Plant Pathol. J. 16(2) : 83-89 (2000)

The Plant Pathology Journal

Symposium

Molecular Plant-Microbe Interactions

December 10, 1999, Taejon, Korea

Genome Sequencing of *Magnaporthe grisea* Using BAC and Shotgun Libraries

Young Jin Koh

Department of Applied Biology, Sunchon National University, Sunchon 540-742, Korea

(Received on February 22, 2000)

The fungus, *Magnaporthe grisea* (Hebert) Barr (anamorph *Pyricularia grisea* Sacc.), is the causal agent of the blast disease which is the most destructive disease of rice (*Oryza sativa*) (Rossman et al., 1990). The fungus has been intensively investigated as a valuable model organism for studying various aspects of host-pathogen interactions in recent years (Valent, 1990; Valent and Chumley, 1991), because it is a well-known haploid, heterothallic Pyrenomycete fungus, producing perithecia containing fusiform, curved ascospores in unordered asci (Kang et al., 1994) and highly fertile laboratory strains have been developed (Chao and Ellingboe, 1991; Kolmer and Ellingboe, 1988; Leung et al., 1988; Valent et al., 1986). As a result, numerous genes providing insight into mechanisms of fungal pathogenicity have been cloned in *M. grisea* over the past few years (Mitchell and Dean, 1995; Sweigard et al., 1992; Talbot et al., 1993; Xu and Hamer, 1996; Zhu et al., 1996). An effective transformation system has been established (Zhu et al., 1999) and several high-density molecular linkage maps have been created containing RFLP and phenotypic markers (Nitta et al. 1997; Romao and Hamer, 1992; Skinner et al., 1993; Sweigard et al., 1993). Extensive studies on electro-karyotyping have been reported for this haploid fungus with a genome size ranging from 37 to 50 Mb depending on strains (Orbach et al 1996; Skinner et al. 1993). Therefore, *M. grisea* is an ideal candidate for a whole genome sequencing project, not only because it is a model organism suited for genetic analysis, but also because it has a relatively small genome size compared to other complex eukaryotes. In this article, genome sequencing of *M. grisea* using BAC and shotgun libraries will be reviewed.

*Corresponding author
Phone, Fax) +82-661-750-3865
E-mail) youngjin@sunchon.ac.kr

BAC library

The construction of overlapping large fragments of DNA (contigs) and map-based cloning are most useful for gaining insight into the biology of *M. grisea* and other eukaryotes. Large-insert libraries play a pivotal role in such research. Two of the commonly available vector systems to clone large insert DNA fragments (>100 kb) are yeast artificial chromosomes (YAC) and the newer bacterial artificial chromosomes (BAC) (Shizuya et al., 1992). In the YAC cloning system, the size-selected large fragments of exogenous DNA are cloned into a linearized YAC vector and maintained in yeast as linear chromosomes. In the BAC cloning system, large fragments of DNA are cloned into an *Escherichia coli* F factor-based plasmid vector and maintained in *E. coli* as circular plasmids.

The two systems are similar in that both can easily handle large pieces (>100 kb) of DNA, maintained at low-copy numbers in the host cells (Zhang et al., 1996). However, compared to the YAC system, BAC libraries are relatively easy to construct with few chimeric clones (Ioannou et al., 1994; Shizuya et al., 1992; Woo et al., 1994). It is also easier to extract a large amount of pure BAC DNA using standard plasmid DNA isolation techniques that take advantage of the separation of supercoiled plasmid DNA from bacterial genomic DNA. Large insert libraries offer a number of advantages over cosmid libraries. For example, if the objective is to assemble a 1-Mb overlapping contig, 49 steps would be required using a cosmid library with the average insert size of 40 kb assuming 50% overlap between clones. However, if a BAC library was used (assuming a 150 kb average size), it would require only 13 steps with a 90% probability of completing the walk (Zhang et al., 1996a).

BAC libraries have been constructed for a number of important plant and fungal species (Woo et al., 1995; Wang et al., 1995; Zhang et al., 1996a; Choi et al., 1995; Shan et al., 1996; Diaz-Perez et al. 1996). Most recently, Zhu et al.

(1997) constructed a high quality novel BAC library from *M. grisea* strain 70-15, a strain derived from the field isolate Guy11 (Chao and Ellingboe, 1991; Skinner et al., 1993). The average insert size of this BAC library is 130 kb, which is comparable to the majority of human and plant BAC libraries constructed to date (Cai et al., 1995; Choi et al., 1995; Ioannou et al., 1994; Kim et al., 1996; Shizuya et al., 1992; Woo et al., 1994; Zhang et al., 1996a). Assuming the genome size of *M. grisea* strain 70-15 is 40 Mb, the library represents >25 genome equivalents that statistically means there is a >99.99% chance of finding at least one specific BAC clone when screening with a specific sequence. Although the insert size of the BAC library is several fold smaller than YAC libraries, it is easier to handle and any deficiencies can be compensated for by the construction of several BAC libraries using different restriction enzymes. The utility of this library is demonstrated through the isolation of a gene of interest and the construction of a 550 kb contig composed of 6 BAC clones that is anchored to *M. grisea* chromosome 2 by RFLP markers (Zhu et al., 1997). BAC clones also seem ideal for producing an accurate contiguous sequence because inserts showing minimal overlaps at each end of BAC clone are sequenced by cloning directly into M13 or plasmid vectors (Venter et al., 1996). Therefore, development of a novel BAC library from *M. grisea* strain 70-15 will provide a key advance in sequencing works of *M. grisea* as a model system to study fungal genomics, because BAC fingerprinting and BAC-end sequence provide framework, i.e. sequence-tagged connectors (STCs) to facilitate whole genome sequencing of *M. grisea* (Zhu et al., 1997). BAC DNA fingerprint techniques are constructing physical maps for the entire genome of *M. grisea*.

Shotgun Library and Sequencing

Recent advances in DNA sequencing technologies have demonstrated that whole genomes can be sequenced accurately and efficiently. In spite of a number of alternative sequencing strategies, most sequencing information about relatively small genome organisms has been generated by traditional shotgun sequencing (Chen et al., 1996; Fraser and Fleischmann, 1997; Radelof et al., 1998). The whole genome of *Haemophilus influenzae* was the first to be completely sequenced by shotgun method (Fleischmann et al., 1995), followed by whole genome sequencing of several free-living microorganisms (Bult et al., 1996; Deckert et al., 1998; Fraser et al., 1995; Klenk et al., 1997). However, for larger genomes including fungi, a necessary condition for whole shotgun sequencing is the existence of a random library of cloned chromosomal DNA fragments (Fraser and Fleischmann, 1997).

The random shotgun approach for sequencing is to subdivide the insert from a cosmid or BAC into a large number of short fragments by mechanical shearing and clone these fragments into plasmid or M13 vectors (Ahmed and Podemski, 1997). BAC clones serve excellent substrates to generate a number of subclones for shotgun sequence analysis because of their large insert size and ease of DNA isolation that lends itself to automation (Shizuya et al., 1992; Venter et al., 1996). A random library is best constructed from randomly sheared fragments of BAC library, since any kind of enzymatic cleavage is generally somewhat nonrandom but mechanical shearing is insensitive to sequence context. Extraction by the usual phenol/chloroform method mechanically shears the genomic DNA to fragments of about 50 kb. The DNA can be further sheared into pieces by sonication or nebulization to maximize the randomness of the DNA fragments. Fragments with narrow range between 1.6 and 2.0 kb in size are usually excised and recovered to minimize variation in growth of clones after subcloning into pBluescript vector (Fleischmann et al., 1995). For shotgun sequencing of a 112 kb BAC clone 6J18 of *M. grisea* chromosome 7, a shotgun library could be practically constructed from BAC DNA fragments nebulized with N_2 at 6.5 psi for 3 min. The shotgun library has an average insert size of 1.9 kb and contains 3,360 individual clones, which show about 12 fold coverage of a 112 kb BAC clone (Koh et al., unpublished).

Fluorescent labeling of DNA fragments generated by the Sanger dideoxy chain termination method has been the mainstay of almost all large scale sequencing projects since the introduction of the first semi-automated sequencer by Applied Biosystems in 1987 and the development of Taq cycle sequencing in 1990 (Venter et al., 1998). Recently new models of the sequencer that can process more samples, Taq polymerase engineered especially for sequencing, and higher sensitivity dyes have improved throughput, accuracy, and operating costs for sequencing.

Random shotgun sequencing requires oversampling to ensure a minimal coverage of under-represented regions because an unequal representation of different parts of the sequence will be expected due to sampling effects. Completed shotgun projects show an 8-12 fold average coverage per base sequence, which is significantly more redundant than necessary to achieve consensus sequence data of sufficient quality (Fraser and Fleischmann, 1997). Therefore, several strategies as cost-effective and efficient approaches have been suggested recently to reduce redundancy in large-scale sequencing projects (Chen et al., 1996; Radelof et al., 1998). A useful intermediate product for genome sequencing might be obtained by 3-4 fold redundancy and nearly complete sequence data could be obtained at 6 fold redundancy through analysis of the quality and utility of

random shotgun sequencing at low redundancies (Bouck et al., 1998). Nearly complete sequence data (95% of the complete sequence data) could be also obatined at about 5 fold redundancy by random shotgun sequencing of a 112 kb BAC clone 6J18 of *M. grisea* chromosome 7 (Koh et al., unpublished).

Contig Assembly and Open Reading Frame Identification

The finishing of the shotgun sequencing is done by combination of computer works, identification of major contigs and finding the insert ends. After the resulting sequence data are usually transferred to a UNIX SUN Sparc Workstation, contig building is initiated using the base-calling program Phred and the assembly program Phrap (Ewing and Green, 1998; Ewing et al., 1998) to find overlaps among sequences and build a framework physical map by connecting groups of clones using the pairwise relationships of sequences from individual shotgun clones. A complete genome sequence is obtained by sequencing across the gaps between contigs. Usually genomes with high GC and AT composition present special problems for sequencing and assembly (Venter et al., 1998). Poly-G sequence makes it difficult to fill the gap and get good quality sequence around the region in finishing stage. Closure of sequence gaps is usually conducted by an appropriate oligonucleotide design followed by a dye terminator sequence walk along the appropriate template (Fraser and Fleischmann, 1997). Physical gaps can be closed by PCR products (both standard and long range) generated by doing all possible reactions with primers designed from the ends of each contig group.

Full identification and annotation of open reading frames (ORFs) is an integral part of finishing the assembly. The final genome sequence is searched with a set of search algorithms incorporating BLAST (Altschul et al., 1990; Madden et al., 1996) against both nucleotide and amino acid databases for the identification of genes. Exon candidates can be sought by the GRAIL (Gene Recognition and Analysis Internet Link), GenScan and WebGene programs and also by comparison with GenBank entries for expressed sequence tags (ESTs) (Adams et al., 1991) and cDNAs. Common names for genes and roles of each putative ORF are stored in the final annotation of the sequence.

Total gene numbers and genome sizes reveal the genetic complexity and relative amount of functional information of organisms. The genome sizes of several ascomycetous fungi along with the expected gene densities and gene complexities are listed in Table 1 (Kupfer et al., 1997). *M. grisea* genome size most likely has a gene density of 225 genes/Mb and a predicted gene number of 8500. Thirty seven ORFs were identified by a BLAST search of the

Table 1. The genome sizes of several ascomycetous fungi along with the expected gene densities and gene complexities

	Genome physical data[a]		
	Size (Mb)	Density (genes/Mb)	Total (complexity)
Unicellular fungi	**13.7**	**470**	**6400**
Saccharomyces cerevisiae	13.5	430	5800
Saccharomyces pombe	14.0	500	7000
Multicellular fungi	**36**	**233**	**8400**
Subclass Plectomycetidae	**33**	**240**	**8000**
Aspergillus nidulans	31	260	8100
Related Anamorphs			
Penicillium chrysogenum	34	240	8200
Trichoderma reesei	33	235	7800
Subclass Pyrenomycetes	**40**	**225**	**8900**
Nectria crassa	42	230	9200
Nectria haematococca	40	220	8800
Magnaporthe grisea	37	225	8500

[a] Data from Kupfer et al. (1997).

GenBank database as a result of shotgun sequencing of a 112 kb BAC clone 6J18 which was selected for its feature of high gene density on chromosome 7 of *M. grisea* based on cDNA hybridization study (Koh et al., unpublished). The distribution of all the ORFs along 112 kb sequence are shown in Fig. 1 as drawn mainly according to GenScan image. Sixteen ORFs predicted by GenScan reveal significant homologous sequences with known genes (Table 2), but it is likely that twenty one ORFs with no significant homology to any entries in public databases represent individual potential genes, pseudogenes, exons, or a transposable elements.

Functional Analysis

The complete genomic sequence is a tremendous asset in the study of *M. grisea*. It is quite likely to reveal essentially all the predicted coding regions as well as a variety of other features, such as ribosomal RNA operons, IS elements, repeat regions, G-C content, origin(s) of replication, operon structure, genome polarity, etc. Currently, genomic sequence analysis primarily involves the identification of the predicted coding regions followed by the assignment of a tentative gene name and functional role based upon a sequence similarity search. The use of sequence similarity as a method for gene identification will at best provide us with functional information for perhaps 50% of the predicted coding regions. The remainder often appear to match genes identified as hypothetical from other organisms or appear to have no match in the existing databases. The inability to identify these genes can rest in two areas, the sequences are

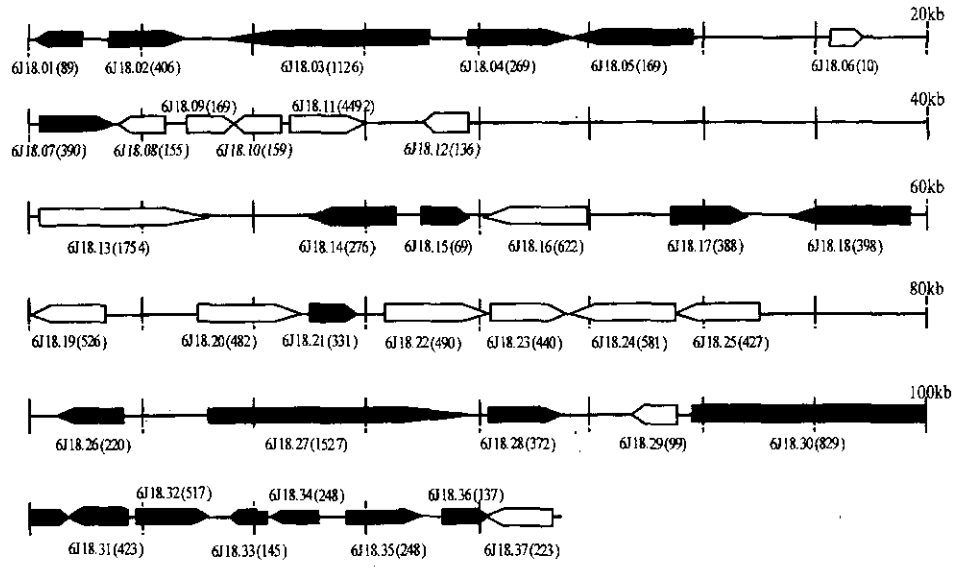

Fig. 1. The distribution of ORFs in BAC clone 6J18 of *Magnaporthe grisea*. A schematic view of the distribution of 37 ORFs predicted by GenScan. Each arrows indicate the direction of translation. Black-filled arrows (▶) indicate ORFs with no homology to other known genes, and open arrows (▷) indicate a presence of homologs in other organisms. The length of the gene products are shown in bracket in the number of amino acids

Table 2. Sequence homology-based assignment of 16 open reading frames in a 112 kb BAC clone 6J18 of *Magnaporthe grisea* chromosome 7

ORF	Coding region	Size (aa)	The most significant sequence homology[a]
6J18.06	18144>18174	10	5S ribosomal RNA gene [*Neurospora crassa*]
6J18.08	22449<23121	155	Hypothetical protein [*Synechocystis* sp.]
6J18.09	24247>24905	169	Peroxisomal-like Protein [*Aspergillus fumigatus*]
6J18.10	25386<26279	159	60S ribosomal protein [*Shizosaccharomyces pombe*]
6J18.11	26882>28402	492	Adenosine deaminase [*S. pombe*]
6J18.12	28890<29425	136	Transcription factor [*S. pombe*]
6J18.13	40259>45532	1754	*ZipA* [*Dictyostelium discoideum*]
6J18.16	50410<52370	622	Acetyl-coenzyme A transporater [*Homo sapiens*]
6J18.19	60031<61638	536	*pot3*, Transposase [*Magnaporthe grisea*]
6J18.20	63851>65844	482	Tetracycline transporter-like protein mRNA [*H. sapiens*]
6J18.22	68597>70165	490	Probable membrane protein [*Saccharomyces serevisiae*]
6J18.23	70854>72173	440	Hypothetical protein [*S. serevisiae*]
6J18.24	72411<74155	581	Putative amidase [*Streptomyces coelicolor*]
6J18.24	74560<75607	427	*krev-1* [*N. crassa*]
6J18.29	93133<93504	99	*rac1*, GTP-binding protein [*H. sapiens*]
6J18.37	110728<111398	223	Putative ubiquinone biosynthesis protein [*S. pombe*]

[a] ORFs were predicted by Genscan. The approximate coding regions were determined also with reference to the Blastn and Blastx results.

dissimilar enough so as to score poorly by the commonly used sequence comparison algorithms, or they encode proteins that represent new biology. Identifying the functional significance of these unknown genes remains a challenge, and developing methods of doing so in high-throughput ways will be important for keeping pace with the large amount of genomic data that will be amassed in the next few years. While elucidating the function of an encoded protein may remain the province of the biochemist for quite some time, the attachment of additional information to each predicted coding region will provide knowledge of its potential role if not its actual function. This knowledge may take the form of identifying whether a gene's expression is essential for viability as well as quantitative differences in its temporal expression, under a variety of environmental stresses, with respect to differentiation, and pathology (Fraser and Fleischmann, 1997).

A number of high-throughput methodologies, developed

Table 3. Strategies for gene identification and functional analysis

Methodology	Comments
Expressed sequence tags (ESTs)	High-throughput gene identification. Provides ten of thousands of sequence-based tags for gene identification, tissue distribution, abundance, and gene family analysis. Requires moderate to large-scale DNA sequencing facility. Unique tags require oligo(dT) priming.
Serial analysis of gene expression (SAGE)	High-throughput technique for measuring transcript abundance based on position and content of small sequence tags. Requires only small DNA sequencing facility to produce thousands of tags. Dependent on polyadenylated mRNA.
Differential display methods	Methods for displaying differences in gene expression levels. Especially valuable when there is no *a priori* knowledge of gene content. Has not been widely used for bacteria.
cDNA microarray analysis	Technology for high-throughput characterization of mRNA population based on hybridization to cDNA microarrays. Potentially valuable for whole genome characterization.
Gene disruption and DNA chip	Technology for high-throughput analysis of uncharacterized ORFs by combining gene disruption with DNA molecular bar coding to be read by DNA chips synthesized by photolithographic techniques.

Summary from Fraser and Fleischmann (1997).

over the past few years and under current development, provide resources for gene identification and functional analysis (Table 3). ESTs provide a rapid method for gene identification by combining sequence-based tags with *in silico* methods of sequence comparison (Adams et al., 1991). ESTs have been applied to eukaryotic organisms where polyadenylated mRNA provides a unique 3' tag, although in principle, randomly primed cDNA could provide a means for rapid gene identification in nonpolyadenylated messages. Most recently, Choi et al. (1998) have initiated an EST project to identify and analyze all possible genes expressed during appressorium formation in *M. grisea*. Serial gene analysis (SAGE) is a rapid method for quantitative mRNA expression level analysis (Velculescu et al., 1995). However, the method has been limited in application to polyadenylated message. Differential display methodologies alone, or in combination with subtractive methods, can provide gene expression data (Liang and Pardee, 1992). While these methods have the advantage that they are not truly high-throughput and require extensive follow-up work to characterize the differentially displayed bands. Methodologies based on DNA chip technology may represent the most promising advances in quantitative gene expression and functional analysis for microbial organisms (Shoemaker et al., 1996). Advances in arraying technology, combined with fluorescent hybridization and sensitive detection methods, make this technology the most robust for moving forward, especially as the sequence of dozens of microbial organisms becomes available in the near future (DeRisi et al., 1996)

Conclusion

The introduction of automated DNA sequence analysis nearly a decade ago, together with more recent advances in the field of bioinformatics, have revolutionized biology and medicine and have ushered in a new era of genomic science, the study of genes and genomes (Fraser and Fleischmann, 1997). These new technologies have had an impact on many areas of research, including the association between genes and disease, in DNA-based diagnostics, and in the sequencing of genomes from human and other model organisms. The demonstration in 1995, that automated DNA sequencing methods could be used to decipher the entire genome sequence of a free-living organism, *Haemophilus infuenzae*, was a milestone in both the genomics and microbial fields (Fleischmann et al., 1995). Since the first report of the complete sequence of *H. influenzae*, these methodologies have been adopted by laboratories around the world. The complete genomic sequences of several have been reported in recent years. It is likely that in the next few years we will see the complete sequence of perhaps as many as 30-40 microbial genomes.

Until now the 112 kb BAC clone 6J18 of *M. grisea* chromosome 7 was completely sequenced using a random shotgun library as a pilot project towards the completion of whole genome sequencing of *M. grisea*. This covers only less than 1/40 of chromosome 7 (4.2 Mb) and 1/400 of the whole genome of *M. grisea*. However, the complete genome sequence will be obtained in the next few years with a high quality novel BAC library of *M. grisea* and a coordinated effort within the *Magnaporthe* genome community. The outcome of such an effort would produce valuable information not only in the genetical and biological studies of *M. grisea*, but also in the areas of fungal biology and plant pathology.

References

Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R. and Ven-

ter, J. C. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252:1651-1656.

Ahmed, A. and Podemski, L. 1997. Use of ordered deletions in genome sequencing. *Gene* 197:367-373.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.

Bouck, J., Miller, W., Gorrel, J. H., Muzny, D. and Gibbs, R. A. 1998. Analysis of the quality and utility of random shotgun sequencing at row redundancies. *Genome Res.* 8:1074-1084.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weidman, J. F., Fuhrmann, J. L., Nguyen, D., Utterback, T. R., Kelly, J. M., Peterson, J. D., Sadow, P. W., Hanna, M. C., Cotton, M. D., Roberts, K. M., Hurst, M. A., Kaine, P., Borodovsky, M., Klenk, H. P., Fraser, C. M., Smith, H. O., Woese, C. R. and Venter, J. C. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii. Science* 273:1058-1073.

Cai, L., Taylor, J. F., Wing, R. A., Gallagher, D. S., Woo, S. S. and Davis, S. K. 1995. Construction and characterization of a bovine bacterial artificial chromosome library. *Genomics* 29:413-425.

Chao, C. T. and Ellingboe, A. H. 1991. Selection for mating competence in *Magnaporthe grisea* pathogenic to rice. *Can. J. Bot.* 69:2130-2134.

Chen, C. N., Su, Y., Baybayan, P., Siruno, A., Nagaraja, R., Mazzarella, R., Schlessinger, D. and Chen, E. 1996. Ordered shotgun sequencing of a 135 kb Xq25 YAC containing ANT2 and four possible genes, including three confirmed by EST matches. *Nucl. Acids Res.* 24:4034-4041.

Choi, S-D., Creelman, R., Mullet, J. and Wing, R. A. 1995. Construction and characterization of a bacterial artificial chromosome library from Arabidopsis thaliana. *Weeds World* 2:17-20.

Choi, W-B., Fang, E. and Dean R. A. 1998. Sequence analysis and characterization of genes expressed during appressorium formation in *Magnaporthe grisea. Phytopathology* 88:S17.

Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Aujay, M., Huber, R., Feldman, R. A., Short, J. M., Olsen, G. J. and Swanson, R. V. 1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus. Nature* 392:353-358.

DeRisi, J., Penland, L., Brown, P. O. (Group 1); Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., Trent, J. M. (Group 2). 1996. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet.* 14:457-460.

Diaz-Perez, S. V., Crouch, V. W. and Orbach, M. J. 1996. Construction and Characterization of a *Magnaporthe grisea* Bacterial Artificial Chromosome Library. *Fungal Genet. Biol.* 20:280-288.

Ewing, B. and Green, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8:186-194.

Ewing, B., Hillier, L., Wendl, M. C. and Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8:175-185.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G. G., Fitz-Hugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L. I., Glodek, A., Kelley, J. M., Weidman, J. F., Philips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O. and Venter, J. C. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496-512.

Fraser, C. M. and Fleischmann, R. D. 1997. Strateries for whole microbial genome sequencing and analysis. *Electrophoresis* 18:1207-1216.

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Keravage, A. R., Sutton, G., Kelly, J. M., Fritchman, J. L., Weidman, J. F., Small, K. V., Sandusky, M., Fuhmann, J., Nguyen, D., Utterback, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J. F., Dougherty, B. A., Bott, K. F., Hu, P. C., Lucier, T. S., Peterson, S. N., Smith, H. O., Hutchison III, C. A. and Venter, J. C. 1995. The minimal gene complement of *Mycoplasma genitalium. Science* 270:397-403.

Ioannou, P. A., Amemiya, C. T., Garnes, J., Kroisel P. M., Shizuya, H., Chen, C., Batzer, M. A. and de Jong, P. J. 1994. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.* 5:84-89.

Kang, S., Chumley, F. G. and Valent, B. 1994. Isolation of the mating-type genes of the phytopathogenic fungus *Magnaporthe grisea* using genomic subtraction. *Genetics* 138:289-296.

Kim, U-J., Birren, B., Mancino, V., Slepak, T., Boysen, C., Kang, H-L., Simon, M. I. and Shizuya, H. 1996. Construction and characterization of human Bacterial Artificial Chromosome library. *Genomics* 34:213-218.

Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., Richardson, D. L., Kerlavage, A. R., Graham, D. E., Kyrpides, N. C., Fleischmann, R. D., Quackenbush, J., Lee, N. H., Sutton, G. G., Gill, S., Kirkness, E. F., Dougherty, B. A., McKenney, K., Adams, M. D., Loftus, B., Peterson, S., Reich, C. I., McNeil, L. K., Badger, J. H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J. D., Weidman, J. F., McDonald, L., Utterback, T., Cotton, M. D., Spriggs, T., Artiach, P., Kaine, B. P., Sykes, S. M., Sadow, P. W., D'Andrea, K. P., Bowman, C., Fujii, C., Garland, S. A., Mason, T. M., Olson, G. J., Fraser, C. M., Smith, H. O., Woese, C. R. and Venter, J. C. 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus. Nature* 390:364-370.

Kolmer, J. A. and Ellingboe, A. H. 1988. Genetic relationship between fertility and pathogenicity and virulence in rice blast fungus *Magnaporthe grisea. Can. J. Bot.* 66:981-987.

Kupfer, D. M., Reece, C. A., Clifton, S. W., Roe, B. A. and Prade, R. A. 1997. Multicellular ascomycetous fungal genomes contain more than 8000 genes. *Fungal Genet. Biol.* 21:364-372.

Leung, H., Borromeo, E. S., Bernardo, M. A. and Notteghem, J. L. 1988. Genetic analysis of virulence in rice blast fungus *Magnaporthe grisea*. *Phytopathology* 78:1227-1223.

Liang, P. and Pardee, A. B. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257:967-971.

Madden, T. L., Tatusov, R. L. and Jhang, J. 1996. Applications of Network BLAST Server. In: *Methods in Enzymology* ed. by R. F. Doolittle, Vol. 226, pp. 131-141. Academic Press, New York.

Mitchell, T. K. and Dean, R. A. 1995. The cAMP-dependent protein kinase catalytic subunit is required for appressorium formation and pathogenicity by the rice blast pathogen *Magnaporthe grisea*. *Plant Cell* 7:1869-1878.

Nitta, N., Farman, M. L. and Leong. S. A. 1997. Genome organization of *Magnaporthe grisea*: integration of genetic maps, clustering of transposable elements and identification of genome duplications and rearrangements. *Theor. Appl. Genet.* 95:20-32.

Orbach, M. J., Chumley, F. G. and Valent, B. 1996. Electrophoretic karyotypes of *Magnaporthe grisea* pathogens of diverse grasses. *Mol. Plant-Microbe Interact.* 9:261-271.

Radelof, U., Hennig, S., Seranski, P., Steinfath, M., Ramser, J., Reinhardt, R., Poustka, A., Francis, F. and Lehrach, H. 1998. Preselection of shotgun clones by oligonucleotide fingerprinting: an efficient and high throughput strategy to reduce redundancy in large-scale sequencing projects. *Nucl. Acids Res.* 26:5358-5364.

Romao, J. and Hamer, J. E. 1992. Genetic organization of a repeated DNA sequence family in the rice blast fungus. *Proc. Natl. Acad. Sci. USA* 89:5316-5320.

Rossman, A. Y., Howard, R. J. and Valent, B. 1990. *Pyricularia oryzae*, the correct name for the rice blast disease fungus. *Mycologia* 82:509-512.

Shan, W., Cao, M., Leung, D., Arrendondo, F., Foster, H., Coffey, M. D. and Tyler, B. 1996. Genetics and cloning of avirulence genes in *Phytophthore sojae*. In Abstract J53: 8th International Congress Molecular Plant-Microbe Interactions. Knoxville, TN.

Shizuya, H., Birren, B., Kim, U-J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-vector-based vector. *Proc. Natl. Acad. Sci. USA* 89:8794-8797.

Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M. and Davis, R. W. 1996. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nature Genet.* 14:450-456.

Skinner, D. J., Budde, A. D., Farman, M. L., Smith, J. R., Leung, H. and Leong, S. A. 1993. Genome organization of *Magnaporthe grisea*: genetic map, electrophoretic karyotype, and occurrence of repeated DNAs. *Theor. Appl. Genet.* 87:545-557.

Sweigard, J. A., Chumley, F. G. and Valent, B. 1992. Cloning and analysis of CUT1, a cutinase gene from *Magnaporthe grisea*. *Mol. Gen. Genet.* 232:174-182.

Sweigard, J. A., Valent, B., Orbach, M. J., Walter, A. M., Rafalski, A. and Chumley, F. G. 1993. Genetic map of the rice blast fungus *Magnaporthe grisea*. In: *Genetic Maps* ed. by S. J. O'Brien, pp. 3.112-3.117. Cold Spring Harbor Laboratory, New York.

Talbot, N. J., Ebbole, D. J. and Hamer, J. E. 1993. Identification and characterization of MPG1, a gene involved in pathogenicity from rice blast fungus *Magnaporthe grisea*. *Plant Cell* 5:1575-1590.

Valent, B. 1990. Rice blast as a model system for plant pathology. *Phytopathology* 80:33-36.

Valent, B. and Chumley F. G. 1991. Molecular genetic analysis of the rice blast fungus, *Magnaporthe grisea*. *Annu. Rev. Phytopathol.* 29:443-467.

Valent, B., Crawford, M. S., Weaver, C. G. and Chumley, F. G. 1986. Genetic studies of fertility and pathogenicity in *Magnaporthe grisea* (*Pyricularia oryzae*). *Iowa State J. Res.* 60: 569-594.

Velculescu, V. E., Zhang, L., Vogelstein, B. and Kinzler, K. W. 1995. Serial analysis of gene expression. *Science* 270:484-487.

Venter, J. C., Adams, M. D., Sutton, G. G., Kerlavage, A. R., Smith, H. O. and Hunkapiller, M. 1998. Shotgun sequencing of the human genome. *Science* 280:1540-1542.

Venter, J. C., Smith, H. O. and Hood, L. 1996. A new strategy for genome sequencing. *Nature* 381:364-366.

Wang, G-W., Warren, R., Innes, R., Osborne, B., Baker, B. and Ronald, P. C. 1995. Construction of an Arabidopsis BAC library and isolation of clones hybridizing with disease-resistance, gene-like sequences. *Plant Mol. Biol. Rep.* 107-113.

Woo, S-S., Gill, B. S., Paterson, A. H. and Wing, R. A. 1994. Construction and characterization of a bacterial artificial chromosome library of Sorghum bicolor. *Nucl. Acids. Res.* 22:4922-4931.

Xu, J-R. and Hamer, J. E. 1996. MAP kinase and cAMP signaling regulate infection structure formation and pathogenic growth in the rice blast fungus *Magnaporthe grisea*. *Gen. & Dev.* 10:2696-2706.

Zhang, H., Choi, S-D., Woo, S-S., Li, Z. and Wing, R. A. 1996a. Construction and characterization of two rice bacterial artificial chromosome libraries from parents of a permanent recombinant inbred mapping population. *Mol. Breeding* 2:11-24.

Zhang, H., Woo, S-S. and Wing, R. A. 1996b. BAC, YAC, and cosmid library construction. In: *Plant Gene Isolation: Principles and Practice* ed. by G. D. Foster and T. Twell, pp. 75-99. John Wiley & Sons Ltd.

Zhu, H., Choi, S., Johnston, A. K., Wing, R. A. and Dean, R. A. 1997. A large-insert (130 kbp) bacterial artificial chromosome library of the rice blast fungus *Magnaporthe grisea*: genome analysis, contig construction, and gene cloning. *Fungal Genet. Biol.* 21:337-347.

Zhu, H., Whitehead, D. S., Lee, Y-H. and Dean, R. A. 1996. Genetic analysis of developmental mutants and rapid chromosome mapping of APP1, a gene required for appressorium formation in *Magnaporthe grisea*. *Mol. Plant-Microbe Interact.* 9:767-774.

Zhu, H. and Dean, R. A. 1999. A novel method for increasing the transformation efficiency of *Escherichia coli* - application for bacterial artificial chromosome library construction. *Nucleic Acids Res.* 27:910-911.