝子断片を PCR 増幅し、Roche 454 で塩基配列を解読した。この結果、既知属に帰属できる PCR 増幅産物は、コントロール土壌由来で約 75%、汚染土壌由来で 65~85%であったものの、Illumina と 454 の解析で明らかになった既知属細菌の菌叢変動様式には大きな矛

盾はなかった。

ここまで述べてきたメタゲノムにおける“菌叢変動”の解析では、抽出した DNA における“割合の変動”を見ていることになる。そこで実際に 3 属ほどの細菌について、real-time PCR を用いて“絶対数”の変化⁶⁾を求めたところ、Illumina リードを使った相対量と高い相関を得た (図 4)。従って、Illumina リードを使ったメタゲノムの“割合の変動”は、ある程度絶対数を反映した結果であるといえる。菌叢解析の結果に戻るが、汚染土壌とコントロール土壌では、変動後に元の菌叢へと収束するスピードが異なっている点も興味深い。コントロール

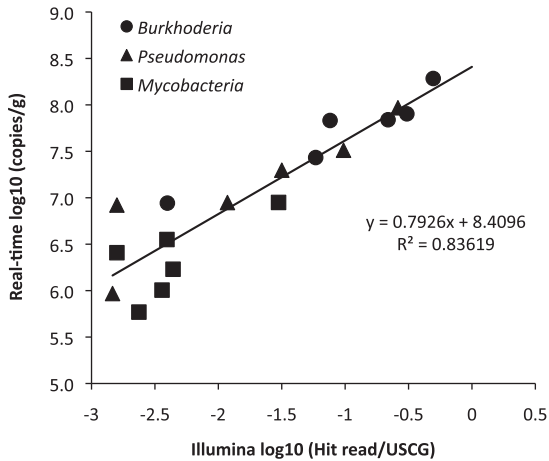


図4. Real-time PCR を用いた各細菌グループの絶対数 (16S rRNA 遺伝子コピー数) と Illumina リードによる菌叢解析の相関。Illumina リードはサンプル間の比較のため universal single copy gene (USCG, Raes ら⁸⁾ のリストに *gyrB* を足したもので標準化した。

土壌での変動は速やかに収束したのに対し、汚染土壌では 24 週目 (汚染物質除去後 12 週間経っても) においても収束していない。この違いは多様度の経時変化においても顕著に現れている (図5)。芳香族化合物の代謝産物、あるいは増殖した分解菌が減少していく際の遺体有機物などがその他の菌の栄養になるなど、芳香族化合物添加によって二次的におきる土壌攪乱が、菌叢変動の収束を妨げているのかもしれない。

4. 芳香族代謝遺伝子群の変動

上記のように、芳香族化合物添加によって土壌微生物群集は大規模な攪乱を受けたことが理解できる。では、土壌微生物群集遺伝子プールの代謝機能ポテンシャルにどのような変動を引き起こしたのだろうか。得られたリードを KEGG データベースに対して BLAST 解析を行った結果、1/4 程のリードが何らかの既知遺伝子に帰属できた。本研究で用いた Illumina では前述のように最大 75 塩基と解読長が短く、その 3/4 のリードが BLAST 解析でもデータベースにヒットしなかった。しかし Illumina では莫大なリード数が得られることから、1/4 でもかなりの情報を得られると予想したが、実際 2 Gb 程度の代謝パスウェイ情報を解析することができた。土壌汚染は微生物の多様な代謝パスウェイに影響を及ぼしたが、ここでは芳香族化合物代謝における変動解析の結果を紹介する。

土壌に添加した芳香族化合物であるビフェニルの分解に関する KEGG パスウェイマップ上において、各代謝反応を担う EC 番号に帰属されたリード存在量および細菌種の組成を円グラフでプロットした (図6)。汚染直前時点においても、遺伝子プールにある程度のビフェニル分解代謝経路の遺伝子が存在していたが、汚染後 1 週間目には各 EC 番号においてヒット数の大幅な増加が認められた。このように経時的なメタゲノム解析を行うことで、遺伝子プール内でビフェニル代謝遺伝子の割合が増加するとともに、それら遺伝子が帰属する細菌

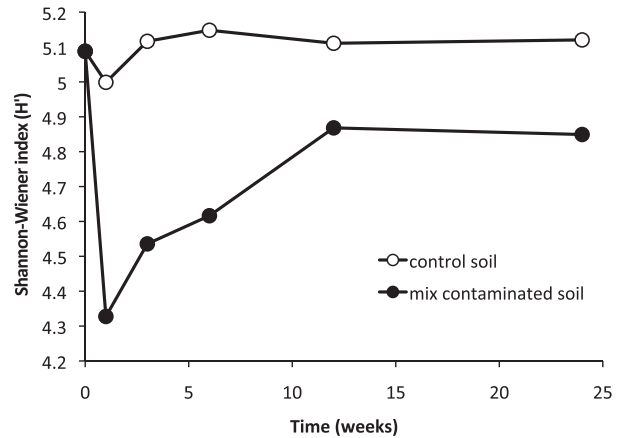


図5. 汚染土壌における多様度の経時変化。多様度は Shannon-Wiener index¹⁰⁾ をもとに計算した。

群の組成も経時的に変動する様子を見ることができた。興味深いことに、ビフェニルが安息香酸および 2-ヒドロキシ-2,4-ペンタジエン酸に酸化され、さらに下流代謝へと流れていく各過程では、関与する細菌群の組成が大きく異なっていた。ある反応では *Proteobacteria* が、別の反応ではグラム陽性の *Actinobacteria* が優先していた。このような現象は、フェナントレン等の他芳香族化合物の酸化的分解経路においても確認された。また、ビフェニルの初発酸化酵素遺伝子は 1 週目において大きく増加するが、実際に分解されるのは中間代謝反応に *Actinobacteria* が増えてくる 6 週目以降である点も興味深い。この時カテコール以降の下流分解代謝には再び *Proteobacteria* の関与が優先した。これら代謝経路での入れ子状態から、土壌環境での複数細菌株による協調的な代謝の可能性が示唆できる。Suenaga らの研究¹²⁾ では、活性汚泥由来メタゲノムライブラリーから、カテコール分解活性を大腸菌に付与する複数 DNA 断片を解析しており、この結果、既知の分離培養可能な細菌株で見られるような高度に組織化されたオペロン構造を持たない DNA 断片が多数見つかった。この知見は、環境中の微生物代謝は既知の分解細菌ほど洗練されておらず、不完全な代謝を持つ細菌が少なからず存在していることを示唆している。本研究の土壌環境においても、不完全な代謝を持つ細菌たちが寄り合って汚染物質を除去しているのかもしれない。

遺伝子プールの変動を代謝経路の上流から下流へ追う「縦のつながり」を解析するだけでなく、広範な他の代謝との関連を探る「横のつながり」の解析も我々は試みている。経時的なメタゲノム解析では、各遺伝子に対して変動率が計算できるので、ある遺伝子とその他遺伝子の変動率の同調性を調べることができる。例えば先のビフェニルジオキシゲナーゼ (EC1.14.12.18) では、本酵素遺伝子と変動率相関係数が 0.9 以上を示す EC 番号酵素をコードする遺伝子が約 400 個存在し、その中には芳香族化合物代謝に関わるものもあれば、そうでないものもある。著者らはこれらを、変動する遺伝子プールにおける「連動遺伝子群」と考えた。このような連動を引き起こす理由として、分解細菌株ゲノムに存在する遺伝子群が、その細胞数増減によって連動する場合や、汚染物

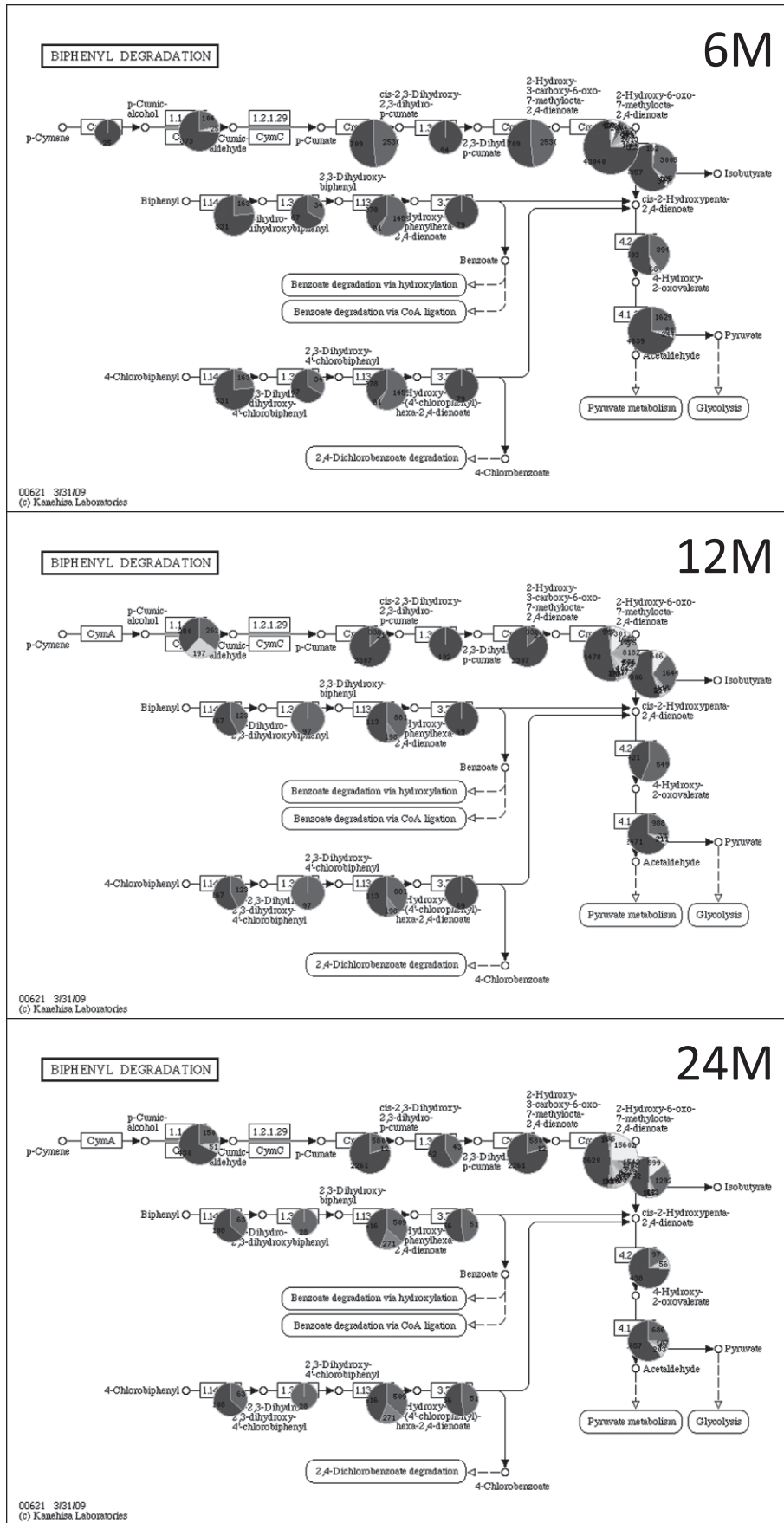


図6. ビフェニル代謝パスウェイにおける汚染土壌の遺伝子プールの変動。サンプル名の0-24は汚染物質が添加されてからの時間(週)を表す。C, Control soil; M, Mix contaminated soil. 円グラフ内の数値は存在量。カラーインデックスは *Proteobacteria*, black; *Actinobacteria*, gray; *Firmicutes*, light gray; *Acidobacteria*, white.

質を直接分解できずとも分解菌株と何らかの協調関係にある別細菌株が持つ遺伝子が連動する場合などを想定している。しかし無理に個体に落とし込まないまでも、遺伝子プールとして連動遺伝子群を想定することで、環境中で分解活性を発現するのに必要な遺伝子のセットについて、新しい知見を得ることができないのではないかと期待している。

5. おわりに

2010年秋に名古屋で生物多様性条約第10回締約国会議(COP10)が開催され、生態系の多様性をどのようにして保全するかの議論が世界的レベルでなされた。土壌は、鉱山排水や海底の底泥のような一定の選択圧がかかり続ける環境ではなく、極めて多様な攪乱に曝されている。攪乱と多様性の研究は、微生物生態学の分野でも近年盛んに行われており、特に生態系が攪乱から回復する性質(レジリエンス)が大きく取りあげられている¹²⁾。しかしながら、レジリエンスを生み出すメカニズムは謎のままである。本研究における菌叢と代謝パスウェイの変動解析においても、両者を結びつけるにはまだ多くの解析と課題が残されているものの、これら問題を一つずつ解決していくことで得られる知見とそのメカニズムは、多様な生態系での「謎」の解明にも大きく貢献できると考えている。

謝 辞

本研究におけるメタゲノムの大規模塩基配列解析は国立情報研究所藤山秋佐教授・国立遺伝学研究所豊田敦特任准教授のグループ、またバイオインフォマティクスは東京工業大学黒川顕教授グループとの共同研究で実施している。また、筆者グループの研究は、文部科学省と日本学術振興会の科学研究費補助金、東北大学「生態適応」グローバルCOEからの助成を受けた。この場にて関係各位に謝意を表したい。

文 献

- Allison, S.D. and J.B.H. Martiny. 2008. Resistance, resilience, and redundancy in microbial communities. *Proc. Natl. Acad. Sci. U.S.A.* 105: 11512–11519.
- Botton, S., M. Van Heusden, J.R. Parsons, H. Smidt, and N. Van Straalen. 2006. Resilience of microbial systems towards disturbances. *Crit. Rev. Microbiol.* 32: 101–112.
- Kurokawa, K., T. Itoh, T. Kuwahara, K. Oshima, H. Toh, A. Toyoda, H. Takami, H. Morita, V.K. Sharma, T.P. Srivastava, T.D. Taylor, H. Noguchi, H. Mori, Y. Ogura, D.S. Ehrlich, K. Itoh, T. Takagi, Y. Sakaki, T. Hayashi, and M. Hattori. 2007. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res.* 14: 169–181.
- Nishiyama, E., Y. Ohtsubo, Y. Nagata, and M. Tsuda. 2010. Identification of *Burkholderia multivorans* ATCC 17616 genes induced in soil environment by *in vivo* expression technology. *Environ. Microbiol.* 12: 2539–2558.
- Ono, A., R. Miyazaki, M. Sota, Y. Ohtsubo, Y. Nagata, and M. Tsuda. 2007. Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches. *Appl. Microbiol. Biotechnol.* 74: 501–510.
- Park, J.K. and D.E. Crowley. 2005. Normalization of soil DNA extraction for accurate quantification real-time PCR and of target genes by DGGE. *BioTechniques* 38: 579–586.
- Qin, J., R. Li, J. Raes, M. Arumugam, K.S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D.R. Mende, J. Li, J. Xu, S. Li, D. Li, J. Cao, B. Wang, H. Liang, H. Zheng, Y. Xie, J. Tap, P. Lepage, M. Bertalan, J.-M. Batto, T. Hansen, D.L. Paslier, A. Linneberg, H.B. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. Zhu, C. Yu, S. Li, M. Jian, Y. Zhou, Y. Li, X. Zhang, S. Li, N. Qin, H. Yang, J. Wang, S. Brunak, J. Doré, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, M. Consortium, P. Bork, S.D. Ehrlich, and J. Wang. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464: 59–65.
- Raes, J., J.O. Korbel, M.J. Lercher, C. von Mering, and P. Bork. 2007. Prediction of effective genome size in metagenomic samples. *Genome Biol.* 8: R10.
- Roesch, L.F.W., R.R. Fulthorpe, A. Riva, G. Casella, A.K.M. Hadwin, A.D. Kent, S.H. Daroub, F.A.O. Camargo, W.G. Farmerie, and E.W. Triplett. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J.* 1: 283–290.
- Shannon, C.E. and W. Weaver. 1949. The mathematical theory of communication. The University of Illinois Press, Urbana, U.S.A.
- Shrestha P.M., M. Kube, R. Reinhardt, and W. Liesack. 2009. Transcriptional activity of paddy soil bacterial communities. *Environ. Microbiol.* 11: 960–970.
- Suenaga, H., Y. Koyama, M. Miyakoshi, R. Miyazaki, H. Yano, M. Sota, Y. Ohtsubo, M. Tsuda, and K. Miyazaki. 2009. Novel organization of aromatic degradation pathway genes in a microbial community as revealed by metagenomic analysis. *ISME J.* 3: 1335–1348.
- Tringe, S.G., C. von Mering, A. Kobayashi, A.A. Salamov, K. Chen, H.W. Chang, M. Podar, J.M. Short, E.J. Mathur, J.C. Detter, P. Bork, P. Hugenholtz, and E.M. Rubin. 2005. Comparative metagenomics of microbial communities. *Science* 308: 554–557.
- 津田雅孝, 永田裕二, 大坪嘉行. 2007. 土壌環境細菌の比較ゲノム. pp. 166–172. 藤山秋佐夫監修. 比較ゲノム学から読み解く生命システム—基本概念から最新ゲノム情報まで—. 秀潤社.
- 津田雅孝, 小野 玲, 宮崎 亮, 府中玄樹, 永田裕二. 2007. 機能発現に基づく環境汚染物質分解酵素遺伝子の生態系からの直接的取得と解析. *環境バイオテクノロジー学会誌.* 7: 75–78.
- 津田雅孝, 西山依里, 永田裕二, 大坪嘉行. 2007. 自然環境で実際に機能する微生物遺伝子の遺伝学的手法による検索と解析. *化学と生物.* 45: 557–563.
- G.W. Tyson, J. Chapman, P. Hugenholtz, E.E. Allen, R.J. Ram, P.M. Richardson, V.V. Solovyev, E.M. Rubin, D.S. Rokhsar, and J.F. Banfield. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428: 37–43.
- Venter, J.C., K. Remington, J.F. Heidelberg, A.L. Halpern, D. Rusch, J.A. Eisen, D.Y. Wu, I. Paulsen, K.E. Nelson, W. Nelson, D.E. Fouts, S. Levy, A.H. Knap, M.W. Lomas, K. Neelson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y.H. Rogers, and H.O. Smith. 2004. Environmental genome shotgun sequencing of the Sargasso sea. *Science* 304: 66–74.