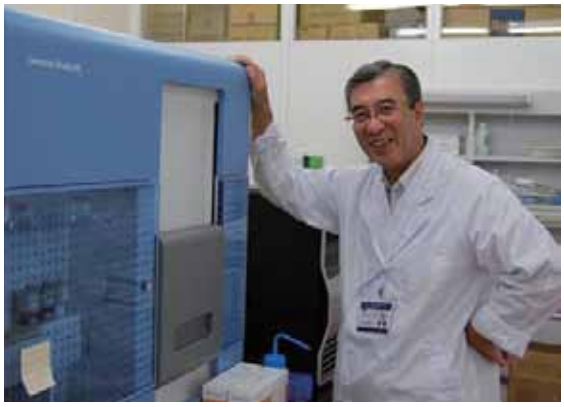
# 「次世代シーケンサー・ラボ訪問 vol.1」

東京農業大学：吉川博文先生

ゲノム研究に大きなパラダイムシフトをもたらした次世代シーケンサーが登場して3年がたちました。この間、この新しいテクノロジーを積極的に導入し、研究に取り組んでいる研究者の先生方も増えてきました。こうしたラボを訪問し、次世代シーケンサーを使った研究の醍醐味についてお話いただく「次世代シーケンサー・ラボ訪問」。初回はバクテリアゲノム解析を行っている東京農業大学の吉川博文先生にお話を伺いました。

## 枯草菌ゲノムのはじまり

私たちがNatureで、枯草菌のゲノム配列決定を発表したのが1997年でした。ゲノムを決めるのが始まったばかりの頃ですから、バクテリアで、いや生物種として、確か全ゲノム配列を決定した生物種の10番目だったと思います。このプロジェクトは研究者たちが実際に自分たちの手で配列を決めたゲノム、という実感がものすごくあります。研究室ごとに分担領域を決めて行ったので、そういう意味でもみんな愛着があるゲノム配列です。プロジェクト初期はやはり大変で、大学院生が丸々1年間ひたすらシーケンスするという時代が何年か続きました。そういうことに対する学生側の不満とか、そんなことで学位を与えていいのかとか、いろいろな周りからの批判のようなものにさらされながらやってきた背景があるので、すごく自分たちも懐かしいです。結局、ゲノム配列を決めるということの意義は、やはりやってみるとすごいなということが実感できる点にあると思います。こうして、ゲノムを決めるということの価値を、身をもって味わうことができたので、非常に貴重な経験でした。



吉川博文先生

東京農業大学応用生物学部　バイオサイエンス学科　微生物分子遺伝学研究室

枯草菌は自然界から分離した株で、初めて形質転換、すなわち遺伝的トランスフォーメーションという手法が使えるという、大腸菌と並んでまさに分子生物学の夜明けの原動力になった代表的な株です。そういった意味では非常に由緒がある生物種です。1997年のゲノム配列決定プロジェクトではオリジナルに近い株を使ったのですが、その後、この株は世界中に散らばり、いろいろなラボが分譲を受けて、少しずつ操作を加えているので、あちこち変わっています。現在米国オハイオのBacillus Genetic Stock Centerにも野生株があるのですが、これもどこかのラボから受けついた株がストックされています。こうして分譲された株は、みんな少しずつ違うだろうということは、なんとなくわかっていました。ラボでいじっているうちに、株の面（つら）が違ってくるんですね。どこか個別にシーケンスを決めても、どこそこ由来の株と、どこそこ由来の株は違うとかいうことは当たり前のようにあるので、全体だって違うだろうということは漠然と感じていたのです。

## 次世代シーケンサーの威力

次世代シーケンサーが登場して、ゲノムを再シーケンスすることが容易になりました。一度決まったものをもう一回配列を読み取るということの意義は、やはり変異の追跡だと思います。ゲノム配列決定で研究が終了ではなく、むしろそこから始まった部分がたくさんあるわけで、自分たちが興味のある株はどういう変異を本当はどうなっているのか、ということが再シーケンスで明らかになります。私たちの研究室でも、イルミナのGenome Analyzerを導入時にテストとして枯草菌の再シーケンスをしました。複数の研究室からサンプルを集めて再シーケンスをしてみたら、ゲノムサイズが400万塩基の枯草菌で、だいたい平均で40カ所ぐらいずつの違いがあることがわかりました。400万分の40ですから、印象としては非常に多いと思いました。ちなみに、この時に使ったリファレンスは1997年のNatureに発表したゲノム配列ではなく、数年前に次世代シーケンサー（454とGenome Analyzer）で読んだ論文がでたので、それと比較しています。というのは、1997年の初版のゲノム配列には、おそらく相当数のシーケンスの間違いも含まれていたと思います。次世代シーケンサーが登場して、簡単に読まれるようになって、どこらへんが怪しいとか、その領域を担当したのがどことか、そういったことまで明らかになってきてしまいます（笑）。日本のチームの成績はやはり優秀でしたけども。

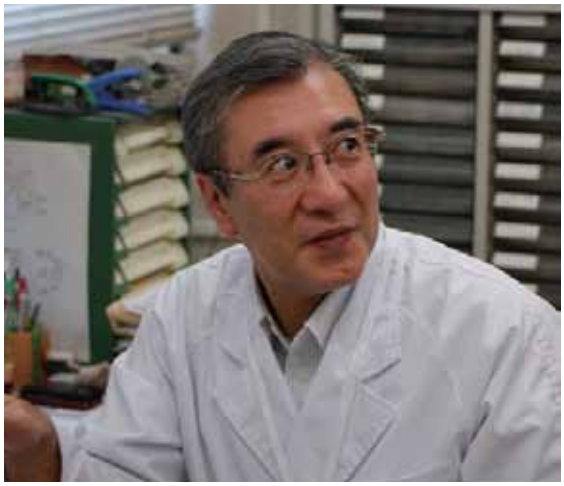
このプロジェクトでひとつ笑い話があります。ある研究室からもらった株を再シーケンスした時のことです。大腸菌などは、みんな－80℃に冷凍保存するのが一般的なのですが、枯草菌の場合は、胞子の状態だととても安定しているので室温に放置して、ひからびようが何しようが生きているので、すごく楽な生物です（笑）。ですので、フリーザー保存しているところもあれば室温のところもあるので、かなり保存状態もまちまちなのです。そういうのを集めて再シーケンスをしたのですが、

ある研究室オリジナルストックと言われてもらった株が、どうみてもオハイオのストックセンター株と非常に似ていました。全く一緒ではなく、明らかに違いもあるのですが、とても似ている配列でした。その先生に問い合わせたのです。「これは本当に先生の研究室から来たオリジナルの株ですか」。そしたら「そうだよ」と。そこでもう一度見たのですが、やっぱりストックセンターの配列にとても類似している。で、その結果を見せて先生に再度聞いたら、「そういえばあのとき、フリーザーが壊れてストックが死んでしまったとか、学生が言ってたな」と（笑）。さらに「うちの研究室の株の他に、別にストックセンターから取り寄せた株も、そういえばあったな…」。おそらく、フリーザーが壊れた時にサンプルの出し入れをして、サンプルが入れ違ったのだと思います。研究者では気づかないミスも、シーケンサーが教えてくれるという。こういう時代、指紋認証ではないですけど、シーケンサーによって株の素性なんかもわかってしまうという、すごい時代になったなと思います。

## バクテリアゲノム複製起点の決定

それ以外のプロジェクトはあと二つあります。ゲノム複製起点の決定と、変異解析です。ゲノム複製起点についてもう少し詳しく話すと、次世代シーケンサーは短鎖リードをリファレンスにマッピングするという方法です。ですので、DNAの量比が、マッピングするリード深度に反映すると考えられます。バクテリアは、十分な栄養環境下では、指数関数的に2、4、8、16、32と倍々に増えていく対数増殖期があります。細胞増殖する場合、ゲノムの複製がおこるのですが、その複製には起点があり、そこから終点までを増やしていきます。枯草菌や大腸菌では複製というのは40分かかります。しかし実際のところ細胞分裂は20分に１回行われています。これは、複製が終了して細胞分裂を起こす前に、すでに次の複製が始まっていることを示します。それぐらい盛んに分裂しているわけです。ですので、ゲノム複製起点周辺でのオーバーラップが発生し、複製起点、オリジンのところは、終点よりも、ゲノムの量としては多いはずなのです。

このことはすでに昔から遺伝的に確かめられているのですが、それが次世代シーケンサーでわかれば、複製の起点を検証することができます。こうしたサンプルをシーケンスしてみると、リファレンス配列の一番端を複製起点にしてゲノムを並べると、きれいにＶの字にマッピングされるはずです。実際、私たちでやってみたらきれいに予想どおりになりました。ただ、残念なことにV字パターンになった図を含む論文はすでに発表されていました（笑）。複製起点を見つけようとした論文ではないのですが、別のバクテリアを再シーケンスしたデータを確認すると、同じようになっている。だから、そのことは理論的には正しかったということが次世代シーケンサーでわかったわけです。



なぜこうしたことに興味を持っているかというと、オリジンのわからないゲノムがいっぱいあるからです。私たちが研究対象としているもうひとつの生物種、シアノバクテリアは、オリジンが予測出来ないということで逆に有名な株です。今回枯草菌で実証できたので、今度はシアノバクテリアでやってみようと思いました。ただ、これが大変でした。通常、枯草菌や大腸菌では、ゲノムは細胞内に1コピーしかありません。ところが、シアノバクテリアは細胞内に10コピーほどのゲノムが存在します。これらの複数ゲノムがそれぞれ独自に複製し、分裂するので、解析も難しいのではないかと。実際、枯草菌のようにシャープなVではなく、なだらかなVが見えてきました。まだ結論は出ていませんが。10コピーあるゲノムが、一斉に同時に複製されていないかもしれないということ、あるいはシアノバクテリアは倍加時間がとても遅いので、ゆっくりと休みながら複製をしているかもしれない。ちょっと次世代の話題と違ってしまうのですけど、次世代シーケンサーがきっかけで、こうした非常におもしろいことがわかってきました。

## 枯草菌の変異解析

もうひとつの変異解析は私たちの研究室のメインでもあります。私たちは遺伝学が専門なので、遺伝的に変異株を取ってきて、そのサプレッサーを取るということをよく行います。最近、ゲノムがわかっているので、何か機能を調べたいというときに、興味ある遺伝子に変異をかけることが比較的容易にできます。バクテリアには細胞の膜と壁が周りにあるのですが、分裂時には膜と壁がきちんと切断され、再合成されるわけです。これができないと分裂できません。ということで、私たちは脂質の合成系と細胞分裂との関係に興味をもっているのです。実験系としては、脂質合成の必須遺伝子に変異をいれた、温度感受性の変異株をまず作りました。温度感受性があるので、この株は高温にすると増殖できません。ですが、そこで無理やり選択すると増えるものがでてきます。そういったものを探し、株のゲノムの違いを探していくのです。実験自体は簡単なのですが、そこから先のマッピング、どこに変異が落ちているかというのを探すのは、たかだかバクテリアといえども大変な作業です。

バクテリアゲノムに対して次世代シーケンサーを使う意義というものを、私たちはこういうことをやっていてよく感じます。変異解析は前々からしたかったことなのですが、ここで重要なのは、最初に行った野生株の配列を再度確認するという作業です。これがないと、変異株シーケンスを行ったときに、どれが新しい有効変異で表現型をもたらしている変異なのかを識別できません。再シーケンスプロジェクトで400万塩基中、SNPとIndelで40か所の変異が見つかったといいましたが、40個もあるとこれまでの手法だと、バクテリアでも検証するのは相当大変です。ですので、直前の親株というものを次世代でシーケンスしたことがすごく効果的で、そこの親株にあった変異というのを、リファレンスとの違いで見つかってきた変異から差し引

くことで、きれいに2つぐらいまで絞ることができました。そこからは通常の検証にはいるわけです。検証ステップではこの変異を人工的に作り、親株にいれます。そうするとサプレスしたという表現型が見えたので、証明できたことになります。大事な話なので繰り返しますが、今、言ったような作業をしないと、バクテリアでは論文になりません。必ず、検証が必要です。これは今日一番言いたいことの一つです。

## 次世代シーケンサーの運用

次世代シーケンサーの運用ですが、プロジェクトは学内公募をしています。メール配信して集まったものをランしています。ながれとしてはゲノム調製からサンプル調製まではみなさんに教えて各ラボで行ってもらっています。きれいなゲノムを抽出した方がいいので、ゲノムがとれたら各ラボで測定してもらい、よさそうならサンプル調製というステップです。その後、Genome Analyzerのランは、こちらのスタッフがやっています。

1回のランで50〜60サンプルながしていますが、次世代シーケンサーには7名の専任スタッフがいます。ほとんどウェットで、ウェットとドライ兼任が1名、ドライの専任が1名です。Genome Analyzerのパイプラインから出てくるデータの一次処理はほとんどみんなハンドルできます。その先は、ウェットとドライ兼任の担当者が行っていますが、今回お話しした内容もすべて彼がプログラムを組んでやってくれています。彼はエキスパートになりつつあり、まさに神様の存在です（笑）。彼も言っていたけれども、ヒトとマウスとラットは、ソフトやツールはかなり充実しているようです。彼のような人材が一人いれば、かなりいろいろなことができるのですが、一人だと、彼が風邪などでダウンするととても困ります（笑）。次世代シーケンサーで可能性はいろいろ広がって、いろいろなことができるようになったけれども、やはりデータ解析は圧倒的に必要です。おそらく、第2世代でも第3世代でも、大量のシーケンスデータがでてくるので、それを取り扱いができる人はどこへ行ってもきっと重宝されるから、大丈夫だろうと思います。

## 次世代シーケンサーのノウハウ

次世代シーケンサーで僕らが最初教わりに行ったのが東大新領域です。東大ではほとんどヒトの仕事をされていると思いますが、生物種を超えて、いろいろと具体的なノウハウを聞くことができ、とても参考になっています。私たちのところでも実験系が徐々に立ち上がっているので、知り合いから少しずつ相談も来るようになりました。イルミナに関して言えば、すごい勢いでいろいろ論文も出ているし、いくつかフリーウェアのようなものも出ています。それと、もちろんイルミナのバージョンアップもあるので、非常に使いやすくなってきていますね。また、イルミナのユーザーミーティングのようなところへ行くと、いろいろな研究者にあえて、おもしろいかなと思っています。

## 今後の展望

次世代は低コストの汎用機がでてきていますが、テクノロジーの広がりを実感しています。ただ、私たちのように一研究施設で大型システムを導入した場合は、そのプロジェクトが終わったあとどうするかも重要です。やはり遊ばせたくないので、それにはランニングコストをどうするのか。別のプロジェクトを立ち上げて取らないといけないとは思っています。予想どおり

うまくいくかどうかわからない。そのために、この成果を上げておく必要があるのです。

イルミナに関して言えば、バクテリア屋さんは1本にしたいという願望が強いので、メイトペアを強化してほしいと思います。確かに、今は99%のデータ量さえあれば、ほとんどのことはわかるし、探したいものは探せるので、それでいいのです。でも、1本にならないと気持ち悪いという人がやはり多いのです（笑）。あと、PCRフリーがもうすぐ出れば、だいぶ変わるかもしれません。ゲノムにはATが20数パーセントというものもありますので。そうしたゲノムの*de novo*配列決定をやりたいという人は、やはり要求としては多くなりつつあるかと思います。

## 編集後記

熱意をもってバクテリアゲノムを研究されているご様子がインタビューから伺えました。詳細までわかりやすくご説明いただいたので、研究の内容も理解しやすく、非常に面白いトピックのものでした。またバクテリアゲノムの研究コミュニティ、そして学内の公募プロジェクトの運営についても話してくださり、東京農大での次世代シーケンサーの活用の様子をうかがうことができました。

January 2011


## イルミナ株式会社

〒108-0014
東京都港区芝5-36-7　三田ベルジュビル22階
Tel (03)4578-2800　Fax (03)4578-2810
www.illuminakk.co.jp


代理店

本製品の使用目的は研究に限定されます。


© 2013 Illumina, Inc. All rights reserved.

Illumina, illumina*Dx*, BaseSpace, BeadArray, BeadXpress, cBot, CSPro, DASL, DesignStudio, Eco, GAIIx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, NuPCR, SeqMonitor, Solexa, TruSeq, TruSight, VeraCode, the pumpkin orange color, the Genetic Energy streaming bases design は Illumina, Inc の商標または登録商標です。
その他の会社名や商品名は、各社の商標または登録商標です。予告なしに仕様を変更する場合があります。


Pub. No. 070-2010-J019 01JAN11

illumina®