

Construction of genomic libraries from *Neurospora crassa* with the BigEasy™ v2.0 Linear Cloning System

Lanelle Connolly and Michael Freitag*

Department of Biochemistry and Biophysics
Center for Genome Research and Biocomputing
Oregon State University
Corvallis, OR 97331

* corresponding author

Introduction

The complete genome sequence of the filamentous fungus *Neurospora crassa* has been difficult to assemble, because a large amount of AT-rich “genomic dark matter” is lost from traditional shotgun, cosmid, and BAC libraries (1-3). This situation is typical of most repeat-rich eukaryotic genomes. Various novel direct sequencing approaches promise to reduce some of these errors. However, preliminary results from 454 pyrosequencing of the *Neurospora crassa* genome (1, 2) suggest that new sequencing methodology alone will likely not be sufficient to generate finished sequence of even a relatively small eukaryotic genome (~40 Mb), simply because assembly of AT-rich repeats proves to be a major obstacle.

Most of the ~2 Mb of missing *Neurospora* sequence maps to centromeric regions. Importantly, the sequence of all seven centromeres contain gaps. As in most other eukaryotes, *Neurospora* centromeric DNA is AT-rich and contains long segments of repeated DNA (Fig. 1). As expected from previous studies (4), we found no tandem repeats upon mapping the seven centromeres of *Neurospora*. Rather, there is an assembly of inactivated retrotransposon relics that have undergone “repeat-induced point mutation” (RIP), a process that induces numerous GC-to-AT mutations throughout the repeats (5). To understand centromere formation and maintenance, we wish to complete the assembly of centromeric DNA by alternative means.

The linear pJAZZ™-OC cloning vector in the BigEasy v2.0 Kit allows cloning and maintenance of long AT-rich fragments; direct and inverted repeats are also tolerated (6, 7). We took advantage of this system to generate medium- to long-insert genomic libraries of *N. crassa* that appear enriched for AT-rich DNA. Here we describe the construction and characterization of such a pilot library. Preliminary results suggest that we succeeded in cloning predominantly AT-rich inserts that will help to complete the assembly of centromeric sequences of *Neurospora crassa*.

Materials and Methods

Neurospora genomic DNA (Strain FGSC#2489) was digested for 1-3 hrs with Not I or Sma I, restriction endonucleases that recognize relatively rare GC-rich sites. An *in silico* digest of the genomic sequence of chromosome VII with Not I+Fse I+Asc I predicted predominantly long fragments in the AT-rich heterochromatin (18-189 kb; average ~48 kb). Shorter fragments were predicted in euchromatin (<15 kb).

To generate a pilot library of AT-enriched sequence, we separately purified genomic Sma I fragments of ~7-10 kb and 10-35 kb from agarose gels. We specifically avoided a strong <6 kb band, which was confirmed to contain most of the rDNA cluster. Gel-purified fragments were extracted by digestion with β-agarase (NEB, Ipswich, MA), and precipitated with isopropanol and ammonium acetate, according to standard procedures. The resulting blunt-ended fragments were directly ligated into the pJAZZ-OC vector. Ligation reactions contained 50 ng of vector plus 1.1 μg of 7-10 kb inserts or 1.8 μg of >10 kb inserts in a volume of 10 μl. Ligation was performed for 3-8 hours at room temperature. Reactions were heat-treated for 15 min at 70°C. For each transformation, 1 μl of the ligation was electroporated into 25 μl of BigEasy-TSA cells on a BTX 610 electroporator with standard *Escherichia coli* conditions (2.0 mm cuvettes, 25 μF, 200 Ohms, 2500 kV. Note that these conditions of ligation and transformation are not the optimal Lucigen protocol.

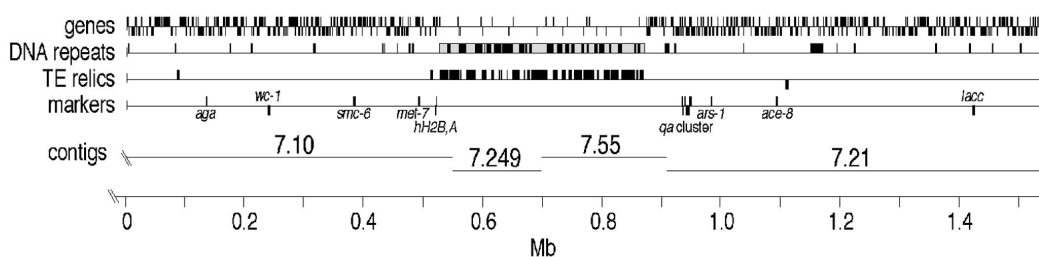


Figure 1: *Neurospora* centromeric DNA is composed of relics of transposable elements (TE). Centromeric DNA of linkage group VII is AT-rich, gene-poor and enriched in short non-tandem repeats of predominantly retrotransposon relics. Unclassified DNA repeats are indicated. Supercontigs with centromeric sequence are 7.10, 7.249 and 7.55.

Lucigen Recovery Medium (1 ml) was added to the cuvettes immediately after electroporation. Cells were mixed by pipetting gently up and down, transferred into a 15 x 100 mm culture tube, and incubated for 45 min at 37°C at 250 rpm. Substituting LB broth for Recovery Medium yielded ~100-fold fewer colonies after transformation. Aliquots of ~100 µl of the transformation mix were plated on YT+CXI plates, as described in the BigEasy™ v2.0 manual. After overnight incubation, clones were picked into LB broth amended with 12.5 µg/ml chloramphenicol and 0.01% (w/v) arabinose.

A master library was generated in 96-well plates from 100 µl of individual cultures mixed with 100 µl of 50% glycerol. The remainder of the culture liquid was used to generate plate, row, and column pools that allow fast identification of specific plasmids (to be described in detail elsewhere). Plasmid DNA was purified by alkaline lysis using standard methods (8) for the plate, row, and column pools. Screening the three DNA pools by either PCR or Southern hybridization is expected to rapidly reveal individual clones that contain previously identified centromeric DNA at the terminus of known supercontigs. The total number of ~5150 clones with an expected median insert size of ~20 kb represents ~2.5 x coverage and a ~95% likelihood of including one complete genome; this estimate is not adjusted for the slight AT enrichment achieved. Pools were screened by PCR with gene-specific primers that map near supercontigs on linkage group VII and with the rDNA primers ITS1 and ITS4-F (9). Twenty-eight individual clones (16 medium-size and 12 large insert clones) were randomly picked from the library and grown individually to assess insert size. Eleven clones with different ranges of inserts were end-sequenced. In eight cases, we were able to unambiguously identify their matched end sequences as previously assembled regions of the *Neurospora* genome by BLASTn (see Table 1, next page).

Results and Discussion

In pilot experiments to generate a genomic DNA library from *N. crassa*, we isolated a total of 1340 clones in three separate transformation experiments. In each case, we plated 100 µl aliquots of the transformation mix onto ten plates and obtained on average 180 colonies per plate. Control plates containing self-ligated pJAZZ-OC vector showed on average five “blue” colonies (i.e., uncut vector). Blue colonies on the experimental plates were rare (on average three per plate). Tests of individual clones revealed that ~85% of clones contained inserts of the expected size range (Fig. 2 a and b). Some clones appeared “empty”, i.e., they show the expected size bands for the pJAZZ-OC plasmid but no additional bands (Fig. 2a, lanes 5 and 12, and Fig. 2b, lanes 3 and 6). Tests of DNA pools revealed that seven out of eight pools contained a mix of inserts, as expected. Even pools with low yields of DNA yielded bands after PCR screening (data not shown). Digestion with restriction endonucleases that recognize AT-rich sites, i.e., MseI (5'-TTAA-3') or SspI (5'-AATATT-3'), revealed smaller fragments than those generated from inserts in standard cosmid vectors (data not shown). This result suggests that the pJAZZ-OC clones were at least slightly enriched for higher AT-content.

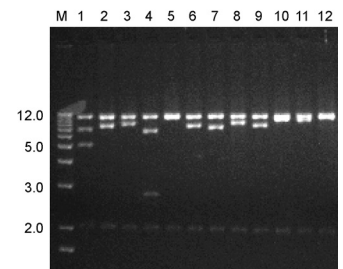
We further tested this pilot library by PCR with 12 individual regions from contig ends of *Neurospora* linkage group VII. In a pool consisting of DNA isolated from all library clones, one positive clone was identified by amplification of the NCU08188.3 locus, which is located next to an AT-rich repeat region at the distal end of contig 52. Based on the estimated genome coverage represented by 1340 clones with ~15 kb inserts, we expected to recover 2-4 positive clones in this experiment. As we assemble a more complete library, we expect to recover more “edge clones”.

We also began verification of the pilot library by amplification of the internal transcribed spacer region (ITS) of the rDNA. The *N. crassa* rDNA cluster is located on the left end of linkage group V and encompasses 100 to 150 direct repeats of a ~7.5 kb region. Incomplete translocated rDNA pseudogenes are also found throughout the genome (3). In combination, rDNA makes up as much as 3% of the *Neurospora* genome. We tested for the presence of the ITS region in 72 column and 24 row pools and in 14 plates of all 1340 clones. By this PCR approach, we were able to unambiguously identify four clones to individual wells (data not shown). In total, out of 1340 clones we identified 10-12 with the ITS region (i.e., <1% of all DNA cloned in the pilot library), which suggests that rDNA is under-represented in our library, a goal we had hoped to achieve.

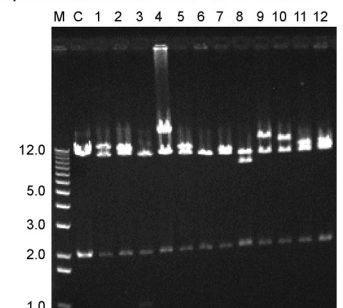
We sequenced the insert ends of several clones of varying size (Fig. 2a, lanes 1, 2, 3, 8, 10 and Fig. 2b, lanes 1, 2, 4, 9, 10, 11). Eight of eleven clones mapped to previously assembled regions of the *Neurospora* genome, including four regions that contain AT-rich segments. For all eight clones the expected insert sizes (see Fig. 2) closely matched those predicted from BLASTn searches with the sequenced insert ends (Table 1).

In summary, preliminary data suggest that our approach for isolation of AT-rich centromeric DNA will prove successful. The pilot library appears slightly enriched for AT-rich DNA, while rDNA appears under-represented. We believe that the BigEasy v2.0 cloning approach will prove useful for similar cloning projects for target DNAs with special structure, even under suboptimal conditions.

a) Individual clones with 7-10 kb inserts



b) Individual clones with >10 kb inserts



c) Pools of plasmid DNA for PCR

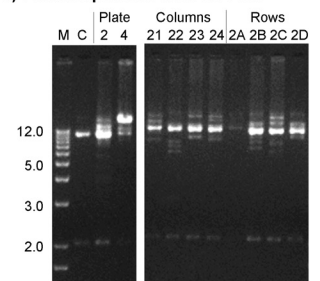


Figure 2. Verification of medium- and large-size inserts in a pilot library of *Neurospora crassa* genomic DNA integrated into pJAZZ-OC.

a) Individual clones containing 7 to 10 kb inserts. In total, 14/16 tested clones (12 shown here) showed inserts of the expected size. Lanes 5 and 12 contain DNA from “empty” vectors, which showed up as white transformants. **b)** Individual clones containing inserts larger than 10 kb. In total 10/12 tested clones showed inserts of the anticipated size. **c)** DNA obtained from alkaline lysis DNA extractions of pooled cultures. Plate pools contain 96 different clones, column and row pools contain 72 different clones. DNA was digested with Not I to release the complete inserts. The 12 kb and 2 kb bands correspond to the expected pJAZZ-OC fragments. DNA marker lanes are shown (M) and in b) and c) purified pJAZZ-OC was run as a control (C). DNA size standards are indicated in kilobases.

Table 1. End-sequencing of individual clones shows expected insert size ranges. Most randomly picked plasmids mapped to previously assembled regions of the *Neurospora* genome.

Clone	LG	Contig	Start	Stop	Insert
2-1*	V	15	137200	ND	ND
2-2	VII	52	119540	127176	7636
2-3	V	14	327696	335774	8078
2-8	I	6	1061681	1070007	8326
2-10	III	1	776434	786071	9637
3-1#	III	17	431425	ND	ND
3-2	IV	20	449255	460743	11488
3-4	I	3	572369	600629	28260
3-9	III	17	147518	163795	16277
3-10*	I	65	63775	ND	ND
3-11	I	39	272073	284311	12238

Notes: One plasmid (#) likely contains a chimeric insert, as one end sequence matches a genomic region that is very well assembled, but the other end contains previously uncloned sequence. Two plasmid ends (*) yielded poor sequence in one-pass sequencing. In these three cases the complete insert size was not determined (ND).

References

- Du, L., Alvarez, P., Desany, B., et al. (2005). Genome Sequencing of *Neurospora crassa* using 454 Sequencing Technology. Available at <http://www.454.com/news-events/publications.asp>.
- Desany, B., Alvarez, P., Rounsley, S., et al. (2005). Assembly of *Neurospora crassa* Genome Using 454 Sequencing Technology and New Assembler. Available at <http://www.454.com/news-events/publications.asp>.
- Galagan, J. E., Calvo, S. E., Borkovich, K. A., et al. (2003). The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422**, 859.
- Cambareri, E. B., Aisner, R., and Carbon, J. (1998). Structure of the chromosome VII centromere region in *Neurospora crassa*: degenerate transposons and simple repeats. *Mol. Cell Biol.* **18**, 5465.
- Selker, E. U. (1990). Premeiotic instability of repeated sequences in *Neurospora crassa*. *Annu. Rev. Genet.* **24**, 579.
- Ravin, N. V., and Ravin, V. K. (1999). Use of a linear multicopy vector based on the mini-replicon of temperate coliphage N15 for cloning DNA with abnormal secondary structures. *Nucleic Acids Res.* **27**, e13.
- Godiska, R., Patterson, M., Schoenfeld, T., and Mead, D. A. (2005). Beyond pUC: Vectors for cloning unstable DNA. In J. Kieleczawa (ed.), *DNA Sequencing: Optimizing the process and analysis*. Jones and Bartlett Publ., Sudbury, MA.
- Birnboim, H. C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**, 1513.
- Gardes, M. and Bruns, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. *Mol. Ecol.* **2**, 113.

ORDER INFORMATION

The BigEasy™ v2.0 Linear Cloning Kit includes: Dephosphorylated pJAZZ™-OC Vector pre-cut at either a SmaI (blunt) or NotI site, CloneSmart® DNA Ligase, CloneDirect™ 10X Ligation Buffer (includes ATP), DNATerminator® End Repair Enzyme & 5X End Repair Buffer (Blunt Kit only), Sequencing Primers, Positive Control Insert DNA, BigEasy-TSA™ Electrocompetent Cells in SOLO packaging (1 transformation per tube), Recovery Medium, Transformation Control DNA, and complete protocols. BigEasy-TSA Electrocompetent Cells are also available separately.

Products	Size	Cat. No.
BigEasy v2.0 Linear Cloning Kit (pJAZZ-OC Blunt Vector)* w/BigEasy-TSA Electrocompetent Cells (SOLOs)	5 reactions	43018-1
	10 reactions	43018-2
	20 reactions	43018-3
BigEasy v2.0 Linear Cloning Kit (pJAZZ-OC NotI Vector) w/BigEasy-TSA Electrocompetent Cells (SOLOs)	5 reactions	43024-1
	10 reactions	43024-2
	20 reactions	43024-3
BigEasy-TSA Electrocompetent Cells (≥ 4 x 10 ¹⁰ cfu/μg) (SOLOs)	6 reactions	60224-1
	12 reactions	60224-2
	24 reactions	60224-3

* The pJAZZ-OC vector is selected using chloramphenicol. An analogous BigEasy Kit using kanamycin selection with the pJAZZ-OK vector will be available soon.