

次世代 DNA シーケンサーのバイオレメディエーションへの利用

Application of Next Generation DNA Sequencers for Bioremediation

養王田 正文
MASAFUMI YOHDA

東京農工大学大学院工学府生命工学専攻 〒184-8488 東京都小金井市中町 2-24-16

TEL & FAX: 042-388-7479

E-mail: yohda@cc.tuat.ac.jp

Department of Biotechnology and Life Science, Tokyo University of Agriculture and Technology,
2-24-16 Naka-cho, Koganei, Tokyo

キーワード: 次世代 DNA シーケンサー, バイオレメディエーション, デハロココイデス, メタゲノム, クロロエテン

Key words: Next Generation DNA sequencer, Bioremediation, *Dehalococcoides*, metagenome, chloroethene

(原稿受付 2014年2月28日/原稿受理 2014年3月4日)

1. はじめに

ヒトゲノム解析で大きな役割を演じた DNA シーケンサーはさらに進化を続け、膨大な DNA 配列情報を低コスト・短時間で解析できるようになった。この次世代 DNA シーケンサーと総称されている新しい DNA シーケンサーは、生命科学と医療で革命を起しており、その革命の波は環境バイオテクノロジーにも及んでいる。バイオレメディエーションは低コストで環境に優しい浄化手段として、環境浄化の一つの分野として位置づけられており、利用拡大が期待されている。しかし、現状ではその利用は伸び悩んでいる。これは、バイオレメディエーションによる浄化の効果・信頼性、及び安全性に関する問題によるものである。こういったバイオレメディエーションの様々な課題も、次世代 DNA シーケンサーで解決される可能性がある。本稿では、特に膨大な DNA 配列解析が可能なショートリード型次世代 DNA シーケンサーのバイオレメディエーションへの利用の可能性について、我々の研究成果を基に議論する。

2. DNA シーケンサーの進歩

ゲノム解析の進歩は、DNA シーケンサーの進化に支えられている。蛍光 DNA シーケンサーの開発により簡単に DNA シーケンスが行われるようになった。さらに、キャピラリーアレイ DNA シーケンサーの発明により大規模なゲノム解析が可能になり、2000年にヒトゲノム解析プロジェクトの実質的成功が発表され、翌年2つのヒトゲノム配列が報告された^{1,17)}。しかし、ヒトゲノム情報を医療などで活用するためには、少数の標準ゲノムの解析だけでは不完全であり、多数のヒトのゲノムを比較解析することが必要である。キャピラリー DNA シーケンサーによるヒトゲノム解析では膨大なコストが

かかることから、DNA シーケンスの低コスト化が求められていた。そこで、ヒトゲノム解析のコストを1000ドル以下にするという1000ドルゲノムプロジェクトが提唱された。NHGRI (National Human Genome research Institute) からの膨大な研究費と、ヒトゲノム解析プロジェクトにおいて開発が進められていた技術シーズにより、高速かつ低コスト DNA シーケンスを可能とする次世代 DNA シーケンサーが予想よりも早く開発された。

2008年4月17日発行の Nature 誌に二重らせん構造の発見者 J.D. Watson 博士の全ゲノム配列が発表された¹⁹⁾。ワトソン博士のゲノムは、454社 (現 Roche 社) が開発したピコタイタープレートとパイロシーケンスという化学発光検出を利用した DNA 解読技術を組み合わせた DNA シーケンサーにより解析され、わずか2カ月で終了している。コストはサンガー法を用いた場合の1/100の約100万ドルであった。Solexa社 (現 illumina 社) と Applied Biosystems社 (現 Life Technologies 社) は、2007年に蛍光検出を利用した次世代シーケンサーをそれぞれ発売した。2008年、Solexa社は6台の装置を用いて、8週間、25万ドルでヒトゲノムを解読した¹⁾。Applied Biosystems社は、SOLiDシステムを用いてヨルバ族アフリカ人男性の全ゲノム配列を6万ドル以下のコストで解析した⁸⁾。その後、3社による解読費用の低減競争が続いており、1000ドルゲノムの実現が近づいている。

サンガー法以外の DNA シーケンス解析技術は、ヒトゲノム解析完成までに目の目を見ることはなかった。サンガー法では、非常に長い配列の解析が可能であり、他のシーケンス技術と一線を画すところである。未知の領域のシーケンスを行う上で解析できる配列が長いほど、その配列情報の質は高くなる。特に、ヒトゲノムのように、繰り返し配列を含む場合には、その効果はさらに大きい。このため、ヒトゲノム解析ではサンガー法の独壇

表 1. 次世代 DNA シーケンサーの性能比較

メーカー	機種名	原理	リード長	リード数/ラン	解読量/ラン	解析時間
Roche	GS FLX Titanium	PS	500	>100 万	400–500 Mb	10 時間
〃	GS Junior	PS	400	10 万	35 Mb	10 時間
illumina	HiSeq2500	SBS	2×100 bp (pair end)	60 億	540–600 Gb	11 日
〃	MiSeq	SBS	2×250 bp (pair end)	3000–3400 万	7.5–8.5 Gb	39 時間
Life Technologies	5500xl SOLiD	SBL	2×60 bp (pair end)	>24 億	270 GB	7 日
〃	Ion Torrent PGM 316/318 Chip	PS	200 bp	100 万 (316) 500 万 (318)	100 Mb (316) 1 Gb (318)	数時間
〃	Ion Proton	PS	200–400 bp	5000 万	10 Gb	数時間
Pacific Bio	PACBIO RS	SM	3000 bp	22000/cell, 12 cell	65 Mb/cell	2 時間 /cell

原理

PS: pyrosequence, SBS: sequence by synthesis, SBL: sequence by ligation, SM: single molecule

場であり、他の技術での解析は不可能であった。しかし、ヒトゲノム解析が完成し、お手本となるシーケンスが得られた現在では、短い配列でもその情報を有効に利用することが可能になり、次世代シーケンサーによるゲノム解析が可能になった。現在最も高性能な DNA シーケンサーは illumina 社の HiSeq2500 であり、1 台で 6000 億塩基の膨大な DNA 配列データを解析できる。単純計算で、ヒトゲノム配列の 200 倍のデータ量である。次世代 DNA シーケンサーの開発は現在も続いており、2011 年に Pacific Biosciences 社は長塩基解読を実現する単分子タイプの次世代シーケンサーを開発した。Ion Torrent 社（現 Life Technologies 社）は、イオン感応型電界効果トランジスタを用いたプロトン検出による DNA 解読技術を発表し、製品化した¹³⁾。さらに、究極の DNA シーケンサーとして、ナノメートルサイズの細孔を通過する DNA による電流値変化に基づく DNA 解読装置ナノポア DNA シーケンサーが発表されたが、製品化には至っていない¹⁰⁾。現在販売されている次世代 DNA シーケンサーの性能を表 1 にまとめた。最初に開発された 3 種類の次世代 DNA シーケンサーの中で、Life Technologies 社の SOLiD 及び illumina 社のシーケンサーは、解析長が短いという欠点があるが、膨大な数の解析が可能である。本稿では、これらのショートリード型次世代次世代 DNA シーケンサーに焦点を当てて、環境バイオテクノロジーへの利用を我々の研究成果を基に提案する。

3. 環境バイオテクノロジーへの次世代 DNA シーケンサーの利用

DNA シーケンスは医療以外の分野でも重要な役割を担っている。特に培養できない環境中の微生物の情報を得るには極めて有効な技術である。ヒトゲノム解析プロジェクトにおける一方の立役者である C.J. Venter 博士はサルガッソ海の海水中の微生物のメタゲノム解析を行い、10 億塩基の配列を決定し、120 万個の新しい遺伝子を発見した¹⁸⁾。キャピラリー電気泳動を用いていたことからデータ量も限られており、予想された微生物の 20%程度しか発見できなかったが、他の実験手法では決

して得ることができない情報を得られることが証明された。未知の微生物資源を獲得する最も有効な方法であり、次世代 DNA シーケンサーの活躍の場として考えられている。

バイオレメディエーションは微生物の能力を利用して環境を浄化する技術であることから、浄化に用いる微生物の特性と、環境中での微生物の状況を理解することが不可欠であり、DNA シーケンスは重要な情報として利用されていた。これまで、単離された浄化微生物のゲノム解析、Real Time PCR による分解微生物の定量分析、16S rRNA の PCR-DGGE 法による分析、マイクロアレイによる菌の特定などが行われてきた。しかし、難分解の汚染物質を浄化する微生物の単離培養は困難な場合が多く、その解析と応用の大きな壁となっている。

テトラクロロエテン (PCE) とトリクロロエテン (TCE) などの揮発性有機塩素化合物の脱塩素化はデハロコッコイデス (*Dehalococcoides*) 属細菌を含む数種類の微生物により行われる³⁾。特に、ジクロロエテン (DCE) 異性体から塩化ビニル (VC)、さらにはエテンまでの分解は、一部のデハロコッコイデス属細菌 (*Dehalococcoides mccartyi* BAV1 株, VS 株など) のみが可能である^{2,4)}。一般的に、エテンまでの完全な脱塩素化には複数の微生物の存在が必要だと考えられている。ポリ乳酸等の水素徐放剤を電子供与体として土壌に供給しても、これらの微生物が存在しなければ、全く分解が進まないか DCE や VC が土壌中に蓄積することになる。このような問題を解決するために、デハロコッコイデス属細菌などの供与を行うことで分解を促進するバイオオーグメンテーションが試みられている。しかし、デハロコッコイデス属細菌を含むクロロエテン類の嫌氣的脱塩素化を行う微生物は単離培養が困難であり、利用可能な微生物が限られている。デハロコッコイデス属細菌の中で唯一 *Dehalococcoides mccartyi* 195 が PCE からエテンまでの完全分解を行うことが可能であるが¹¹⁾、*D. mccartyi* 195 には VC 分解酵素は存在せず、VC は共代謝で分解されるだけなので分解能が低い。*D. mccartyi* GT は VC 分解酵素である *vcrA* を有し、TCE からエテンまでの完全脱塩素化が可能であるが、TCE 分解酵素は分かっていない¹⁵⁾。また、デハロコッコイデス属細菌には協調して働

く微生物が存在することが示唆されているが、それに関する知見も限られている。これらの問題は、PCE または TCE を完全に分解できる混合培養系を構築し、バイオオーグメンテーションに利用することで解決される可能性がある。しかし、この混合培養系に存在する微生物群に関する情報を明らかにし、その安全性などを示すことができなければ実用化は不可能である。この問題は、従来の技術では解決不可能であったが、膨大な解析力を有する次世代 DNA シーケンサーにより解決が可能であると考え、研究を行った。

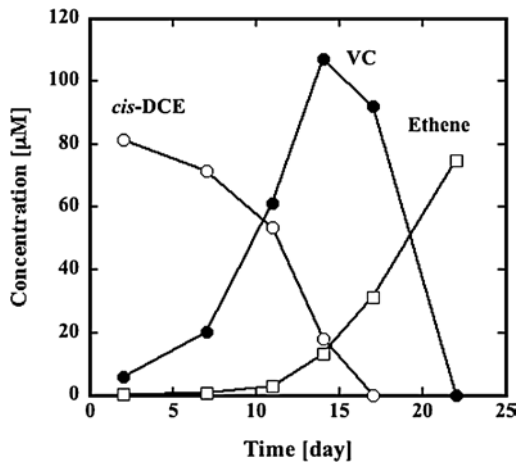


図 1. 混合培養系における cis-DCE の脱塩素化

4. cis-1,2 DCE 分解微生物混合培養系の構築と SOLiD 3 による解析

TCE で汚染されており、事前の調査でデハロココイデス属細菌が存在し水素徐放剤によりエテンまで浄化が進むことが分かっている地下水を微生物源として、cis-1,2 DCE を嫌氣的に脱塩素化する微生物混合培養系を構築した。図 1 は第 4 世代培養液中の cis-1,2 DCE, VC 及びエテンの濃度変化を示したものである。cis-1,2 DCE の分解に伴い、VC が生成し、約 3 週間でほぼ完全にエテンまで脱塩素化が進んだ。以後の継代では、ほぼ同じ速度で分解が進むことが確認された。Real Time PCR による 16S rRNA 遺伝子の解析の結果、デハロココイデス属細菌が主に占有していることが明らかになった。そこで、この培養液からメタゲノムを精製し、次世代 DNA シーケンサー SOLiD 3 により解析を行い、約 3 億個の 50 塩基のタグ配列を得た。これは約 150 億塩基分の配列情報に相当する。解析ソフトウェア Corona Lite により、得られた配列情報を既知のデハロココイデス属細菌ゲノム配列: NC_002936 (*D. mccartyi* 195)¹⁴, NC_009455 (*D. mccartyi* BAV1), NC_007356 (*D. mccartyi* CBDB1)⁶, NC_013552 (*D. mccartyi* VS), NC_013890 (*D. mccartyi* GT) にマッチングを行った。その結果を図 2 に示す。横軸はゲノム配列に対応しており、マッチングされたタグ配列の数を Coverage 数として示している。その結果、混合培養系に存在するデハロココイデス属細菌のゲノムは *D. mccartyi* 195 株のゲ

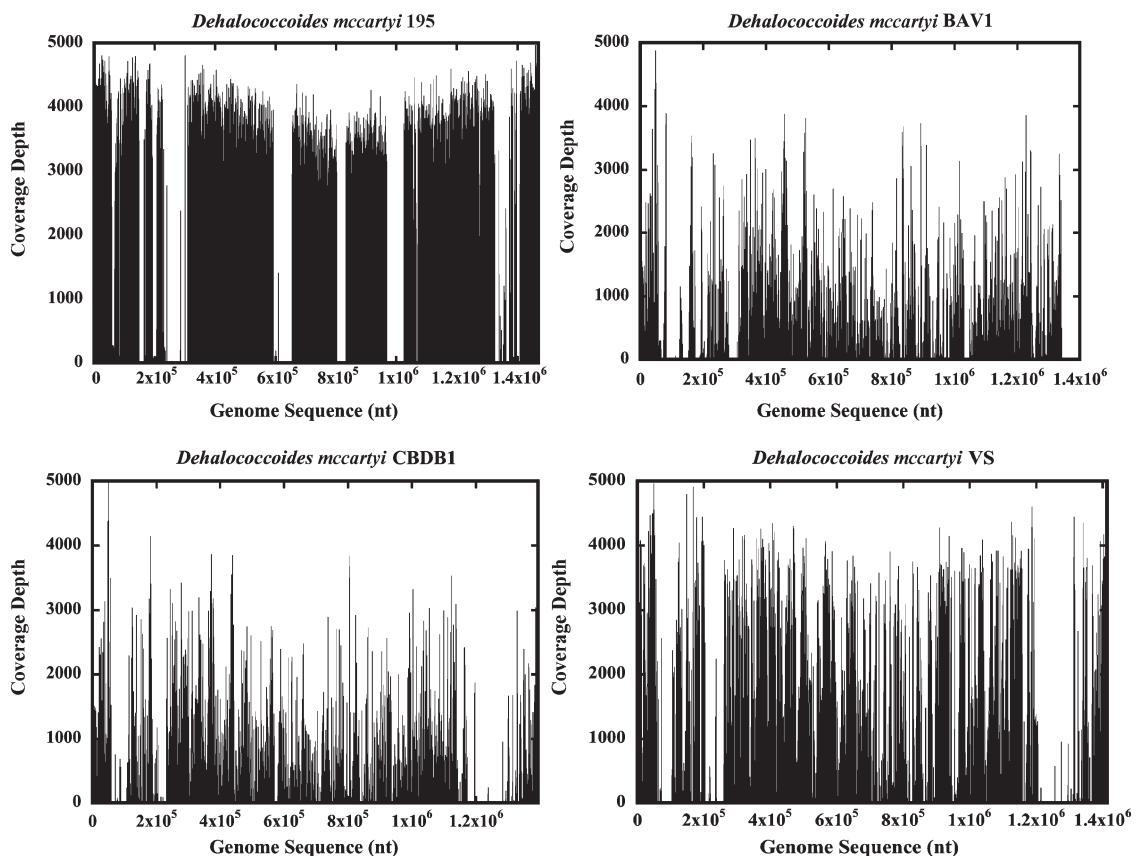


図 2. SOLiD 3 解析データとデハロココイデス細菌ゲノム配列のマッチング

ノムと全長に渡って高い相同性を示すことが明らかになった。

さらに、どのような還元脱塩素化酵素遺伝子が含まれるかを明らかにするために、各デハロココイデス属細菌のゲノムにおける還元脱塩素化酵素遺伝子とのマッチングを解析した。その結果、*D. mccartyi* 195 のゲノムに

表 2. *Dehalococcoides mccartyi* 195 の還元脱塩素化遺伝子との比較

Gene	Strand	Start	Stop	Coverage
DET0079 (TCEase)	-	77229	78893	2042
DET0173	+	167859	169391	836
DET0180	+	173382	174749	3718
DET0235	+	226290	227762	54
DET0302	+	290049	291593	1
DET0306	+	294215	295732	0
DET0311	+	298706	300253	0
DET0318	+	304666	306153	0
DET0876	-	803653	805185	0
DET1171	-	1067812	1069410	0
DET1519	-	1371471	1372988	0
DET1522	-	1374234	1375757	905
DET1528	-	1379479	1380888	3373
DET1535	-	1384117	1385601	2500
DET1538	-	1386684	1388162	8
DET1545	-	1392249	1393751	409
DET1559	-	1404340	1405788	0

表 3. デハロココイデス細菌 16S rRNA 及び還元脱塩素化酵素遺伝子の coverage の比較

	16S rRNA	<i>tceA</i>	<i>pceA</i>	<i>bvcA</i>	<i>vcrA</i>
Coverage	3141	2042	0	2432	3271

存在する 17 種類の還元脱塩素化酵素遺伝子の内、TCE 分解酵素遺伝子である *tceA*⁹⁾ を含む 7 種類の還元脱塩素化酵素遺伝子が存在することが分かった (表 2)。他のデハロココイデス属細菌の還元脱塩素化酵素とはほとんど高い相同性は示さなかったが、*D. mccartyi* BAV1 及び *D. mccartyi* VS の VC 酵素遺伝子である *bvcA*⁵⁾ 及び *vcrA*¹²⁾ とは高い相同性を有する遺伝子が含まれることが確認された。Real Time PCR の結果から、*bvcA* 及び *vcrA* を有するデハロココイデス属細菌が混合培養系の微生物の大部分を占めていることが明らかになっており、*bvcA* 及び *vcrA* の coverage の平均値が、存在が確認された *D. mccartyi* 195 の *tceA* とデハロココイデス属細菌の 16S rRNA 遺伝子の coverage の平均値と近いことから、この混合培養系を主に構成するデハロココイデス属細菌は *D. mccartyi* 195 株とゲノム配列が極めて高い相同性があり、*tceA* の他に *bvcA* や *vcrA* も有することが示唆された (表 3)。一方、PCE 分解酵素遺伝子である *pceA*¹⁰⁾ は存在しなかった。*tceA*、*bvcA* 及び *vcrA* については、PCR でクローニングを行い、シーケンスの結果、既知の遺伝子と高い相同性があることを確認した。

次に共存する微生物の解析を行った。SOLiD 3 の解析データは膨大なので、データベース上の多数の配列データとマッチングすることは困難である。そこで、データベース上の 16S rRNA 遺伝子を一定の長さで切断して連結することで構築した仮想的なゲノム DNA 配列を構築し、SOLiD のタグ配列をマッチングする方法を考案した (図 3)。それぞれの 16S rRNA 配列の coverage の平均値を計算することで、それぞれの存在の割合を推算することが可能である。今回は、coverage の平均値が 30 以上を示す 16S rRNA 断片を選出し、混合培養系内に存在する微生物の大まかな選定を行った。次に、選出された 16S rRNA 断片について、それらが由来する配列の全長について、再度マッチングを行い coverage の平均値を計算した。選出された配列の中には同じ微生物属由来

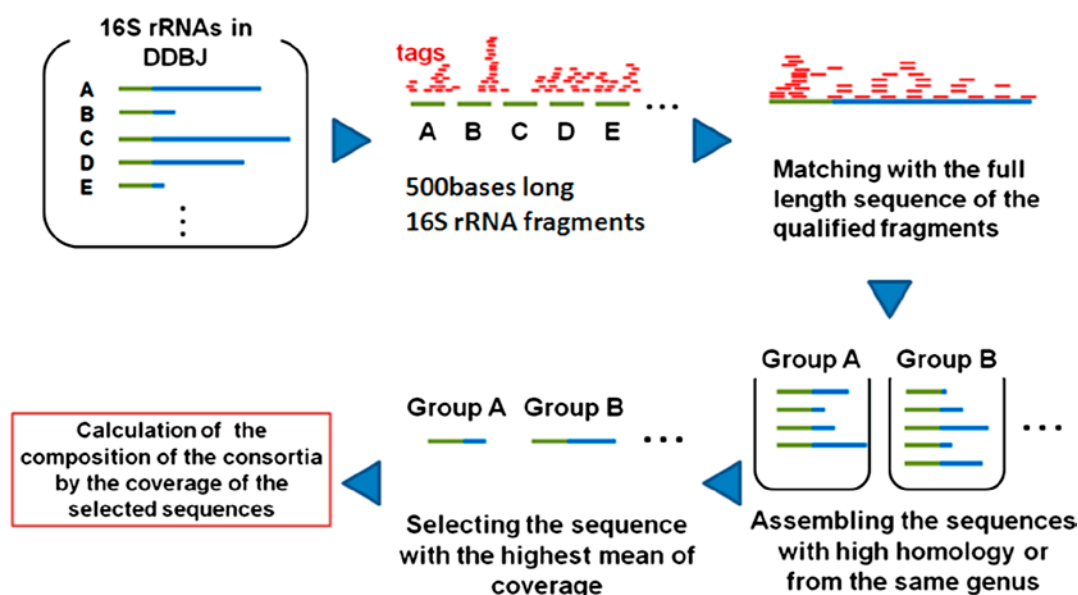


図 3. 第 3 世代 DNA シーケンサーのデータを用いた菌叢解析方法

表 4. 集積培養系に存在する微生物とその存在割合

微生物	Coverage の平均値	占有率 (%)
<i>Dehalococcoides</i>	3141	47.1
<i>Scarparcha</i>	1556	23.4
<i>Azospira</i>	1027	15.4
<i>Chorobi</i>	500	7.5

の配列が多数存在した。本研究では微生物の株まで解析する必要はないので、同じ微生物属に由来する配列をグループにまとめた。各グループから最も coverage の高かった配列を選出した。選出された各配列が、それぞれのグループを代表するものとした。各配列に対して配列の修正を行い、配列をよりテンプレートに即した配列に修正した。これらの coverage を基に、微生物の同定及びコンソーシア内での存在割合を計算した (表 4)。デハロココイデス属細菌の存在割合は 47.1% と計算された。培養系中の微生物群の主な構成微生物であることが、ゲノム解析データからも証明された。この方法で計算されたデハロココイデス属細菌の存在割合は Real Time PCR から推定された割合よりも低かったが、これは検量線の作成に用いた DNA と若干配列が異なるために、Real Time PCR の増幅効率が異なったことが原因であると考えられる。

5. 総括と今後の展望

TCE で汚染されている地下水から *cis*-1,2 DCE をエテンまで完全に分解する混合培養系の構築に成功した。この培養系から精製したメタゲノム DNA を SOLiD 3 で解析したところ、*D. mccartyi* 195 と高い相同性を有するデハロココイデス属細菌が存在していることが分かった。この混合培養系メタゲノムには TCE 分解酵素遺伝子である *tceA* 他に VC 分解酵素遺伝子である *bvcA* や *vcrA* が存在しており、同一の菌にこれらの遺伝子が存在する可能性が示唆された。この混合培養系が、*cis*-1,2 DCE を電子受容体として継体培養を行っていたにも関わらず、高い TCE 分解活性を有していたことから、この混合培養系に存在するデハロココイデス属細菌が TCE 分解酵素遺伝子と VC 分解酵素遺伝子を有し、TCE からエテンまでの完全分解が可能な新規なデハロココイデス属細菌である可能性があり、現在研究を進めている。デハロココイデス属細菌は各遺伝子の相同性が極めて高く、一部の遺伝子の相同性だけを解析してもその特性は明らかにできない可能性が大きい。次世代 DNA シークエンサーによる解析を行うことで、そのゲノム全体の情報から菌の特性を明らかにすることが可能であることが分かった。また、16S rRNA 遺伝子を対象とした解析により、集積培養系に共存する微生物の種類の特定とその存在量を推算することが可能であった。この解析方法は、データベース上の全ての遺伝子を対象とすることが可能であり、毒素の遺伝子などを対象とすることで、有毒微生物の有無を明らかにすることが可能である。この結果は、次世代 DNA シークエンサーによる解析は、集積培養系の安全性の評価などにも有効であると

考えられる。解析に要するコストの問題が指摘されるが、従来行われていた安全性試験試験などと比較すれば、低コストであり、短時間で明確な結果が得られることから、バイオレメディエーションの効率化と安全性評価に有効な手段であると考えられる。

6. 謝 辞

本研究の一部は沖縄先端ゲノムプロジェクトにより行われました。クロロエテン分解混合培養系の構築と SOLiD 3 での解析にご協力いただいた以下の皆様 (敬称略) に感謝いたします。矢木修身 (日本大学)、北嶋瑞樹 (東京農工大学)、岩本めぐみ、田村紀義 (PaGE Science)、臺場昭人 (Accelrys)、宮島隆 (TM Software)、塚原正俊、照屋盛実、喜久里育也、藤森一浩、今田有美、鼠尾まい子、佐藤友紀、矢野修一、三輪友希乃、町田雅之、平野隆 (沖縄先端ゲノムプロジェクト)

文 献

- Bentley, D.R., S. Balasubramanian, H.P. Swerdlow, G.P. Smith, J. Milton, et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456: 53–59.
- Cupples, A.M., A.M. Spormann, and P.L. McCarty. 2003. Growth of a *Dehalococcoides*-like microorganism on vinyl chloride and *cis*-dichloroethene as electron acceptors as determined by competitive PCR. *Applied and environmental microbiology* 69: 953–959.
- Damborsky, J. 1999. Tetrachloroethene-dehalogenating bacteria. *Folia microbiologica* 44: 247–262.
- He, J., K.M. Ritalahti, K.L. Yang, S.S. Koenigsberg, and F.E. Löffler. 2003. Detoxification of vinyl chloride to ethene coupled to growth of an anaerobic bacterium. *Nature* 424: 62–65.
- Krajmalnik-Brown, R., T. Holscher, I.N. Thomson, F.M. Saunders, K.M. Ritalahti, and F.E. Löffler. 2004. Genetic identification of a putative vinyl chloride reductase in *Dehalococcoides* sp. strain BAV1. *Applied and environmental microbiology* 70: 6347–6351.
- Kube, M., A. Beck, S.H. Zinder, H. Kuhl, R. Reinhardt, and L. Adrian. 2005. Genome sequence of the chlorinated compound-respiring bacterium *Dehalococcoides* species strain CBDB1. *Nature biotechnology* 23: 1269–1273.
- Lander, E.S., L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.
- Levy, S., G. Sutton, P.C. Ng, L. Feuk, A.L. Halpern, et al. 2007. The diploid genome sequence of an individual human. *PLoS biology* 5: e254.
- Magnuson, J.K., M.F. Romine, D.R. Burris, and M.T. Kingsley. 2000. Trichloroethene reductive dehalogenase from *Dehalococcoides ethenogenes*: sequence of *tceA* and substrate range characterization. *Applied and environmental microbiology* 66: 5141–5147.
- Manrao, E.A., I.M. Derrington, A.H. Laszlo, K.W. Langford, M.K. Hopper, N. Gillgren, M. Pavlenok, M. Niederweis, and J.H. Gundlach. 2012. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nature biotechnology* 30: 349–353.
- Maymo-Gatell, X., Y. Chien, J.M. Gossett, and S.H. Zinder. 1997. Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene. *Science* 276: 1568–1571.
- Muller, J.A., B.M. Rosner, G. Von Abendroth, G. Meshulam-Simon, P.L. McCarty, and A.M. Spormann. 2004. Molecular

- identification of the catabolic vinyl chloride reductase from *Dehalococcoides* sp. strain VS and its environmental distribution. *Applied and environmental microbiology* 70: 4880–4888.
- 13) Rothberg, J.M., W. Hinz, T.M. Rearick, J. Schultz, W. Mileski, et al. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475: 348–352.
 - 14) Seshadri, R., L. Adrian, D.E. Fouts, J.A. Eisen, A.M. Phillippy, et al. 2005. Genome sequence of the PCE-dechlorinating bacterium *Dehalococcoides ethenogenes*. *Science* 307: 105–108.
 - 15) Sung, Y., K.M. Ritalahti, R.P. Apkarian, and F.E. Löffler. 2006. Quantitative PCR confirms purity of strain GT, a novel trichloroethene-to-ethene-respiring *Dehalococcoides* isolate. *Applied and environmental microbiology* 72: 1980–1987.
 - 16) Suyama, A., M. Yamashita, S. Yoshino, and K. Furukawa. 2002. Molecular characterization of the PceA reductive dehalogenase of *Desulfotobacterium* sp. strain Y51. *Journal of bacteriology* 184: 3419–3425.
 - 17) Venter, J.C., M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, et al. 2001. The sequence of the human genome. *Science* 291: 1304–1351.
 - 18) Venter, J.C., K. Remington, J.F. Heidelberg, A.L. Halpern, D. Rusch, et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304: 66–74.
 - 19) Wheeler, D.A., M. Srinivasan, M. Egholm, Y. Shen, L. Chen, et al. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452: 872–876.