

Ⅲ. イネゲノム塩基配列完全解読の成果

2004 年末をもってイネゲノム全塩基配列の完全解読が完了した。ここに完成した塩基配列情報は、今後広く遺伝学ならびに植物の品種開発に大きく貢献していくことが予測される。さらに、このプロジェクトの成果は単に塩基配列情報だけでなく、解読を進めるに当たって構築した基盤情報である遺伝地図、発現遺伝子地図、物理地図ならびに cDNA ライブラリーなども今後広範囲の利用が予測される重要な成果である。

ここでは、このプロジェクトが達成した成果を、単に塩基配列情報だけでなく基盤的地図情報についても解説するとともに、地図情報と対になって利用される DNA ライブラリーについても紹介することとする。

1. 遺伝地図

遺伝地図は遺伝子や DNA マーカーの染色体上の位置を示したものであり、ゲノム解析研究を進める上では極めて重要な基本情報である。染色体 DNA を制限酵素で切り出した場合、DNA 断片の長さが品種間で異なってくることがある。このような違いは多型と呼ばれ、二者を区別することができるため、遺伝子の違いとして認識できる。これを用いて連鎖分析をすることによりその多型が染色体のどの部分に存在するのかを決めることがで

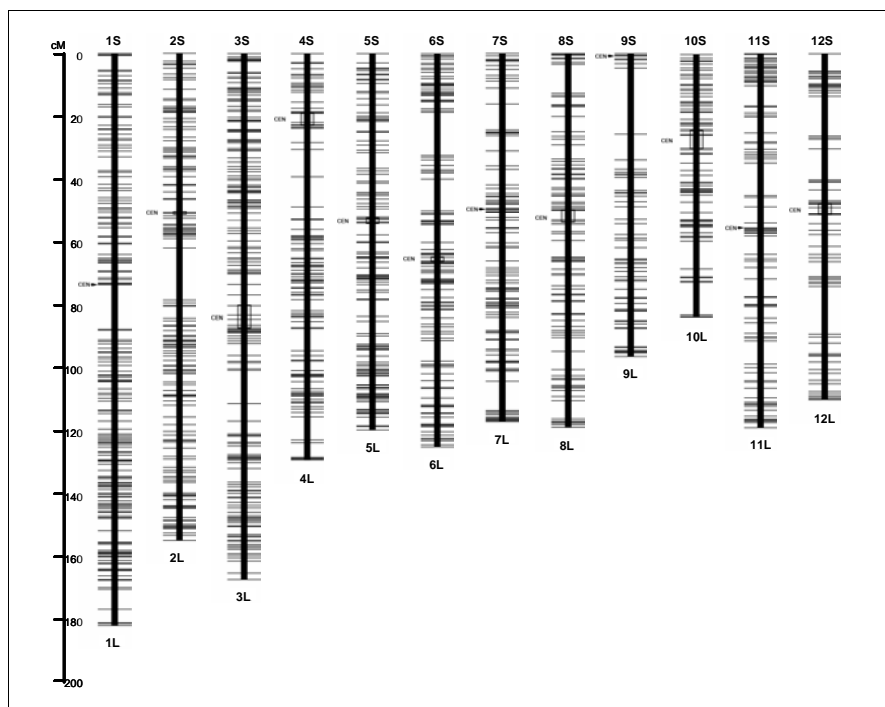


図7. 遺伝地図

遺伝地図は、染色体に目印を付け DNA 断片や遺伝子が染色体上のどの位置にあるかを知るための指標となる地図である。この図に示した遺伝地図は 1991 年から 1997 年の 7 年間をかけて 2,275 個の目印を 12 本の染色体に付けたものである。さらに、2000 年には約 1,000 個のマーカーが追加され、3,267 個としてホームページ上に公開した。

きる。この多型は制限酵素断片長多型（RFLP）と呼ばれ、最初の分子遺伝地図を作製するマーカーとして用いられた。

イネゲノム解析研究では、日本型品種「日本晴」とインド型品種「Kasalath」との交配 F₂ 集団 186 個体を用いて RFLP を検出し連鎖解析を行い、1998 年までに、図 7 に示すような総数 2,275 個の RFLP マーカーを染色体上に位置づけた高密度連鎖地図を完成させた。その後徐々にマーカーを増やし、現在では 3,267 個のマーカーによる地図が完成し、RGP のホームページ (<http://rgp.dna.affrc.go.jp/>) から公開している。

この地図は、物理地図を完成させるのに必須の道具であり、その後続く階層化ショットガン法による完全解読の基礎となる成果で、かつ遺伝子の単離や機能解明などに有効なばかりでなく、重要形質と密接に関連する DNA マーカーによる間接選抜など実用技術にも利用でき、イネゲノム研究最初の大きな成果であった。

2. 発現遺伝子地図

イネゲノムの塩基配列の解読や遺伝子単離のためには、先に述べた遺伝地図の遺伝マーカーだけでは不十分であり、さらに多くの遺伝マーカーを配置した高密度な遺伝地図が必要となる。異なる条件下で発現している mRNA から cDNA を作製し、これらの部分塩基配列を調べ染色体上に位置付ければ密度の高い遺伝地図となる。これらは発現遺伝子地図と呼ばれ、イネゲノム解析研究では、種々の組織、生育ステージ、各種ストレス条件下で

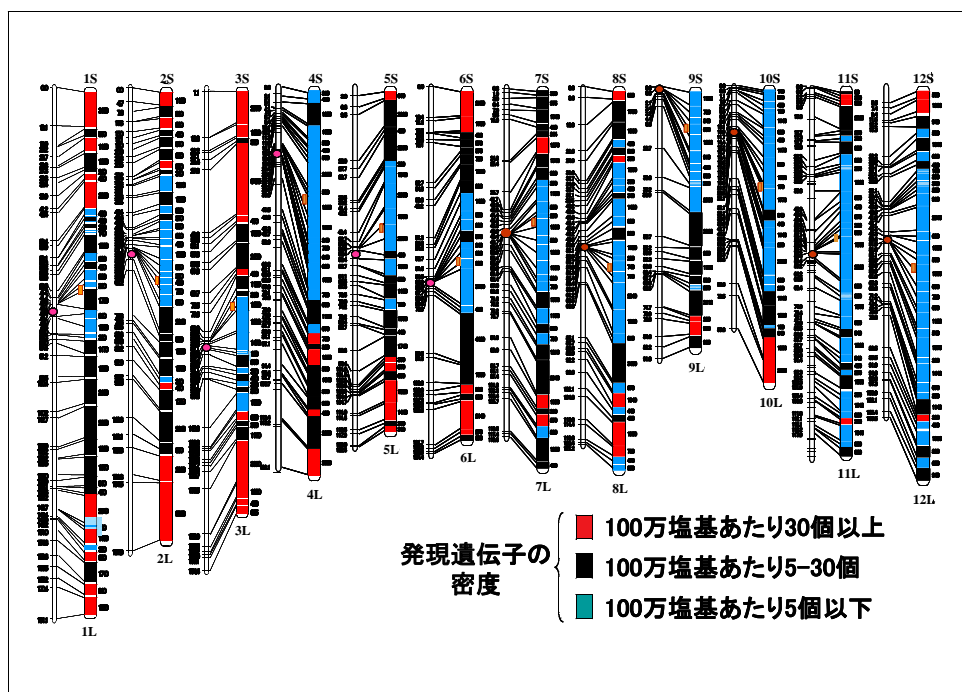


図8. 発現遺伝子地図

イネで実際に機能している遺伝子を染色体上に位置づけた地図であり、塩基配列の解読、遺伝子を単離する時や近縁の作物の遺伝マーカーを探索する時に利用できる。この発現遺伝子地図は第1期イネゲノム解析研究のテーマの1つであった cDNA カタログ作成の中で構築した約4万個の発現遺伝子の中で重複のない約 6,591 個を位置づけたものである。

発現している mRNA を集め、cDNA を作製し、その末端配列を利用してプライマーを合成し、YAC ライブラリーに対して PCR (ポリメラーゼチェーンリアクション) スクリーニングを行い、YAC 物理地図を参照して、**図 8** に示すような 6,591 個の発現遺伝子マーカーをイネゲノム全体にわたって配置した発現遺伝子地図を完成させた。

このようにして作製した高密度な発現遺伝子地図によって、ゲノム全体における遺伝子の分布状況が明らかになり、いずれの染色体ともセントロメアに近い部分の遺伝子密度は両腕末端部に比べて低い傾向があり、また遺伝子密度は染色体によって異なっていることも明らかになった。

この情報はイネの全塩基配列を解明する上で分析材料となる PAC/BAC を選抜するのに大きな役割を果たしたばかりでなく、遺伝子単離や機能解明に役立っている。

3. 物理地図

物理地図は、制限酵素で切断した DNA 断片を、マーカーを指標にしてそれが染色体上に並んでいたと同じ順序に並べ直し、断片番号などで表示したものであり、これらの DNA 断片がゲノム塩基配列や遺伝子単離の試料になるという役割を持っている。

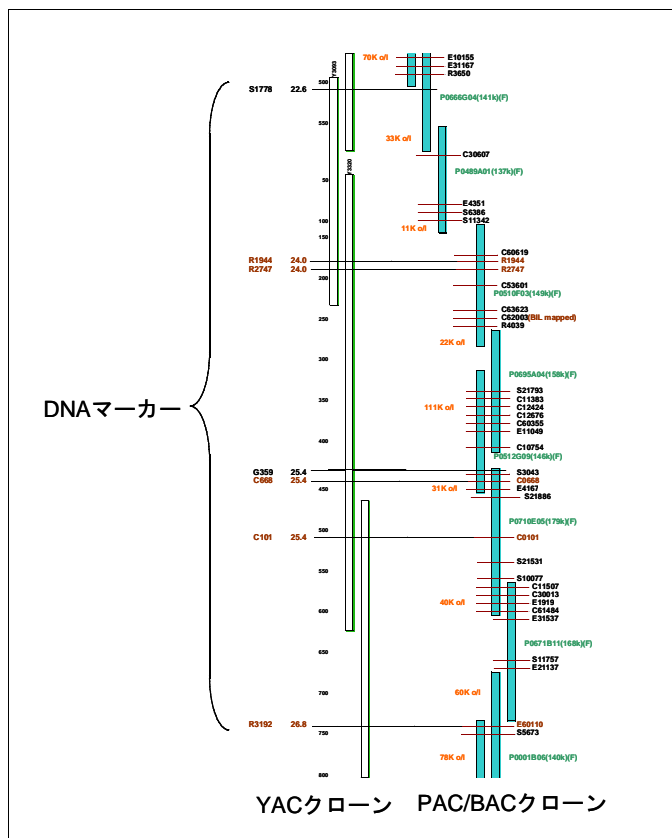


図9. 物理地図

物理地図は DNA の断片を DNA マーカーを目印に並べた地図であり、塩基配列の解読や遺伝子の単離はこの地図を指標として行う。この図は第1染色体の YAC と PAC/BAC 物理地図の一部分を示したものであり、このような地図が 12 本の染色体全体にわたって完成している。

イネゲノム解析研究では、まず 350kb 以上の長い DNA 断片を組み込むことができる YAC ライブラリーの作製から始め、イネゲノムの約 70% をカバーする YAC 物理地図を作製した。しかし、YAC クロームは分析試料として使うには長過ぎるため、引き続き PAC と BAC ライブラリーを作製し、YAC 物理地図の基盤の上に発現遺伝子マーカーを指標にして PAC/BAC 物理地図を作製した。また、クレムゾン大学とモンサント社から BAC クロームを導入し、物理地図の精密化をはかった。この地図は 3,000 を越える PAC/BAC クロームで構成されており、すべてをここに示すことはできないが、参考までに**図 9** に第 1 染色体の一部分を示した。物理地図上の各 PAC/BAC クロームについては順次配列解析を行い、最終的に全

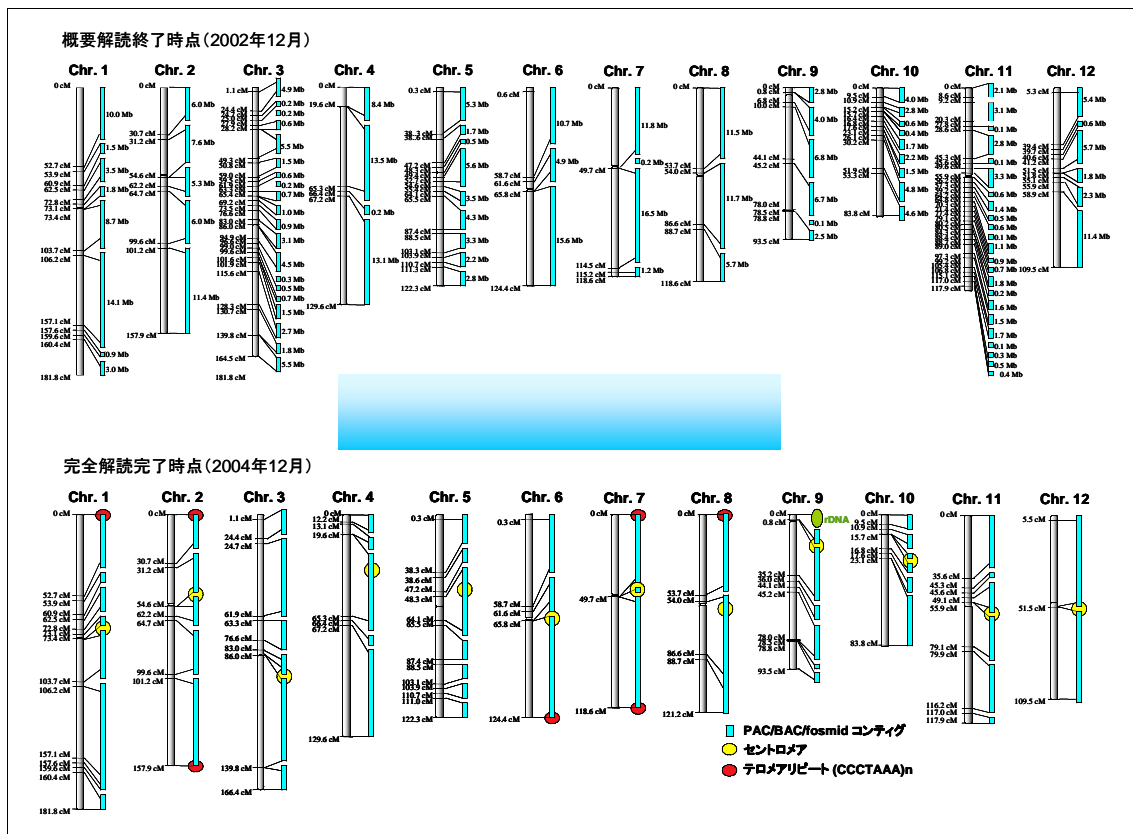


図10. 概要解読終了時と完全解読完了時の物理地図ギャップの比較

物理地図作製に当たっては、最初は DNA クローンの見つからない部分があり、多数のギャップが残ってしまうが、いくつかの方法によりギャップを埋めていく。概要解読終了時点では 88 個のギャップがあったが、完全解読完了時点ではこれを 45 個にまで減らすことができた。また、第 4、8 染色体以外のセントロメア領域はギャップとなっている。

ゲノムの塩基配列を解読した。なお、物理地図はゲノム全体を完全にカバーするのは極めて困難なことであり、DNA 断片で埋めきれない部分が残ってしまう。概要解読終了時点と完全解読完了時点で物理地図の構築の程度を比較したものが図 10 である。概要解読終了時点では、全体で 88 個であったギャップが完全解読完了時点では 45 個に減少し、未カバー領域も 34Mb から 29Mb に減少している。

4. ゲノム塩基配列情報

イネゲノムの全塩基配列解読は非常に大きなプロジェクトであり、解読を達成するために必要な多くの基盤研究を行い、上に述べたような成果を上げたのであるが、このプロジェクトの最終目的はイネゲノム全塩基配列の完全解読であり、解読された配列情報が最も重要な成果である。2年前の概要解読終了時とこの度の完全解読完了時のイネゲノムの塩基配列解読結果を比較すると、表 1 に示すようになる。

イネゲノムの推定全塩基配列数は約 390Mb で、解読前に推定されていた約 430Mb、概要解読終了時に推定されていた約 400Mb よりも少なくなった。

表1. イネゲノム塩基配列の解読結果

概要解読終了時と完全解読完了時で大きく変わったのは物理地図のギャップ数であり、塩基数やカバー率などに大きな変化はない。しかし、PAC/BAC間のギャップを減少させ、全PAC/BACをフェーズ3にした完全解読結果は今後の利用にとっては、概要解読の配列情報に較べると格段に利用価値の高いものとなっている。

	概要解読終了時 (2002.12)	完全解読完了時 (2004.12)
決定塩基数	約 366 Mb	約 371 Mb
推定全塩基数	約 400 Mb	約 390 Mb
カバー率 (決定塩基数/推定全塩基数)	約 92 %	約 95 %
物理地図ギャップ数	88 個	45 個
未カバー領域	約 34 Mb	約 19 Mb
うちテロメア・セントロメア領域	約 10 Mb	約 8 Mb
うちギャップ領域	約 24 Mb	約 11 Mb
データ精度	99.99%	99.99%

2002 年末の概要解読終了時点では、解読精度 99.99%で 366Mb まで解読できていた。未解読で残っていたのはセントロメアとテロメアに相当する部分 10Mb、反復配列やループ状構造などがあるため解読が困難でギャップとなった部分 24Mb の合計 34Mb であった。

完全解読完了時点では、概要解読終了時点に比べて、解読に用いた人工染色体のすべてで未解読部分を少なくしており、また PAC/BAC 間の未結合部分も少なくしている。このため 12 本の各染色体について、シュードモレキュールといわれる一本につながった塩基配列を正確に作製することができ、これにより塩基配列からの遺伝子予測の精度が著しく高まり、遺伝子単離実験がより精度高くできるようになるなど、配列情報の信頼性と利用価値が著しく高まった。

セントロメアは、細胞分裂時に染色体を両極に分離させていくために紡錘糸が付着するなど重要な機能を持つ染色体領域とされるが、機能の詳細についてはまだ明らかにされていない。機能を明らかにするにはまずセントロメアの塩基配列を解読することが必須であるが、セントロメアには長大な反復配列が含まれており、その解読は極めて難しい。セントロメアの塩基配列を完全に解読した例はこれまでになく、その機能の解明も十分にはなされていなかった。

イネにおいてもセントロメアの解読は難しく、ほとんどの染色体でギャップとして残っていたが、第 8 染色体に関しては完全解読に成功した。そのセントロメアの詳細構造を図 11 に示した。従来、セントロメアには遺伝子は含まれていないというのが定説であったが、この図に示すようにいくつかの遺伝子が含まれていることを明らかにした。今後さらに研究を進めることにより、他の染色体におけるセントロメアの構造、機能および進化の過程が明らかになっていくであろう。

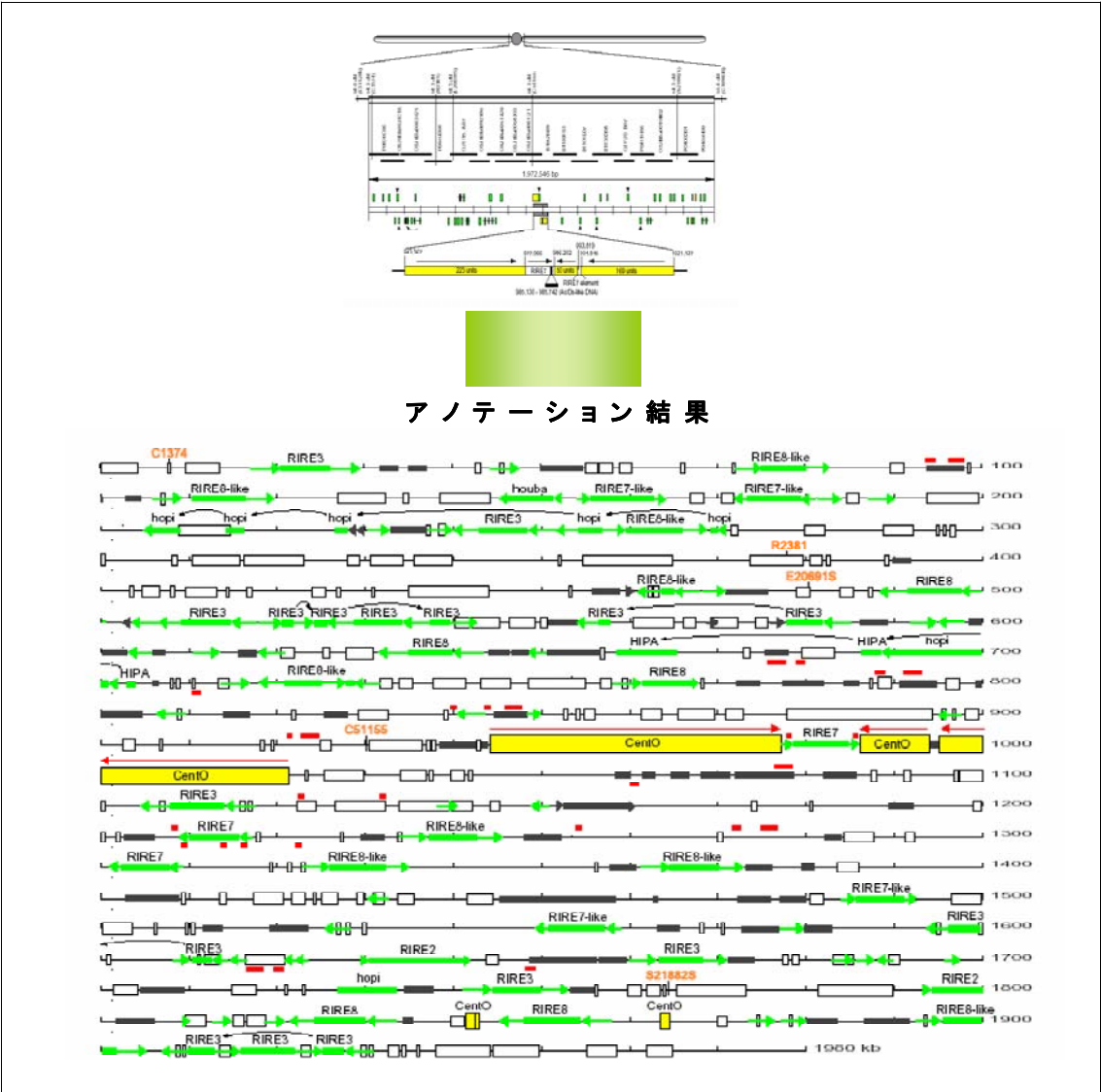


図11. 第8染色体セントロメアの構造

この領域には CentO と呼ばれるサテライト配列が3つのクラスターになって存在すると共に、220 個以上に及ぶトランスポゾン様配列が推定された。タンパク質遺伝子も発見され、そのうちの 48 個については既知のタンパク質や cDNA 配列との相同性が見られた。

5. 塩基配列情報からの遺伝子予測

ゲノムの塩基配列が解読されると、その配列の中にどのような遺伝子が存在するかを予測することができる。遺伝子の始まりと終わりの塩基配列には一定の法則があり、始まりは ATG で、終わりは TAA、TAG および TGA のいずれかである。そこで、解読されたゲノム配列の中からこれらの配列を見つけ出すことにより、一個の遺伝子の開始地点と終了地点とを知ることができ、遺伝子の配列と存在位置がわかる。

さらに、遺伝子領域の塩基配列の中にはタンパク質に翻訳されるエクソンと翻訳されな

いイントロンが含まれているが、イントロンの始点は GT であり、その終止点は AG であることからイントロンとエキソンの境界点を探し出し、遺伝子の存在箇所を確認するとともに、遺伝子の産物であるタンパク質の機能を推定することも可能になる。

予測された遺伝子からタンパク質の機能を推定するには、類似性検索プログラムを使って、公的データベースに登録されたあらゆる実在および予測遺伝子とのアミノ酸配列の類似性を調べ、この比較結果から機能を推定する。

アノテーションについては、RGP ではフェーズ 3 になった PAC/BAC から順次開始することになっているので、完全解読完了時点で日本担当分の 6 本の染色体については、ほぼ完了している。RGP ではこのようにして得られたアノテーション結果を表示するシステム INE (INtegrated rice genome Explorer) を開発し、ゲノム塩基配列が終了した領域について遺伝地図上のマーカーや EST と対応させた物理地図情報とともに、予測遺伝子の詳細な情報を提示している。

2004 年末までの遺伝子予測の結果を要約すると、予測遺伝子の密度は 6.4 ~ 7.4 kb に 1 個で、イネのゲノムサイズ 400Mb から推定される全ゲノム中の遺伝子数は 54,000 ~ 62,000 個となり、シロイヌナズナの約 25,500 個から較べるとかなり多いことが推定される。また、類似性検索の結果、既知遺伝子と類似性を示すものと示さないものの比はほぼ半々であった。

なお、日本以外の国・地域が担当している残りの 6 本については、アノテーションは未完成のものが多く、今後完全長 cDNA との照合等も含めて、日本、米国の協力のもとに完成に向けて努力していく必要がある。

6. DNAライブラリー

イネゲノム塩基配列の解読に当たっては、ゲノム DNA を各種の方法で断片化し、保存あるいは増殖し、塩基配列解読の実験材料にするライブラリーという形のものを、目的に応じて作製した。ライブラリーは-85℃の冷凍庫に保管し、必要に応じて取り出し研究に



写真1. フリーザー室

各ライブラリーのクローンは大腸菌プラスミドに組み込まれ、-85℃の冷凍庫に保管し、必要に応じて取り出し、増殖し実験に使えるように整えてある。また、外部の研究者にも譲渡ができる体制が整えられている。

供する。また、将来長期間にわたって利用することが予想されるので、細心の注意を払い完全な形での永久保存をしていかねばならない。

このようなライブラリーは各断片が染色体のどの位置から切り出されたものかを調べ、地図の形で配置図が作成され、その塩基配列が解読される。したがって、ライブラリーはその位置情報と配列情報とが3点セットになっている。これらのライブラリーもイネゲノム研究から生まれた大きな成果であり、財産である。

今後イネの配列情報を利用し、遺伝子の機能研究あるいは単離、あるいは DNA マーカー

育種等を進めるに当たっては、目的とする領域のクローンが必要になることが多くなる。

世界の研究者のクローン分譲の要請に対応するため、農業生物資源研究所と農林水産先端技術研究所はこのような DNA ライブラリーのクローン譲渡の体制を徐々に整えてきている。現在までに譲渡体制ができあがっているライブラリーは遺伝地図に対応する RFLP と物理地図に対応する YAC ライブラリー、発現遺伝子地図に対応する cDNA ライブラリーである。また、塩基配列解読の際に材料として使った PAC/BAC ライブラリーなどについても永久保管をし、外部の研究者が利用できるような体制が検討されていくものと考えられる。