Hybridization-Based Mapping of *Neurospora crassa* Linkage Groups II and V

Verena Aign,* Ulrich Schulte† and Jörg D. Hoheisel*

**Functional Genome Analysis, Deutsches Krebsforschungszentrum, D-69121 Heidelberg, Germany and*
†Institute of Biochemistry, Heinrich-Heine Universität, D-40225 Düsseldorf, Germany

Manuscript received October 6, 2000
Accepted for publication November 30, 2000

ABSTRACT

As part of the German *Neurospora crassa* genome project, physical clone maps of linkage groups II and V of *N. crassa* were generated by hybridization-based mapping. To this end, two different types of clone library were used: (1) a bacterial artificial clone library of 15-fold genome coverage and an average insert size of 69 kb, and (2) three cosmid libraries—each cloned in a different vector—with 17-fold coverage and 34 kb average insert size. For analysis, the libraries were arrayed on filters. At the first stage, chromosome-specific sublibraries were selected by hybridization of the respective chromosomal DNA fragments isolated from pulsed-field electrophoresis gels. Subsequently, the sublibraries were exhaustively ordered by single clone hybridizations. Eventually, the global libraries were used again for gap filling. By this means, physical maps were generated that consist of 13 and 21 contigs, respectively, and form the basis of the current sequencing effort on the two chromosomes.

THE filamentous fungus *Neurospora crassa* has been used as a model organism in basic science for more than 50 years. It was instrumental in the work resulting in the development of the one-gene-one-enzyme hypothesis (Beadle and Tatum 1941) and served in many other fields of basic research. This work has accumulated a large body of knowledge witnessed by more than 5000 published articles. A detailed genetic map was developed that includes >800 genes (Perkins *et al.* 1982; Perkins 1992), and >300 genes were cloned. A large collection of strains and mutants is maintained at the Fungal Genetic Stock Center in Kansas City, MO. Two cDNA sequencing projects at the University of New Mexico and the Oklahoma State University created >20,000 expressed sequence tags (ESTs; http://molbio.ahpcc.unm.edu/search/ngp.html; http://www.genome.ou.edu/fungal.html). Blast analysis of this data indicated that a large portion of all identified genes have no clear homologue elsewhere (Nelson *et al.* 1997).

All these efforts combined, however, characterized only a portion of the estimated 10,000–13,000 genes of *N. crassa*. To overcome this limitation to the understanding of the fungus's biology, a genome initiative was established (Bennett 1997). The entire genomic sequence of *N. crassa* will be analyzed in a collaboration between laboratories in Germany and the United States. It has a length of ~43 Mb split among seven chromosomes. The content of G:C base pairs of 52% and low number of repetitive sequences make the Neurospora genome amenable to normal large-scale sequencing procedures (Radford and Parish 1997). Because of the ability to separate the chromosomes physically by pulsed-field gel electrophoresis (PFGE; Orbach *et al.* 1988), a chromosome-oriented approach is possible. The German project aims at the sequencing of the two chromosomes that correspond to linkage groups II and V, estimated to be 4.6 and 9.2 Mb in length (Orbach *et al.* 1988). The first step in this project was the creation of physical maps. This was accomplished by hybridization-based mapping, which had been established as an efficient method to such end (*e.g.*, Hoheisel *et al.* 1993; Johnston *et al.* 1997; Hanke *et al.* 1998; Frohme *et al.* 2000). Prior to the start of this project, a cosmid library of 13-fold genome coverage (Orbach 1994; Kelkar *et al.* 2001, this issue) was already available. An additional cosmid library of some four genome equivalents and a library of bacterial artificial chromosomes (BACs) of 15-fold coverage were constructed to complement this resource. Using clone hybridization procedures on these libraries, physical clone maps were produced that correspond to linkage groups II and V and form the basis of the currently ongoing sequencing of the two chromosomes.

Corresponding author: Verena Aign, Functional Genome Analysis, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 506, D-69121 Heidelberg, Germany. E-mail: v.aign@dkfz-heidelberg.de

MATERIALS AND METHODS

DNA preparation: For the production of clone libraries, DNA from *N. crassa* strain 74-OR23-1A was used. High molecular weight genomic DNA was isolated as described (Herrmann and Frischauf 1987). For preparation of intact chromosomal DNA (Orbach *et al.* 1988), 1×10^9 conidia were inoculated in 150 ml Vogel's medium and grown at 30° and 200 rpm until nearly all conidia germinated, usually 3–4 hr. They were harvested, pelleted, washed twice with 30 ml ice-cold water, and taken up in 0.5 ml 50 mm EDTA. The cells were mixed with 1.7 ml of molten 1.7% low-melting-point agarose in 125

mM EDTA and 50 mM sodium-citrate, pH 4.0. To this, 333 μl Novozym-234 (20 mg/ml in 1 M sorbitol) were added, and the mixture was pipetted into a plug former. Solidified plugs were placed in 0.4 M EDTA, 50 mM sodium-citrate, pH 4.0, 7.5% 2-mercaptoethanol, and incubated 24 hr at 37°. Subsequently, they were washed extensively with 50 mM EDTA at 50°. For protein removal, plugs were incubated 24 hr at 50° in NDS buffer (0.4 M EDTA, 10 mM Tris-HCl, pH 9.5, 1% *N*-lauroylsarcosine) containing 2 mg/ml proteinase K. The plugs were washed as before and stored in 50 mM EDTA at 4°.

Clone libraries: A cosmid library of 16,800 individual clones with an average insert size of 34 kb in the vectors pMOcosX (Orbach 1994) and pLORIST6Xh (Kelkar *et al.* 2001) was obtained from the Fungal Genomic Stock Center (University of Kansas Medical Center, Kansas City, MO). For the construction of an additional cosmid library, the vector Lawrist-4 was used as previously described (Hanke and Hoheisel 1999). High-density clone filter grids (Hybond-N^+; Amersham, Braunschweig, Germany) were arrayed using a commercial robotic device (BioGrid; BioRobotics, Cambridge, UK); filter processing was as described in detail elsewhere (Hoheisel *et al.* 1995).

A BAC library with 15-fold coverage and an average insert size of 69 kb was constructed in cooperation with LION Bioscience AG (Heidelberg, Germany) using the pBeloBAC-Kan vector (Mozo *et al.* 1998). A total of 9216 individual clones were picked into microtiter dishes and arrayed on nylon filters.

Sublibrary selection: Specific cosmid sublibraries of linkage groups II and V were selected as described (Frohme *et al.* 1998). Intact chromosomal DNA (provided by Jonathan Arnold, University of Georgia, Athens, GA) was separated by PFGE, and the individual chromosomal fragments were isolated from the gel. About 70 ng of DNA of each fragment was directly used for labeling by random hexamer priming (Feinberg and Vogelstein 1983). Hybridization to the clone filters and detection of positive clones followed standard protocols (Hoheisel *et al.* 1995). Positive clones were rearrayed into new microtiter dishes. A BAC sublibrary of linkage group II was selected by hybridizing 190 cosmids to the BAC-library filters—in pools of up to six clones at a time—which themselves had been positive in hybridizations with chromosome II.

Mapping procedures: Cosmid DNA was isolated by alkaline lysis and phenol/chloroform extraction (Sambrook *et al.* 1989). One-third of each preparation was labeled by random hexamer priming (Feinberg and Vogelstein 1983) in the presence of digoxigenin-11-dUTP (Roche, Germany). BAC DNA was isolated using the QIAprep Spin Miniprep Kit (QIAGEN GmbH, Germany) as recommended by the manufacturer except for some modifications. About 7 ml of overnight culture was pelleted and resuspended in 250 μl P1 buffer supplemented with 200 μg/ml RNaseA. After adding 250 μl P2 buffer and incubation at room temperature for <5 min, 350 μl P3 buffer was added instead of the recommended N3 buffer. This proved to be more successful for BAC DNA isolation, since P3 contains no chaotropic salts as does N3, thus improving the elution efficiency of large DNA. After cell lysis, the DNA was precipitated with isopropanol and digested with 5 units *Eco*RI (New England Biolabs, Germany) for 1 hr at 37°. Five volumes of PB buffer were added to 1 volume of DNA solution, mixed, and applied to the column. Washing was as recommended. Elution of DNA was with 30 μl 10 mM Tris-HCl, pH 7.5, prewarmed to 65°. With this modified protocol, a sufficient amount of high quality BAC DNA was obtained with minimal effort. For introduction of a label, 22 μl of the eluate was subjected to random hexamer priming in the presence of digoxigenin-11-dUTP (Roche). Hybridization was done as reported (Hoheisel *et al.* 1995). Detection was by anti-digoxigenin antibody-alkaline phosphatase conjugate using the chemiluminescent alkaline substrate CSPD (Roche). Results were scored manually and analyzed with an established software package (Mott *et al.* 1993).

RESULTS

Although an older cosmid library of *N. crassa* existed and was available, a complementary library of four genome-equivalents was constructed to assure the colinearity of the earlier library with the DNA used in construction of the BAC library. Thus, 21,800 clones were used in the experiments, equivalent to ~17-fold coverage of the *N. crassa* genome. From this total, 2163 cosmids belonging to chromosome V were identified by hybridizing radioactively labeled chromosomal DNA to the clone arrays. This sublibrary statistically represented an 8.5-fold coverage, which is about half the redundancy of the global library. Therefore, it could be concluded that only one-half of the clones deriving from this chromosome were actually identified by this procedure, and gap closure had to be performed on the total genomic library at a later stage. Hybridization-based mapping of chromosome V started with the cosmid sublibrary only, since the BAC clones were not available at the time. Probe labeling was performed by incorporation of digoxigenin-11-dUTP and detection via a chemiluminescent substrate, since this procedure proved to be as sensitive as radioactive labeling. First, individual cosmid clones were picked at random and used as probes following an iterative process of sampling without replacement (Mott *et al.* 1993). Probes were picked at random from the ever decreasing number of library clones that were not positive in any prior hybridization and should, by definition, originate from still unmapped areas. By this strategy, the probes should also be spaced relatively evenly. This process was continued until essentially all clones had been positive at least once. Then, cosmid clones located at contig ends were chosen as probes for gap closure but, as expected, little improvement of the map was made using the sublibrary filters. Therefore, hybridizations of cosmids from contig ends were carried out on the global genomic cosmid and BAC libraries. This proved to be more successful in reducing the initially >100 contigs to a final number of 21 contigs. This number could not be reduced further, although eventually a total of 1192 cosmids and 82 BACs were used as probes, which is 55% more than anticipated. During about the last couple of hundred hybridization experiments, no new data was produced; only existing contigs were confirmed, indicating an exhausting mapping of the library. The resulting physical map is shown in Figure 1; probes that yielded redundant data were mostly omitted for clarity. The majority of the signals are located along the diagonal of the two-dimensional presentation of clones and probes, representing the best correlation between the ordered probes and the fitting clone order. The background of false positive hybridiza-

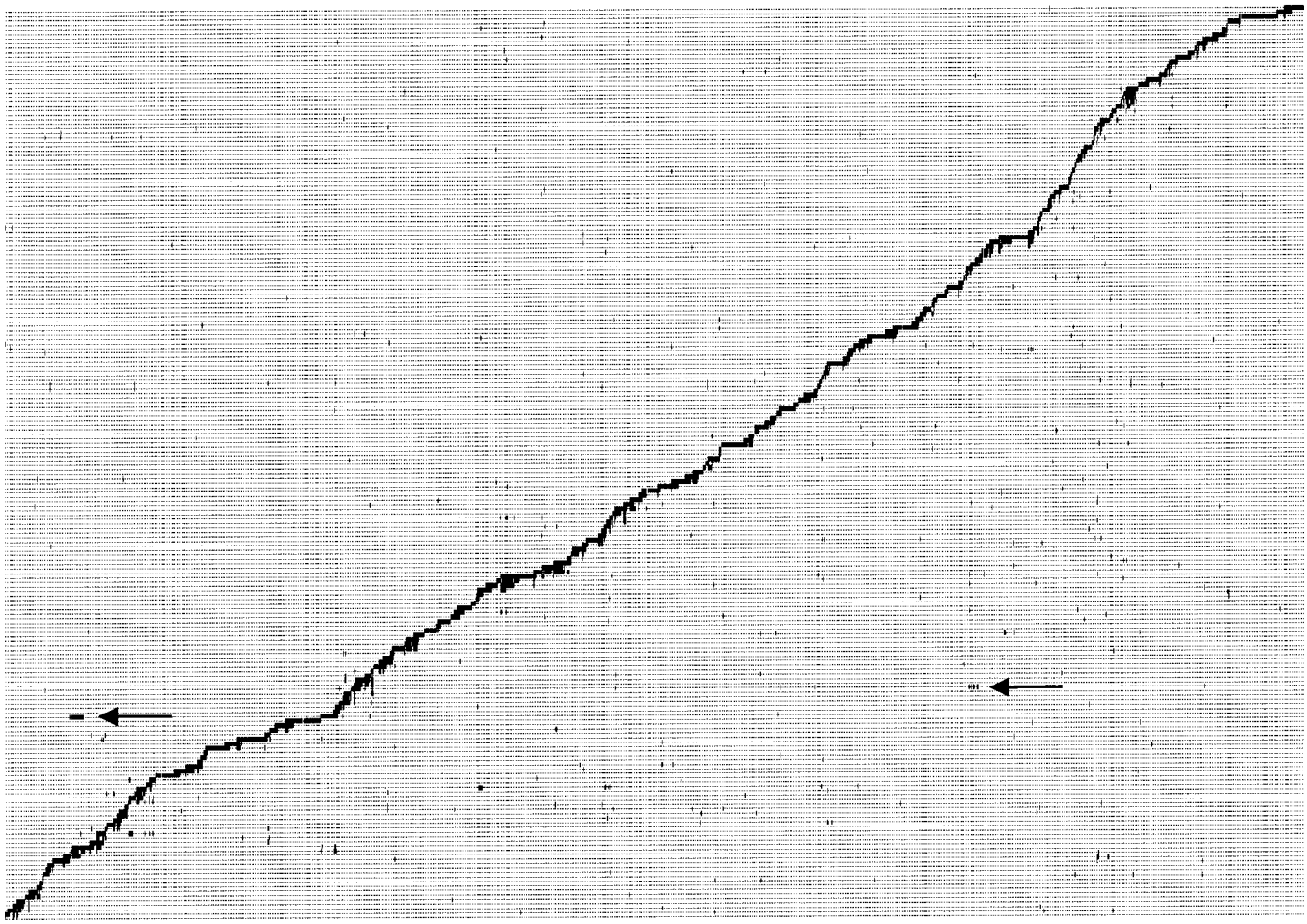

FIGURE 1.—Physical clone map of chromosome V. The hybridization results are shown in the form of a two-dimensional matrix. Probes correspond to rows and clones to columns. Positive hybridization is indicated by a solid bar at the respective intersection. A subset of 203 probes is shown covering the entire chromosome, which is represented by 2978 cosmid and BAC clones. The arrows mark two chimeric probes. Upon request, the authors will provide a detailed version of the map containing all the probes used in experimentation.

tion results was low with <3% of the scored signals being located outside of the diagonal.

For the mapping of chromosome II, a slightly different strategy was followed. A total of 1369 cosmid clones—equivalent to a 10-fold coverage—were identified in the global library by hybridizing the chromosomal band. Since hybridization of the chromosomal DNA to the BAC-library filter yielded only very weak signals, which were difficult to impossible to analyze, a BAC sublibrary was selected by hybridizing a random set of 190 of the chromosome-specific cosmid clones. This process resulted in 751 BAC clones, statistically an 11-fold coverage. Physical mapping was started in parallel on both the cosmid and BAC library, applying the same strategy reported above on chromosome V. Again, gap closure was performed on the two global libraries. Altogether, 191 BACs and 273 cosmids were used as probes, resulting in the map of 13 contigs presented in Figure 2; as before, probes that yielded redundant data were omitted for the purpose of presentation.

Close analysis of the hybridization data revealed that some 10% of the cosmid clones were probably chimeras, which for cosmids is an unusually high percentage. In other libraries made from genomic DNA of similar base composition and complexity (HOHEISEL *et al.* 1993; JOHNSTON *et al.* 1997), a rate of <1% was found. However, for the preparation of these libraries, partial digests of the genomic DNA had been used whose average fragment sizes were much bigger than the maximally possible insert size. Although this led to a sharp drop in the efficiency of packaging and thus transfection, it simultaneously reduced the probability that two linked genomic fragments would fit into a single phage head and thus could be transfected into a host cell (HOHEISEL *et al.* 1995). In the analysis, chimeric clones were identified by their specific hybridization pattern when being used as probes. They uniquely hybridized to clones that on the basis of all other, always redundant, results belonged to two different regions. While most chimeric clones were removed from the map, two typical ones

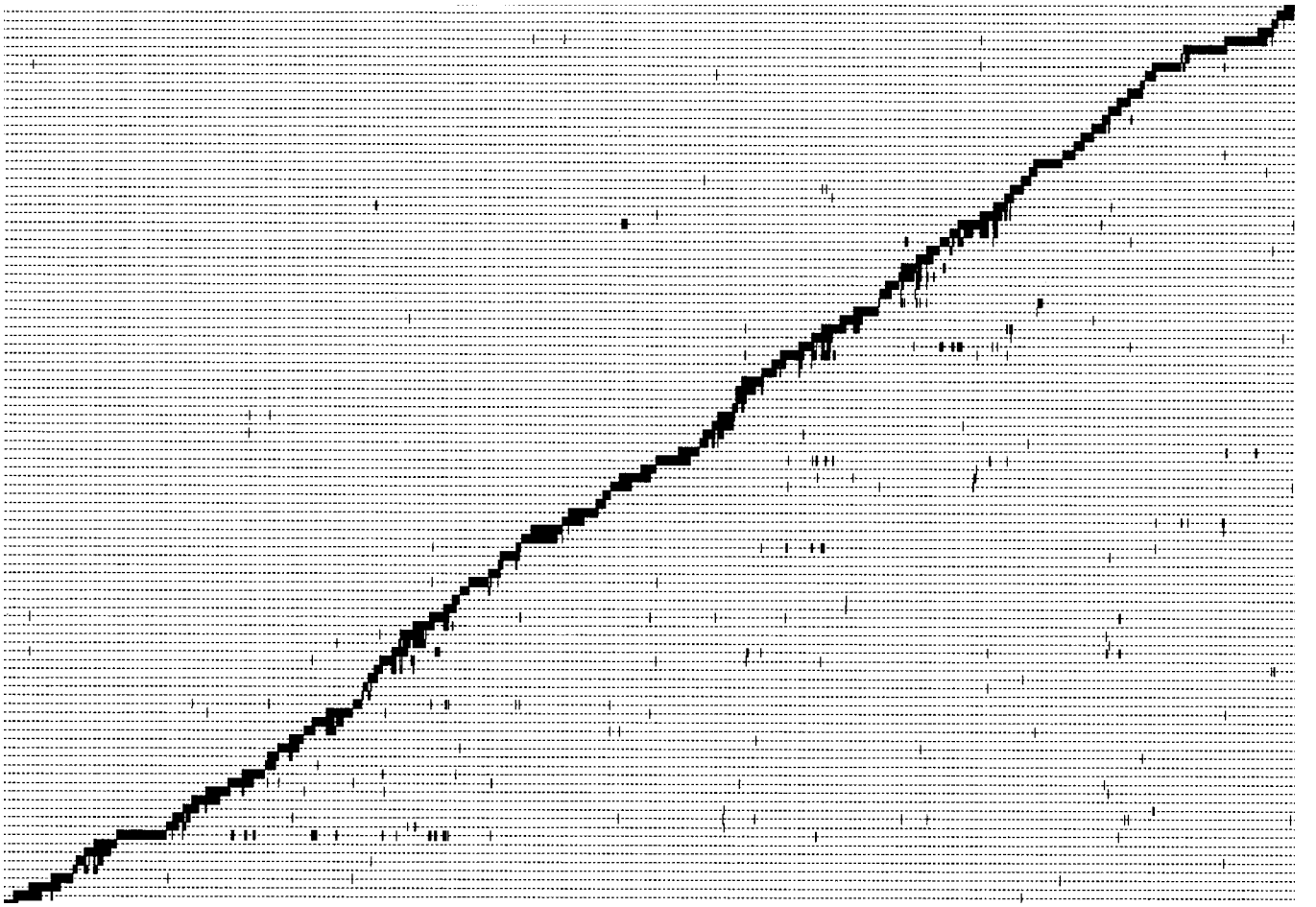

FIGURE 2.—Physical clone map of chromosome II. Map presentation is as described for Figure 1 but shows 104 probes and 2219 clones.

were left in the map of chromosome V (Figure 1) and are indicated by arrows. Because of the high redundancy of the information acquired during the mapping process and the fact that no apparently overlapping regions were accepted as such unless confirmed by more than two independent clones, it is unlikely that the map was significantly disturbed by chimeras.

As expected, there is no indication for the presence of large repetitive areas. They would be seen in the map as clusters of signals outside the main diagonal, similar to the signals that chimeric clones produce but more pronounced since occurring in more than a single clone. Few probes exhibit signals of that sort in numbers above the average background noise, but none is sufficiently confirmed by other probes to point with any significance to the presence of repeats.

DISCUSSION

Construction of the physical clone maps of chromosomes II and V of *N. crassa* was carried out by hybridization procedures. Earlier mapping projects had demonstrated that a 10-fold clone representation is usually sufficient for continuous coverage of a genome. Since for various reasons the actual coverage fluctuates strongly along a chromosome (HOHEISEL *et al.* 1995; FROHME *et al.* 2000), a lower overall redundancy could be insufficient for the generation of an unambiguous map in some regions. Employing more than one type of library for contig assembly should also increase the degree of uniformity in representation. Therefore, cosmid and BAC libraries of high coverage were used in this project. Not only did the cloning vectors differ, but the fragmentation of the insert DNA was performed with two different restriction endonucleases: *Mbo*I was used for construction of the cosmid libraries and *Eco*RI for the BAC library.

The mapping procedure started with the selection of chromosomal sublibraries, thereby increasing the efficiency of the mapping procedure. During the initial phase, only clones with a high probability of containing DNA of chromosomes II and V, respectively, were chosen. Gap closure was then performed on the global libraries. From practical experience (*e.g.*, HOHEISEL *et al.* 1993) as well as theoretical prediction (GRIGORIEV 1993), a probe number that is three times the clone

number of a minimal clone coverage should be sufficient for completing a map. Nevertheless, a large number of extra hybridizations—up to an eventual excess of 55% in the case of chromosome V—were carried out during the final stages in an attempt to bridge gaps. Still, the remaining ones could not be closed in either library type, and only redundant information was produced in the end. This and the fact that independently produced libraries of high clone coverage were analyzed suggest that the relevant genomic regions were actually missing in the libraries. As has been done before for other genomes (*e.g.*, Frohme *et al.* 2000), upon the availability of the final sequence, the gene content of these regions will be analyzed in an attempt to correlate potential gene function with the lack of clonability in large-insert vector systems of the *Escherichia coli* host.

The map is very likely to be accurate for most of the two chromosomes, since data redundancy is high and only a few false positive or negative hybridization events were recorded. Still, some very local disorder can be expected. In a few cases, individual clones could have been placed at homologous regions instead of their real map position because of cross-hybridization of the corresponding sequences. Mixing clone libraries of significant length difference—34 and 69 kb average insert size for the cosmids and BACs, respectively—led to a slight perturbation of the graphical representation of the maps. However, the data produced on the different clone types were studied separately as well as combined to exclude any potential local disturbance caused by this fact. Overall, a combined analysis is based on better statistics. Novel computational tools for physical mapping (*e.g.*, Heber *et al.* 2000; Bhandarkar *et al.* 2001, this issue) allow an evaluation of the map quality by algorithms different from the ones used in its production. Experimentally, there will be independent data originating from another mapping effort (http://gene.genetics.uga.edu). The ultimate quality test, however, will be a comparison to the final DNA sequence.

The maps described here are the basis of the ongoing German sequencing project (http://www.mips.biochem.mpg.de/proj/neurospora/). Because of this, a comparison to the genetic maps was not performed; this will be done upon completion of the sequence, since this will be more informative with respect to the presence and structural organization of markers. Although not all gaps in the maps could be bridged as yet, the number of base pairs sequenced to date clearly indicates that most of the two chromosomes are represented in the clone maps. Therefore, most stretches lacking DNA should be small enough to be sequenced on the basis of PCR fragments or sequence walking approaches. An exception could be the rDNA repeat region located on chromosome V. It consists of several individual units. A complete unit is generally difficult to impossible to clone (Johnston *et al.* 1997; Frohme *et al.* 2000), so the rDNA is likely to be missing except for peripheral parts (Kelkar *et al.* 2001). Also, the centromere could be absent, although two out of three centromeres of *Schizosaccharomyces pombe* were covered in a similar project (Hoheisel *et al.* 1993). The very ends of the telomeres are definitely missing from the map. This is not surprising, since the clone libraries used for mapping contained only genomic fragments that were isolated by restriction digestion of genomic DNA. At least, the DNA beyond the most telomeric restriction sites cannot therefore be present in the libraries. Sequencing these very telomeric ends of the chromosomes will require a special effort as demonstrated in other projects (Goffeau *et al.* 1996).

The technical assistance of Mareike Grees and help by Carmen Fischer is gratefully acknowledged. We thank Jonathan Arnold and James Griffith for providing material and information, and Berthold Fartmann (MWG Biotech) for helpful discussions. This work was funded by the Deutsche Forschungsgemeinschaft.

LITERATURE CITED

Beadle, G. W., and E. L. Tatum, 1941 Genetic control of biochemical reactions in Neurospora. Proc. Natl. Acad. Sci. USA **27:** 499–506.

Bennett, J. W., 1997 White paper: genomics for filamentous fungi. Fungal Genet. Biol. **21:** 3–7.

Bhandarkar, S. M., S. A. Machaka, S. S. Shete and R. N. Kota, 2001 Parallel computation of a maximum likelihood model for chromosome reconstruction. Genetics **157:** 1021–1043.

Feinberg, A., and B. Vogelstein, 1983 A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:** 6–13.

Frohme, M., J. Hanke, L. Åslund, U. Pettersson and J. D. Hoheisel, 1998 Selective generation of chromosomal cosmid libraries within the *Trypanosoma cruzi* genome project. Electrophoresis **19:** 478–481.

Frohme, M., A. A. Camargo, S. Heber, C. Czink, A. D. J. Simpson *et al.*, 2000 Mapping analysis of the *Xylella fastidiosa* genome. Nucleic Acids Res. **28:** 3100–3104.

Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon *et al.*, 1996 Life with 6000 genes. Science **274:** 546–567.

Grigoriev, A. V., 1993 Theoretical predictions and experimental observations of genomic mapping by anchoring random clones. Genomics **15:** 311–316.

Hanke, J., and J. D. Hoheisel, 1999 Construction and screening of cosmid libraries, pp. 405–414 in *The Nucleic Acid Protocols Handbook*, edited by R. R. Rapley. Humana Press, Totowa, NJ.

Hanke, J., M. Frohme, J. P. Laurent, J. Swindle and J. D. Hoheisel, 1998 Hybridization mapping of *Trypanosoma cruzi* chromosomes III and IV. Electrophoresis **19:** 482–485.

Heber, S., J. D. Hoheisel and M. Vingron, 2000 Application of bootstrap-techniques to physical mapping. Genomics **69:** 235–241.

Herrmann, B. G., and A. Frischauf, 1987 Isolation of genomic DNA. Methods Enzymol. **152:** 180–183.

Hoheisel, J. D., E. Maier, R. Mott, L. McCarthy, A. V. Grigoriev *et al.*, 1993 High resolution cosmid and P1 maps spanning the 14 Mb genome of the fission yeast *S. pombe*. Cell **73:** 109–120.

Hoheisel, J. D., E. Maier, R. Mott and H. Lehrach, 1995 Integrated genome mapping by hybridization techniques, pp. 319–346 in *Analysis of Non-mammalian Genomes: A Practical Guide*, edited by B. Birren and E. Lai. Academic Press, San Diego.

Johnston, M., L. Hillier, L. Riles, K. Albermann, B. Andre *et al.*, 1997 The complete sequence of *Saccharomyces cerevisiae* chromosome XII. Nature **387** (Suppl.): 87–90.

Kelkar, H. S., J. Griffith, M. E. Case, F. Covert, R. D. Hall *et al.*, 2001 The *Neurospora crassa* genome: cosmid libraries sorted by chromosome. Genetics **157:** 979–990.

Mott, R., A. Grigoriev, E. Maier, J. D. Hoheisel and H. Lehrach,

1993 Algorithms and software tools for ordering clone libraries: application to the mapping of the genome of *Schizosaccharomyces pombe*. Nucleic Acids Res. **21:** 1965–1974.

Mozo, T., S. Fischer, H. Shizuya and T. Altmann, 1998 Construction and characterization of the IGF Arabidopsis BAC library. Mol. Gen. Genet. **258:** 562–570.

Nelson, M. A., S. Kang, E. L. Braun, M. E. Crawford, P. L. Dolan *et al.*, 1997 Expressed sequences from conidial, mycelial, and sexual stages of *Neurospora crassa*. Fungal Genet. Biol. **21:** 348–363.

Orbach, M. J., 1994 A cosmid with a HyR marker for fungal library construction and screening. Gene **150:** 159–162.

Orbach, M. J., D. Vollrath, R. W. Davis and C. Yanofsky, 1988 An electrophoretic karyotype of *Neurospora crassa*. Mol. Cell. Biol. **8:** 1469–1473.

Perkins, D. D., 1992 *Neurospora crassa* genetic maps. Fungal Genet. Newsl. **39:** 61–70.

Perkins, D. D., A. Radford, D. Newmeyer and M. Bjorkman, 1982 Chromosomal loci of *Neurospora crassa*. Microbiol. Rev. **46:** 426–570.

Radford, A., and J. H. Parish, 1997 The genome and genes of *Neurospora crassa*. Fungal Genet. Biol. **21:** 258–266.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Communicating editor: J. Arnold