

II. 塩基配列解読の基本方針

イネゲノム塩基配列の解読に関して、基本方針は階層化ショットガン法によって全染色体を解読し、シーケンシングや編集作業も含めて高精度な解読に努め、未解読部分を限りなく少なくし、完全解読に近づけるということであった。イネ品種「日本晴」のゲノムが持つ塩基配列を可能な限り忠実に解読しようとするためのものである。

1. 階層化ショットガン法

ゲノムの塩基配列を解読する方法は、図 5-(a)に示すように、ホールゲノムショットガン法と階層化ショットガン法とがある。ホールゲノムショットガン法は1つの生物のゲノム DNA を酵素等で裁断し、全体のゲノムサイズの6～10倍のDNA量に相当する断片について塩基配列を解析し、重複部分を目安に高性能コンピュータを用いてつなぎ合わせる方法であり、ゲノムサイズの小さい生物の全塩基配列を迅速に解析するのに適する方法である。

一方、階層化ショットガン法は、段階を踏んで染色体の位置情報も含めて、核ゲノムの塩基配列を完全に解読する方法である。まず始めにゲノム DNA を制限酵素によりある程度長い断片に切断し、PAC あるいは BAC などに組み込み、ライブラリーを作製し、遺伝地図上のマーカーを使って断片をつなぎ合わせ、元の染色体を構築する、いわゆる物理地図を作製する。既知となったこれら PAC/BAC を裁断し、2kb から 5kb ほどの小さなサブクローンを作製し、このサブクローンごとに両端約 500 塩基を解読する。これらをつなぎ合わせ、PAC/BAC の塩基配列を決定する。次に物理地図情報をもとにして PAC/BAC をつなぎ合わせ、染色体の塩基配列を再現する。この方法はゲノムサイズの大きな生物の全塩基配列の解読に適し、より精度の高い解読が可能となる。

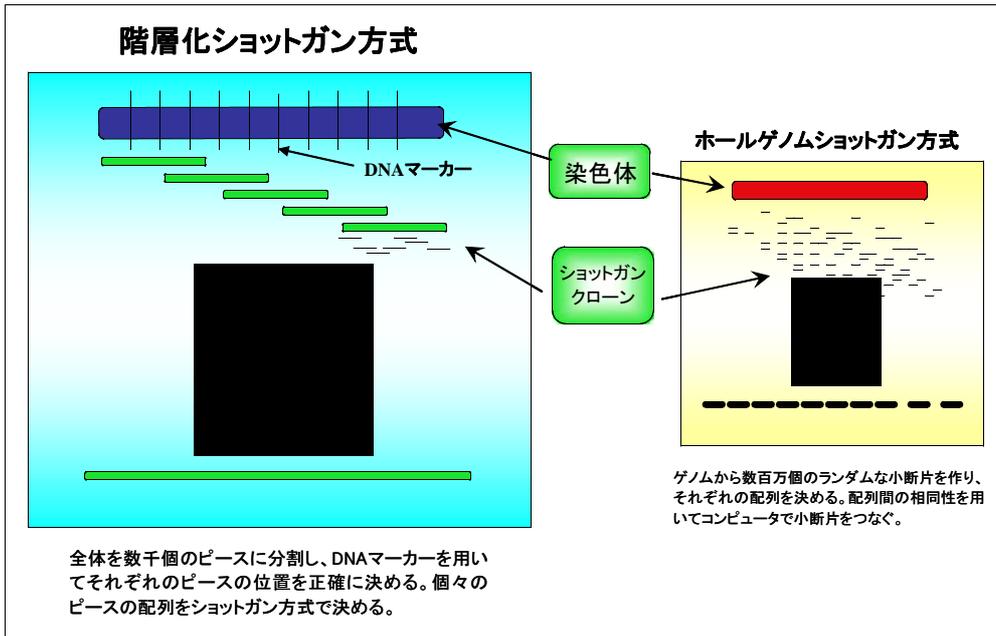
図 5-(b)は、RGP において、遺伝地図、発現遺伝子地図、YAC 物理地図を使用して PAC/BAC を整列化し、高精度解読を行った経過について、使用した主要器機類の写真とともに示したものである。

2. 高精度解読

解読されたゲノム塩基配列は、最終的には公的データベースに登録され、世界各国の研究員の研究開発のために使われるので、実際に利用する場合を考えると、正確な解読が必要である。このためには、いかに精度の高い機器類を使い、高品質の試薬を用いるか、各 PAC/BAC クローンについていかに多くのサブクローンを解析し、冗長度を高めるか、連続する断片の重なり部分をいかに厳しく判定し誤りを少なくするか、さらにはギャップといわれる未解読部分をいかに少なくするか、などを解決し、精度の高い解読結果を得なければならない。

このプロジェクトでは精度 99.99%を基準として解析を進めた。すなわちこの基準は、間違いが 10,000 塩基に 1 塩基以下であることを意味する。遺伝子は数百～数万の塩

(a)



(b)

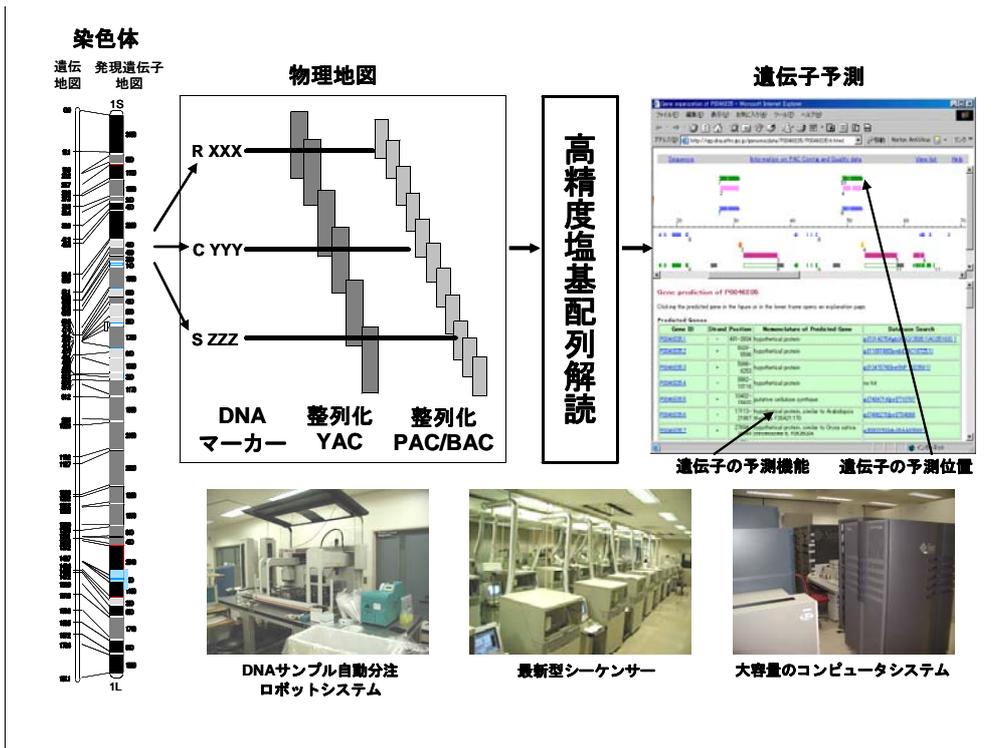


図5. 階層化ショットガン法による解読の経過

(a) 階層化ショットガン法とホールゲノムショットガン法との比較

階層化ショットガン法では、遺伝地図、発現遺伝子地図、物理地図を事前あるいは並行的に作製し、これらを参考にして塩基配列を精度高く解読する。なお、参考に示したゲノム塩基配列のもう1つの解読方法であるホールゲノムショットガン法では、地図は作製せず大量の小断片をコンピュータでつなぎ合わせる方法であり、完全なゲノム全体のつながりや高い精度を得ることは難しい。

(b) 階層化ショットガン法による解読

階層化ショットガン法では、遺伝地図、発現遺伝子地図、YAC物理地図、PAC/BAC物理地図を使って、PAC/BACを選抜し、塩基配列を解読していく。

基配列を持つため、この精度においては、1個の遺伝子の中にある間違いは1塩基以下である。

また、解析に使用する機器類や薬品類も解読精度に大きく反映する。このプロジェクトでは、シーケンサー、DNA抽出、分析用ロボットなど主要な機器類は最も精度の高い最新鋭のものを使い、使用試薬類も解読精度に重点を置いて選定した。

さらに、1つの塩基を何回重複して読むかも精度には大きく影響する。このプロジェクトでは1つの塩基を平均10回読むことになる10倍の冗長性で読むことにより、精度99.99%の基準達成を目指した。

3. 完全解読

ゲノム全塩基配列を解読するに当たっては、未解読部分、低精度部分を全くなくすることは難しく、これらを最小限にすることを目指して、PAC/BAC物理地図の充実および個々のPAC/BACのギャップフィリングおよび低精度部分の解読精度向上を進めた。

1) PAC/BAC物理地図の精度向上

PAC/BAC物理地図は、解読に当たって分析試料を作製する基本となるものであり、図6に示すように、解読精度を高めるためには常に充実を図っていかねばならない。

遺伝地図および発現遺伝子地図上のマーカーを用いてPAC/BACを選抜し、物理地図を作製していくが、その領域のPAC/BACがライブラリーには含まれておらず、ギャップになる場合やマーカーの存在しない領域でPAC/BACの選抜ができずギャップになってしまう場合などがある。これらのギャップを埋め、全体的により正確な塩基配列解読を目指すことも完全解読に

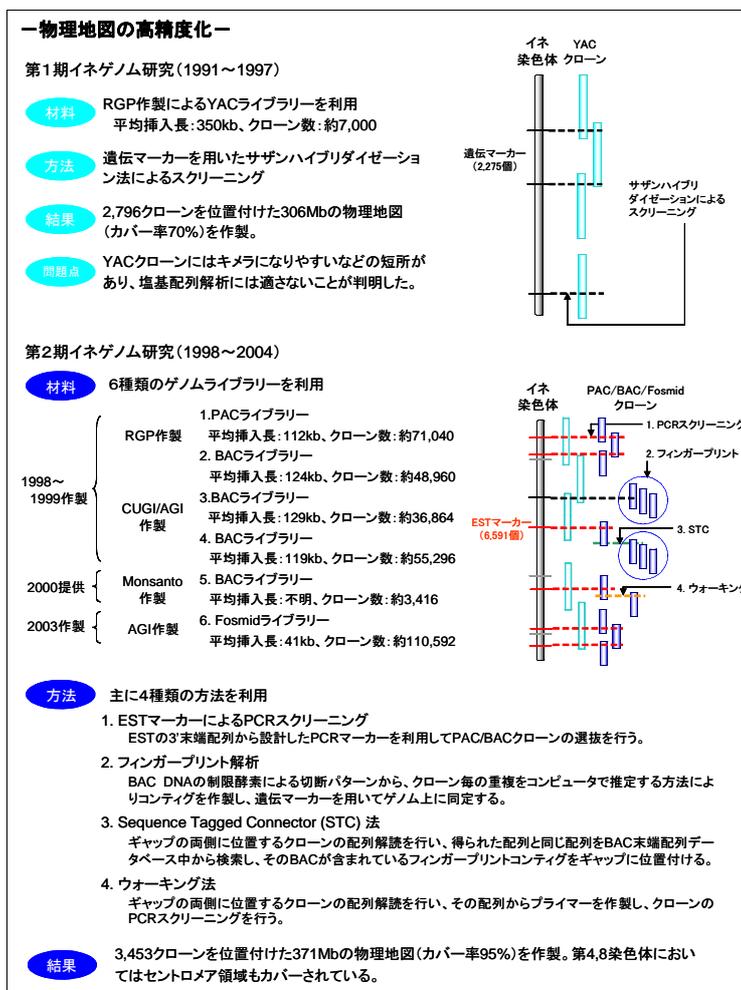


図6. 物理地図の高精度化

PAC/BAC物理地図は作製開始以降、新たなライブラリーを加えるとともに解読手法を開発、駆使し、高精度化を図ってきた。

は欠かせないことである。

そこでシーケンスタグコネクタ（STC）法やウォーキング法と呼ばれる方法によりギャップを埋める。STC 法では、ギャップの両側に位置するクローンの配列解読を行い、得られた配列と同じ配列を BAC 末端配列データベースの中から検索する。そしてその BAC が含まれているフィンガープリントコンティグをギャップに位置づけ、ギャップを埋める。ウォーキング法では、両側クローンの配列からプライマーを作製し、PAC/BAC クローンのスクリーニングを行い、ギャップを埋める。

さらに、単一の DNA ライブラリーの STC 法やウォーキング法だけではギャップフィリングには限界がある。そこでさらにギャップフィリングを進めていくには、単に 1 種類のライブラリーから解析に用いる DNA クローンをサンプルするだけでなく、多くのライブラリーからサンプルし、ギャップフィリングをすることが必要となる。最初 RGP では自ら作製した PAC/BAC のみを用いていたが、その後モンサント社から提供された BAC を加え、概要解読終了後はアリゾナ大学が作製したフォスミドライブラリーなどを用いて、ギャップフィリングを行った。

2) PAC/BACクローンのギャップフィリングと解読精度向上

PAC/BAC クローン内のギャップフィリングと配列精度向上は、図 1 で示したフェーズ 1 からフェーズ 3 に段階を上げていくことにより達成される。PAC/BAC の個々について、〔サブライブラリー作製〕－〔シーケンシング〕－〔編集〕の一連の作業を一回実施しただけの場合は、ほとんどのクローンがフェーズ 1 かフェーズ 2 の段階でとどまり、ギャップや解読精度の悪い部分が残る。そこでフェーズを上げるために、再度 PAC/BAC からサブクローンを作り、この中からギャップを埋めるブリッジクローンを探しギャップを埋める。また、解読精度の低かった領域のサブクローンについても再度シーケンシングを行い、解読精度を高める。

解読精度が低い領域は、反復配列がある場合が多く、普通のサブクローンの作製方法ではサブクローンが取れない場合が多い。そこで、このような場合にはトランスポゾンを利用したサブクローンの作製を行い、均等にサブクローンが取れるよう工夫した。

3) セントロメア、テロメアなど解読困難領域の解読

セントロメアは各染色体の中間部分にあつて、細胞分裂の際、新たにできる細胞に染色体を正確に分配する役目を持っている。また、テロメアは、染色体の両末端部分にあつて染色体の複製や安定化に必要な特殊な構造を持っている。両者とも極めて長い範囲にわたり反復配列が含まれており、そのために解読は非常に難しいと言われている。

また、機能は不明であるが、セントロメアやテロメア部分以外にも染色体の多くの部分で反復配列が存在する。反復配列は、サブクローニングしたとき前後の区別や反復回数を正確に判定することが極めて難しい。したがって、反復配列を正確に解読するためには、特殊な分析技術や試薬の開発が必要となる場合もある。セントロメアやテロメアは反復配

列部分が著しく長いことから、塩基配列の解読にはそれ相応の技術開発が必要となる。