EUKARYOTIC CELL, June 2011, p. 724–733 Vol. 10, No. 6
1535-9778/11/$12.00 doi:10.1128/EC.00016-11

Bulk Segregant Analysis Followed by High-Throughput Sequencing Reveals the *Neurospora* Cell Cycle Gene, *ndc-1*, To Be Allelic with the Gene for Ornithine Decarboxylase, *spe-1*▽†

Kyle R. Pomraning,[2] Kristina M. Smith,[1] and Michael Freitag[1]*

Department of Biochemistry and Biophysics[1] and Program for Molecular and Cellular Biology,[2] Center for Genome Research and Biocomputing, Oregon State University, Corvallis, Oregon 97331

Received 31 January 2011/Accepted 14 April 2011

With the advent of high-throughput DNA sequencing, it is now straightforward and inexpensive to generate high-density small nucleotide polymorphism (SNP) maps. Here we combined high-throughput sequencing with bulk segregant analysis to expedite mutation mapping. The general map location of a mutation can be identified by a single backcross to a strain enriched in SNPs compared to a standard wild-type strain. Bulk segregant analysis simultaneously increases the likelihood of determining the precise nature of the mutation. We present here a high-density SNP map between *Neurospora crassa* Mauriceville-1-c (FGSC2225) and OR74A (FGSC2489), the strains most typically used by *Neurospora* researchers to carry out mapping crosses. We further have demonstrated the utility of the Mauriceville sequence and our approach by mapping the mutation responsible for the only existing temperature-sensitive (ts) cell cycle mutation in *Neurospora*, *nuclear division cycle-1* (*ndc-1*). The single T-to-C point mutation maps to the gene encoding ornithine decarboxylase (ODC), *spe-1* (NCU01271), and changes a Phe to a Ser residue within a highly conserved motif next to the catalytic site of the enzyme. By growth on spermidine and complementation with a wild-type *spe-1* gene, we showed that the defect in *spe-1* is responsible for the ts *ndc-1* mutation. Based on our results, we propose changing *ndc-1* to *spe-1*^{ndc}, which reflects that this mutation results in an ODC with a specific nuclear division defect.

Single nucleotide polymorphism (SNP) maps between organisms with different genetic backgrounds are useful for identifying point mutations that result in an observable phenotype. By mating a mutant with a strain of a different genetic background and then selecting progeny with and without a phenotype, one is able to map the underlying mutation(s) by bulk segregant analysis (29). After undergoing recombination, each progeny exhibiting a specific phenotype will have the same genetic background as the mutant parent in the vicinity of the mutation but a mixture of DNA inherited from either parent in regions of the genome that are unlinked to the mutation. By analyzing DNA from a mixture of phenotypically mutant progeny from the cross, it is possible to define the genetic region containing the mutation and even to identify the mutation itself. This is what is meant by "bulk segregant analysis."

Most recently, molecular mapping strategies based on restriction fragment length polymorphisms (RFLP) (28), cleaved amplified polymorphic sequences (CAPS) (20, 22), or microarray-based restriction-site-associated DNA mapping (RAD mapping) (1, 23) have been used to map mutations in *Neurospora crassa*. These methods can be time-consuming and laborious and therefore expensive. In addition, once the mutation is mapped to a specific region of the genome, additional primer pairs for CAPS mapping need to be ordered and additional SNPs identified to more precisely map the mutation. Typically, researchers will then attempt complementation of phenotypes by candidate genes. Once (or if) complementation is successful, the mutated allele of the candidate gene still needs to be sequenced by traditional Sanger sequencing. In contrast, methods based on microarrays (5) or high-throughput DNA sequencing (3, 4, 11, 44) make it possible to quickly generate high-density SNP maps that allow direct identification of specific point mutations or indels by sequence analysis. Moreover, this approach does not rely on a specific set of reference strains.

We tested mapping by high-throughput Illumina sequencing in studies with the filamentous fungus *N. crassa*, one of the most widely used genetic model organisms. The two most commonly used genetic backgrounds of *Neurospora* for molecular mapping purposes are OR74A (FGSC2489) (6, 10, 14) and Mauriceville-1-c (FGSC2225) (2), hereafter called OR and MV. The OR strain was used for the *N. crassa* genome sequencing project (14). The genome has been assembled into seven chromosomes, represented by supercontigs one to seven. There are 13 additional supercontigs (http://www.broadinstitute.org/annotation/genome/neurospora/Regions.html) that mostly contain ribosomal DNA (rDNA) and AT-rich regions mutated by repeat-induced point mutation (RIP) (K. M. Smith and M. Freitag, unpublished data). We have now sequenced and partially assembled the *N. crassa* MV genome to improve upon the available SNP maps between the OR and MV strains (22, 23). As proof of principle for mutation mapping by high-throughput sequencing, we sequenced DNA from mutant progeny resulting from a cross between MV and the classic *N. crassa ndc-1* (*nuclear division cycle-1*; FGSC3441) mutant. We identified one specific candidate SNP that likely

* Corresponding author. Mailing address: Department of Biochemistry and Biophysics, 2011 ALS Bldg., Oregon State University, Corvallis, OR 97331-7305. Phone: (541) 737-4845. Fax: (541) 737-0481. E-mail: freitagm@cgrb.oregonstate.edu.

† Supplemental material for this article may be found at http://ec.asm.org/.

▽ Published ahead of print on 22 April 2011.

caused the temperature-sensitive (ts) *ndc-1* defect, which presents as arrest just prior to DNA synthesis during the nuclear division cycle (38). This single T-to-C point mutation maps to *spe-1* (NCU01271), the gene encoding ornithine decarboxylase (ODC) (27, 34), and changes a Phe to a Ser residue within a highly conserved motif near the ODC catalytic site. We were able to complement the *ndc-1* phenotypes by growth on spermidine and transformation of the mutant with a wild-type *spe-1* gene.

MATERIALS AND METHODS

Strains, crosses, and growth conditions. The MV genome sequence was produced from *N. crassa* Mauriceville-1-c (FGSC2225; NMF37) (2). Strains were grown under standard conditions (10) in Vogel's minimal medium (VMM) with 1.5% sucrose at 32°C, except for ts *ndc-1* strains, which were grown at 22°C, and *inl* strains, which were supplemented with 50 mg/liter of *myo*-inositol. The mapping cross was inoculated with *N. crassa* MV as the recipient strain on synthetic crossing medium supplemented with 0.5% sucrose and 50 mg/liter *myo*-inositol. Conidia from the *ndc-1* mutant (FGSC3441; NMF164) (38) were added 3 days later and spread across the recipient culture. After 2 weeks, random ascospores were recovered from the cross and heat shocked at 65°C for 60 min (standard heat shock) or 60°C for 30 min (gentle heat shock) on VMM with 50 mg/liter *myo*-inositol (VMMI) and FGS (0.5 g/liter fructose, 0.5 g/liter glucose, and 20 g/liter sorbose) as the carbon source. After overnight growth at room temperature, 200 viable progeny that were treated according to each of the heat shock protocols were collected and incubated in slants with VMMI with 1.5% sucrose. The progeny were tested for *inl* by spotting conidia on VMM or VMMI with FGS, and they were tested for the ts *ndc-1* allele by spotting conidia on VMMI with FGS followed by growth at 22°C or 37°C. Complementation of the *ndc-1* mutant was done on VMMI with 1.5% sucrose, and strains were grown at 22°C or 37°C with or without 1 mM spermidine trihydrochloride.

DNA sequencing. Genomic DNA was extracted from *N. crassa* tissue as previously described (35). Genomic DNA from MV and different progeny from the MV × *ndc-1 inl* cross was quantified using a NanoDrop2000 spectrophotometer and diluted to 100 ng/μl with Tris-EDTA (TE) buffer. DNA from progeny was mixed in two pools, the first containing equal amounts of DNA from 54 progeny (NMF343 to NMF396) with both the *inl* and *ndc-1* mutations, the second containing equal amounts of DNA from 9 progeny (NMF334 to NMF342) with only the *inl* mutation. The pooled DNA was placed in a 4°C circulating water bath and sheared into 200- to 1,000-bp fragments in a Bioruptor (Diagenode, Denville, NJ) set on "high" for 18 min (30-s on/off cycles). MV DNA was sheared with a microtip-equipped Branson (Danbury, CT) S450A sonicator by seven successive 10-s pulses (output, 1.2; duty cycle, 80), followed by a 30-s rest on ice after each cycle. Illumina sequencing libraries were constructed by ligating paired-end adaptors (see Table S1 in the supplemental material) to the sheared DNA and amplifying the library essentially according to our previously published protocol (35); the latest version of the protocol is available upon request. Paired-end 80-nucleotide (nt) or single-end 40-nt reads were generated on an Illumina GAII sequencing system at the Oregon State University Center for Genome Research and Biocomputing (OSU CGRB) for the MV and pooled strains, respectively. DNA sequences derived from MV, the *ndc-1 inl* strain, and the *inl* pools are available upon request.

SNP analysis and genome assembly. The 80-nt *N. crassa* MV genomic sequence reads were trimmed to 63 nt and converted from the Solexa FASTQ to the Sanger FASTQ quality score format using the "Mapping and Assembly with Quality" software program (MAQ, v0.7.1) (24). The processed reads were aligned to assembly 10 of the *N. crassa* OR reference genome (14) with the MAQ easyrun command (24). Called SNPs were quality filtered with conservative settings to reduce the likelihood of calling false-positive SNPs. For this, only SNPs called from reads mapped to a single site in the genome with a minimum phred-like quality score of 42, which requires at least 5× coverage, and N as the second-best call were retained. For genome assembly, the 80-nt MV reads were trimmed to 76 nt to remove the 4-nt GGGT barcode that was used for Illumina sequencing. Reads were then assembled *de novo* into contigs with the Velvet v1.0.12 software program (46). Contig alignments were performed using NUCmer, a part of the MUMmer v3.0 software program (21).

SNP validation. SNPs were mapped as perfect matches to a data set of 248,997 *N. crassa* MV expressed sequence tags (ESTs) (22; M. Basturkmen, J. Xu, M. Shi, J. Loros, M. Nelson, M. Henn, C. Kodira, N. Lennon, L. Green, J. Galagan, B. Birren, J. Dunlap, and M. S. Sachs, unpublished data) of 71-nt reads (35 nt flanking both versions of each SNP that was farther than 35 nt from the nearest called SNP) using the Hashmatch software program (12). For validation, 47 SNPs that lie in a TaqI restriction endonuclease cleavage site in the OR background but not in the MV background were selected and amplified by PCR from the OR and MV strains with primer pairs obtained from the *Neurospora* Functional Genomics Project (see Table S1 in the supplemental material). PCR products were digested with TaqI and RFLP patterns analyzed by agarose gel electrophoresis.

Mutant mapping by high-throughput sequencing. The "SNPome," i.e., all the SNPs, between two *N. crassa* strains was generated by running the Perl script "SNPome_builder" (available upon request). The output SNPome consists of a FASTA-formatted file that contains both versions of each SNP and the reference genome coordinates. SNPome_builder extracts the genomic sequence around each SNP site from the reference genome and makes a FASTA-formatted sequence for each version of the SNPs. When called SNPs are closer than the read length, sequences are concatenated (Fig. 1). Processed 63-nt Illumina reads were mapped to the SNPome as perfect matches with Hashmatch (12). A second Perl script, "SNP_profiler," was run on the Hashmatch output to count the number of reads hitting each version of each SNP. This script then calculates a ratio of reads matching each version of a SNP as a fraction between 0 and 1 to indicate which parental genome is the likely origin of the underlying DNA sequence. SNP profiles were viewed in the Argo genome browser (R. Engels; www.broadinstitute.org/annotation/argo2/) to identify contiguous genomic regions in which SNPs originated from a specific parental strain.

MAQ (24) was used to identify SNPs between mutant and nonmutant strains in the regions identified by bulk segregant analysis. Exclusion of spurious SNPs as candidates for point mutations was performed by aligning reads from genomic sequences from FGSC322 (17) and FGSC3562 (42) to the OR assembly with MAQ (24) and subtracting SNPs found in these *ndc-1*$^{+}$ *inl*$^{+}$ strains from SNPs found in mutant progeny. The SNPs remaining that had phred-like quality scores of ≥42 were aligned to the OR genome in the GFF3 format and visually scanned in the Argo browser for specific regions that generated nonsynonymous mutations in predicted or known coding regions.

Sequencing of the *ndc-1* allele. To verify that *ndc-1* is an allele of *spe-1*, partial DNA sequences of the *spe-1*^{ndc} allele were obtained by traditional Sanger sequencing from three *ndc-1* strains (NMF344, NMF345, and NMF346) and an MV sibling (NMF398). We designed primers within the *spe-1* coding region (OMF2243 and OMF2244; see Table S1 in the supplemental material), amplified this region from the strains in three independent reactions, and carried out DNA sequencing at the CGRB core facility using the primer OMF2243.

Complementation of *ndc-1*. The *ndc-1* mutant (FGSC3441; NMF164) was transformed by electroporation with pKP19, a derivative of pGS1, which contains most of the *spe-1* region (45). pKP19 carries the *hph* gene, which can confer resistance to hygromycin, integrated between the BamHI and EcoRV sites of pGS1. Transformants were obtained on VMMI plus 100 μg/ml hygromycin B at 22°C. Heterokaryotic transformants were selected and grown at 37°C to score for complementation of the *ndc-1* phenotype. To verify transformation and complementation, DNA from around the mutated site of *spe-1* was amplified by PCR in three independent reactions using a mixture of three primers (OMF2244, NCU1271.5F, and NCU1271.3R; see Table S1 in the supplemental material) from the *ndc-1* mutant (FGSC3441; NMF164) and a heterokaryotic transformant (NMF401). The reactions from each strain were pooled and purified by gel electrophoresis using a Qiaquick gel extraction kit (Qiagen, Valencia, CA) and sequenced with the primer OMF2244 (see Table S1).

RESULTS

Mauriceville genome sequencing. The *N. crassa* MV genome was sequenced as paired-end 80-nt reads on an Illumina GAII system. In total, 9,384,230 reads were obtained, resulting in a nominal coverage of ~17-fold across the ~41-Mb *Neurospora* genome. Adapter-trimmed 76-mer reads were assembled *de novo* into 10,435 *N. crassa* MV contigs with Velvet software (46), which resulted in ~80% (34,698,139 nt) coverage of the currently assembled *Neurospora* genome, with a minimum contig length of 200 nt and an n50 of 50,295 nt. MV contigs were aligned to the OR reference genome with the NUCmer software program (21), and the alignment was visualized as a percent similarity plot (Fig. 2, bottom panel). The majority of the contigs are >95% identical to the OR reference genome,

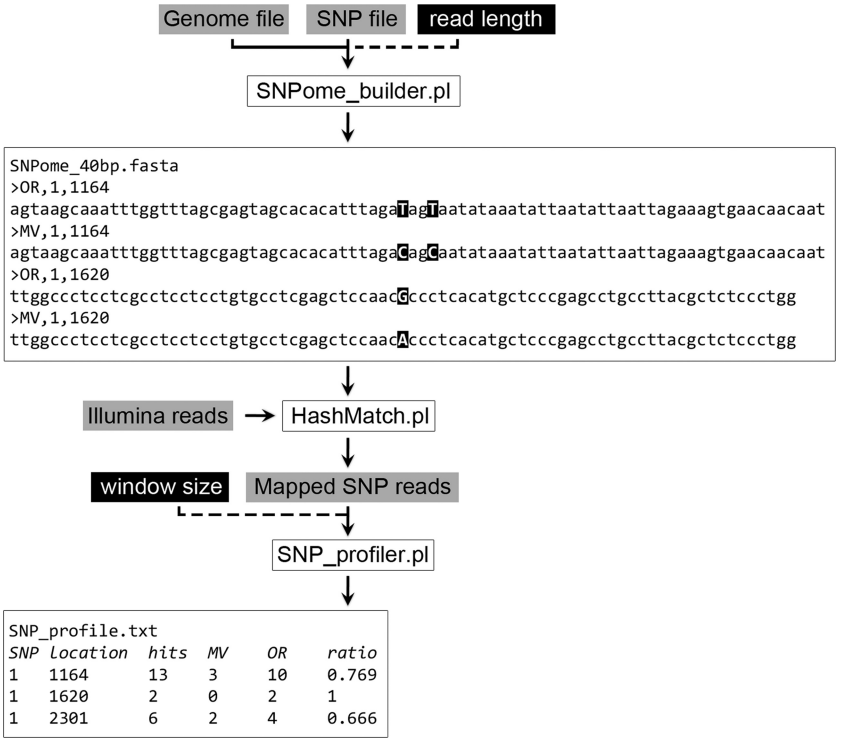

FIG. 1. Pipeline to generate SNPomes and profile SNPs in bulk segregants. The Perl script SNPome_builder creates a SNPome file (shown here for two example SNPs) given a FASTA-formatted genome file, SNP position file, and user-specified read length. SNPs within the SNPome file between OR and MV are highlighted in capital letters. Hashmatch (12) was used to map Illumina high-throughput sequencing reads, the output of which was analyzed by the Perl script SNP_profiler using a 5-kb window size. The output of SNP_profiler indicates the chromosome and position of each SNP, the number of times each version of the SNP was sequenced, and the ratio of the versions in a tab-delimited file.

with the highest similarity occurring in genic regions. In contrast, contigs aligning to the centromere and telomere regions are shortest, as we expected for *de novo* assembly of repetitive DNA. These contigs show the least similarity to OR and are coincident with regions where no MV reads matched the OR reference genome (Fig. 2). Whole-genome alignments showed that the majority of the OR sequence absent from MV is confined to AT-rich regions. This lack of reads is not caused simply by poor assembly or low coverage of AT-rich MV contigs. Rather, we predicted that many short AT-rich reads from MV are sufficiently different to not match the AT-rich OR regions. To lend support to this idea, we examined repetitive regions previously found by Southern analyses to be present in OR but absent in MV (methylated segments 8:A6, 8:F10, and 8:G3) (37). Here we show that we found no reads in these regions (Fig. 3).

SNP analysis of MV versus OR. It is of particular interest to the *Neurospora* community to identify SNPs between OR and MV because these are the most commonly used strains for molecular mapping. Reasonably detailed RFLP maps (31) and primer pairs for CAPS mapping (20, 22) have been generated. Nevertheless, even with these techniques, several iterations of PCR to obtain fragments covering SNPs are required to map mutations into a small region that lends itself to complementation analyses. Moreover, currently known MV SNPs were found by EST sequencing (22; M. Basturkmen, J. Xu, M. Shi, J. Loros, M. Nelson, M. Henn, C. Kodira, N. Lennon, L. Green, J. Galagan, B. Birren, J. Dunlap, and M. S. Sachs, unpublished data), thus covering only the most conserved regions of the MV and OR genomes. One of our goals was to assemble a more complete SNP map of MV to facilitate mapping of both novel and classical mutations.

We identified 168,579 SNPs between the OR reference genome and MV genomic reads after quality-filtering SNP calls from MAQ alignments (24), resulting in an average SNP density of 1 SNP per 243 nt. Of these, 64,066 (38.0%) occur in the coding region of genes, leaving the remaining 104,513 in intergenic space. Of the coding mutations, 19,898 (31.1%) cause nonsynonymous changes. SNPs predicted to cause changes in protein sequence are presented in Table S2 in the supplemental material. The SNPs are spaced relatively homogeneously along the coding regions, with somewhat decreased SNP density around the centromeres, the pericentric regions (Fig. 4A). As mentioned above, few SNPs within centromeric core regions are detected, because MV centromeric DNA is different from OR centromeric DNA (Fig. 1 and 4A). The complete set of SNPs is attached here as Table S3 in the supplemental material and is also available on the *Neurospora* homepage (http://www.fgsc.net/neurospora/SNPs/SNP_map.htm).

For validation, SNPs were aligned as either the OR or MV version to a data set containing 248,997 MV ESTs obtained by traditional Sanger sequencing (22; M. Basturkmen, J. Xu, M. Shi, J. Loros, M. Nelson, M. Henn, C. Kodira, N. Lennon, L. Green, J. Galagan, B. Birren, J. Dunlap, and M. S. Sachs, unpublished data). With a requirement of perfect alignment, 29,612 of the SNP calls aligned to the ESTs. As expected,

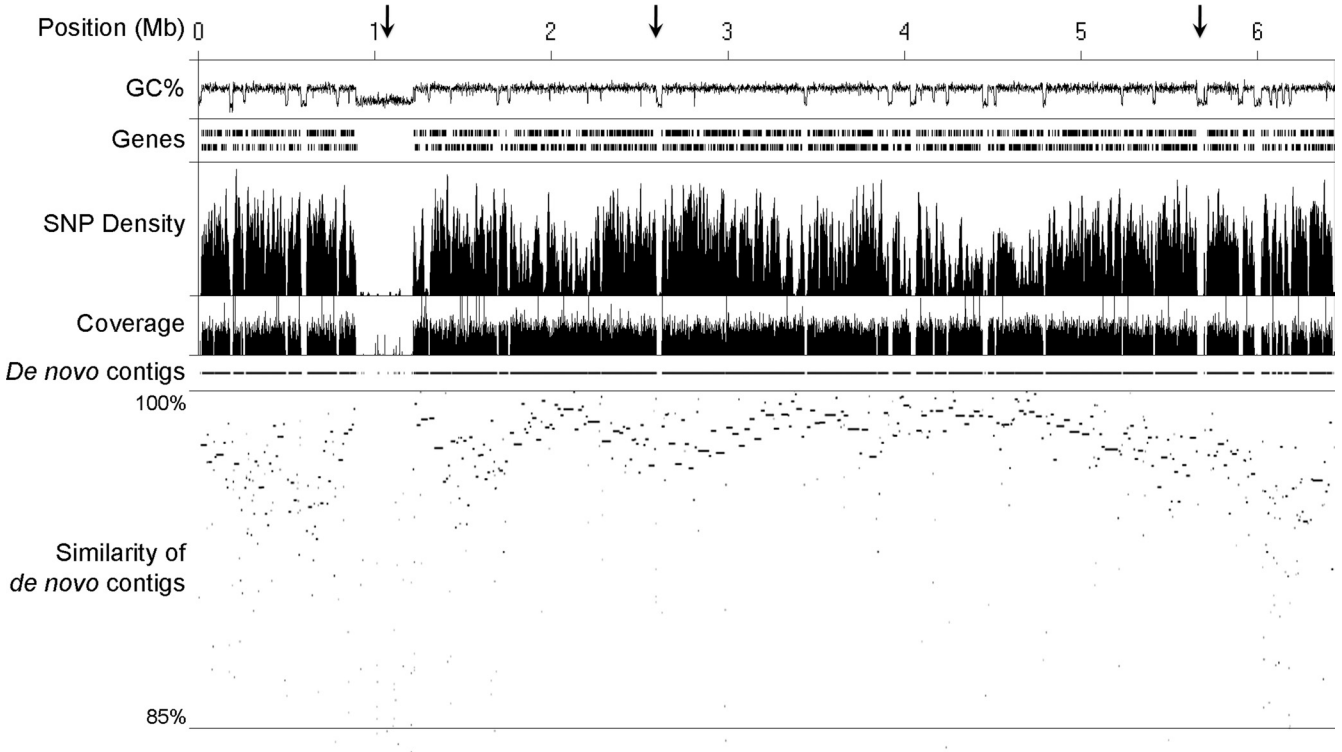

FIG. 2. Genome sequence of *N. crassa* MV. The SNP density (0 to 15 SNPs/kb), short read coverage (0 to 50×), and *de novo* contig coverage (as contigs) of MV Illumina sequencing of LG V were compared to LG V of the OR reference genome. Known and predicted genes and the GC percentage (0 to 100%) are shown above the mapped reads. Note the absence of mapped reads in most of the centromeric region (large gene sparse block at 1 Mb) and some dispersed AT-rich regions that are composed of RIP-mutated relics of transposons (arrows). The bottom panel shows the similarity of aligned MV contigs that were derived by *de novo* assembly (85 to 100% nucleotide similarity).

almost all (27,865 or 94.1%) were MV, while 1,747 were OR, suggesting a false-positive SNP call rate of 5.9%. We experimentally confirmed 47 randomly selected SNPs by RFLP after TaqI digest and found that in this case all (i.e., >97%) were correctly called SNPs (Fig. 4B). The small discrepancy between mapping to ESTs and experimentally validated SNPs suggests that Sanger sequencing of ESTs resulted in slightly higher error rates in SNP calls than direct Illumina sequencing of genomic DNA.

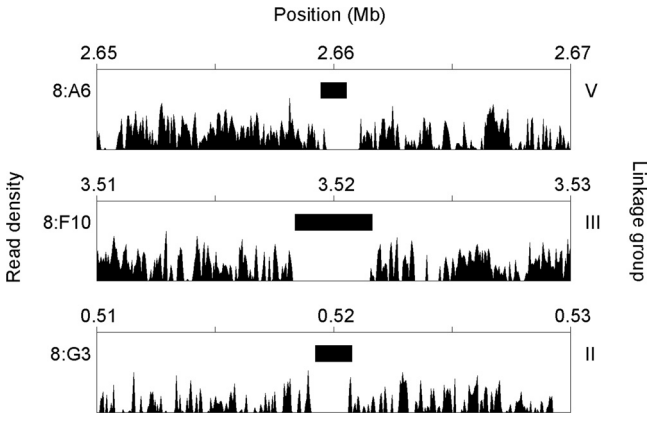

FIG. 3. Previously identified segments of transposon relics are absent from MV. Short read coverage (0 to 50×) of the MV sequencing compared to the OR reference genome for three regions (8:A6, 8:F10, and 8:G3) previously identified as absent in MV by Southern analyses (37) is shown. The Southern blot probes previously used are indicated by black rectangles.

Molecular mapping of *ndc-1*. The *ndc-1* strain (NMF164; FGSC3441) carries a ts mutation that results in arrest of the nuclear division cycle just prior to DNA synthesis when grown at 37°C (38). This strain also carries a linked lesion in *myo*-inositol-1-phosphatase (*inl*), which requires supplementation with *myo*-inositol for growth (38). The *ndc-1* gene had previously been mapped by classical three-point crosses to the left of *arg-4* (NCU10468.5) and *inl* (NCU06666.5) on the right arm of linkage group (LG) V (Fig. 5B) (38). To map *ndc-1* and to uncover the molecular cause of the ts phenotype, we crossed NMF164 with MV (NMF37; FGSC2225) (2). Random ascospores were collected from plate lids and germinated by heat shock at 65°C for 60 min. This treatment yielded virtually no (<0.05%) *ndc-1* progeny, confirming the previous suggestion that normal heat shock was lethal to *ndc-1* ascospores (38). Consistent with prior linkage analysis (38), five (~5%) *inl* progeny were recovered. We changed our conditions to heat shock at 60°C for 30 min, which resulted in recovery of 64 (~32%) *ndc-1* progeny, still somewhat lower than the expected 50%.

We extracted DNA from 63 strains to generate two separate pools of DNA for Illumina sequencing. The first pool consisted of equal amounts of DNA from 54 *ndc-1 inl* progeny, while the second pool consisted of equal amounts of DNA from nine *inl*.

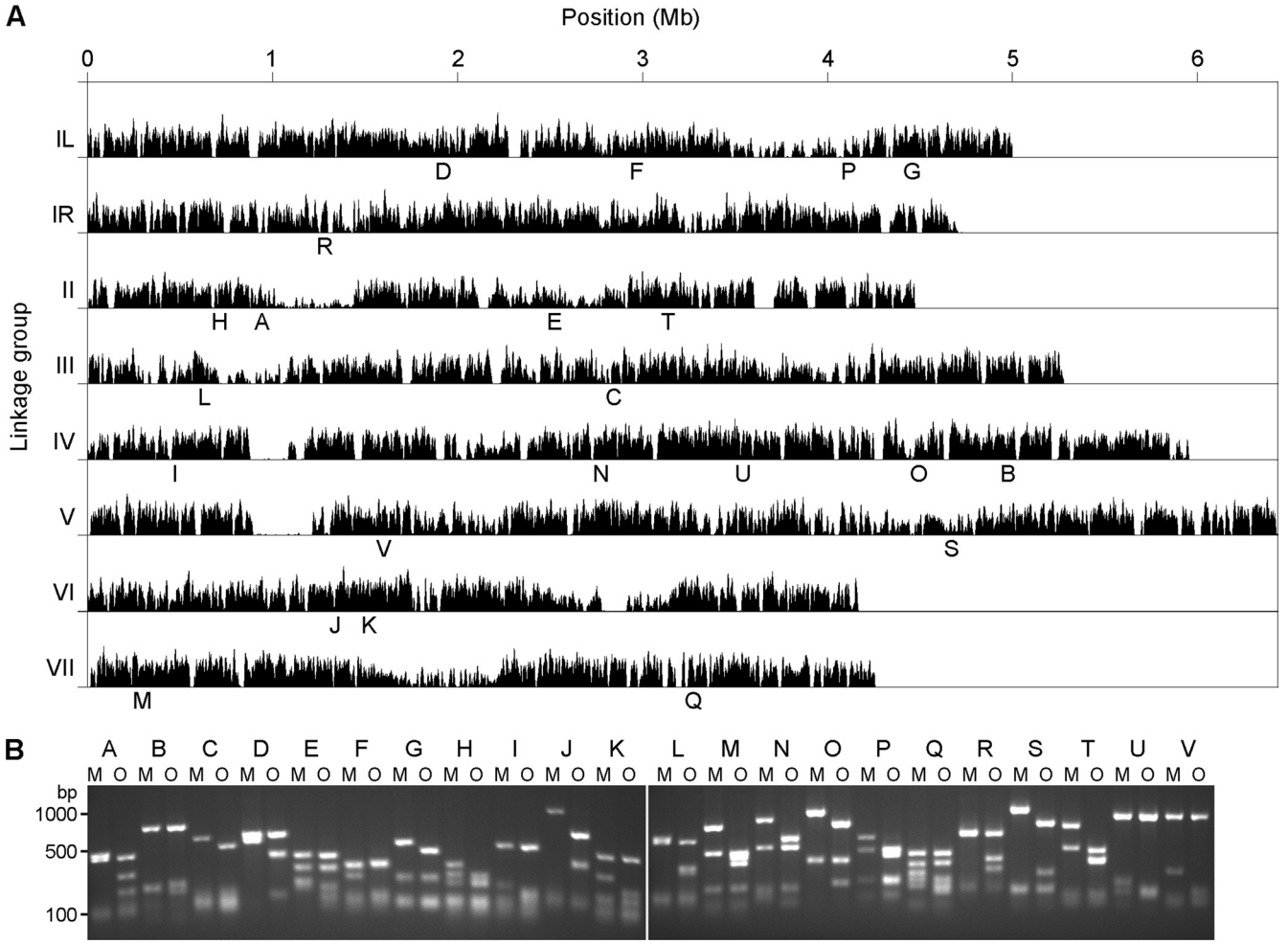

FIG. 4. SNP map of *N. crassa* MV compared to the OR reference genome. Between MV and OR, 168,579 SNPs were identified. (A) SNP density plot along all seven of the *Neurospora* chromosomes (0 to 26 SNPs/kb). (B) A subset of SNPs within a TaqI site present in OR (O) but not MV (M) was confirmed by RFLP. Their locations in the genome are indicated by letters in panel A.

This second pool was used to map the recombination track that must contain *ndc-1*, since this region of the genome should have MV genomic sequence at *ndc-1*$^{+}$. In contrast, the region with *inl* should contain DNA from the mutant background, while the remainder of the genome should be a mixture of DNA from both parents. The pool composed of DNA from the 54 *ndc-1 inl* progeny was used to find the specific mutation responsible for the *ndc-1* phenotype.

We generated a "SNPome," a FASTA file with both versions of each SNP between OR and MV, by running "SNPome_builder" (Fig. 1). We mapped processed 36-nt reads from the *ndc-1 inl* and *inl* pools to the OR and MV SNPome as perfect matches with Hashmatch (12) and ran SNP_profiler on the Hashmatch output to count the number of reads matching each version of each SNP (Fig. 1). This Perl script generates a ratio of reads matching each version of a SNP as a fraction between 0 and 1 to indicate the amount each parental genome contributes to the progeny in bulk. The resulting SNP profiles were viewed in the Argo genome browser to identify contiguous genomic regions in which SNPs originated from a specific parental strain (Fig. 5A).

As described above, the region containing the *ndc-1* locus should contain most—if not all—MV-type SNPs in the pool of nine *inl* strains. By this logic, we narrowed the location of *ndc-1* to a 0.85-Mb region on LG V, the only region in the genome with an extended stretch of MV-only SNPs (Fig. 5). We next mapped the *ndc-1 inl* reads to the OR reference genome with MAQ (24) and found 2,724 high-quality SNPs in the region containing *ndc-1* (between 2.70 and 3.55 Mb on LG V). NMF164 (FGSC3441) is of mixed genetic background. We found that the *ndc-1* region is highly polymorphic compared to the OR reference genome (14), which accounts for the high number of mutation candidates found. To exclude as many SNPs as possible, we subtracted SNPs that were also present in two additional *Neurospora* strains, FGSC322 and FGSC3562, two non-*ndc-1* strains of mixed and related backgrounds (K. McCluskey and S. Baker, unpublished data). This procedure reduced the number of candidate SNPs in the *ndc-1* region to 102. Of these, six SNPs caused nonsynonymous mutations within exons.

Three predicted SNPs resulted in changes in protein residues that were not conserved (in NCU01215, NCU01114, and

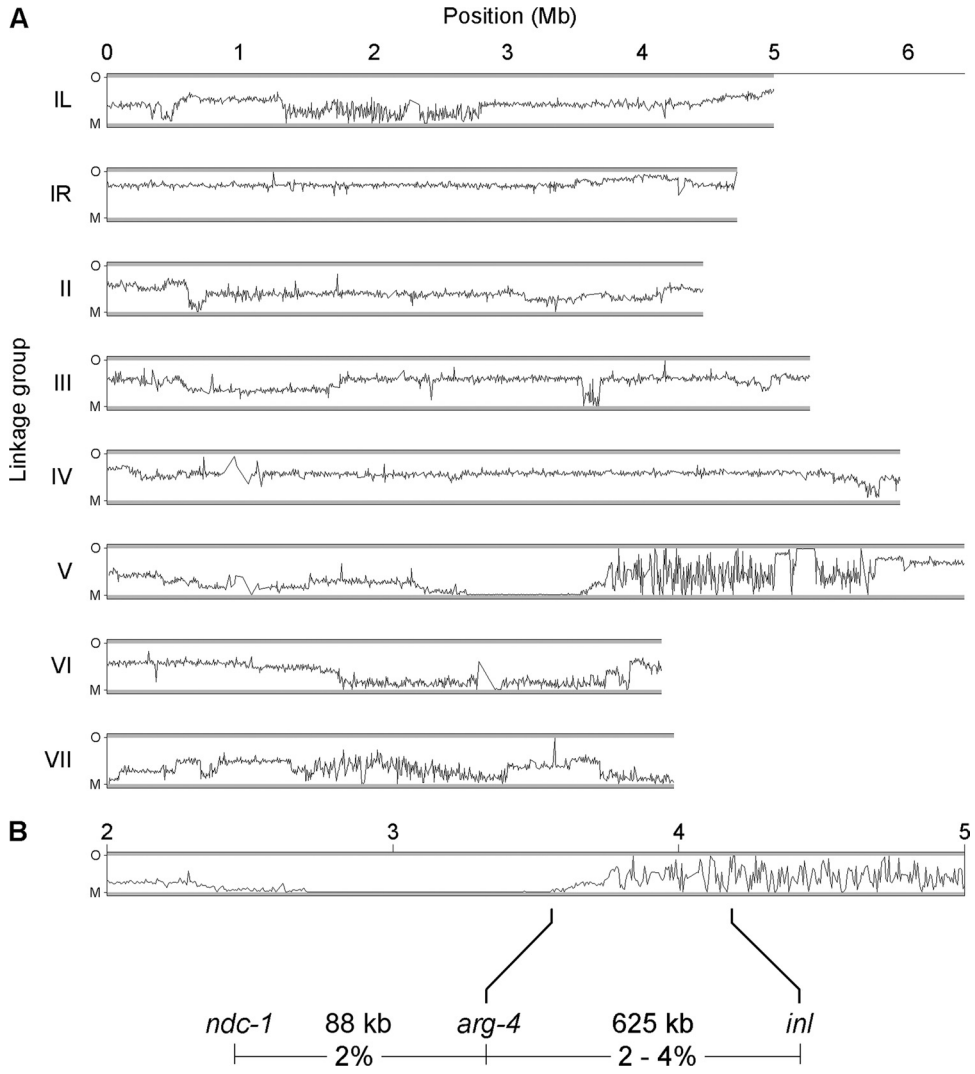

FIG. 5. Mapping the *ndc-1* mutant. (A) Genomic DNA from 54 *ndc-1*$^+$ *inl* bulk progeny was mapped to the OR/MV SNPome. The number of reads obtained from either the OR or MV background was tallied at each SNP, and a ratio of OR to MV reads was calculated to indicate the contribution from each genetic background in the bulk progeny. The ratio is shown as 100% OR (O) to 100% MV (M) along all seven *N. crassa* chromosomes as a 5-kb sliding window of the average ratio. Note the 0.85-Mb block on LG V where reads originate entirely with MV; thus, this region must contain *ndc-1*. (B) The region of LG V originating from the MV parent is shown in detail. Linkage analysis had previously mapped *ndc-1* to the left of *arg-4* (38).

NCU01090). Two SNPs resulted in changes from Thr to Ser or Gly to Glu in regions that were in more conserved protein motifs in NCU01172 and NCU01027, respectively. The most obvious candidate gene, however, was *spe-1* (NCU01271), the gene encoding ornithine decarboxylase (27, 34), since it has well-documented roles during the cell cycle (for a review, see reference 18). The predicted mutation causes a Phe-to-Ser change at residue 132 (F132S). This phenylalanine is conserved in all ODC proteins we investigated, including those from fungi, vertebrates, flies, trypanosomes, nematodes, cnidarians, and plants (Fig. 6A). To confirm that the predicted SNP was present in individual strains with the *ndc-1* mutation, we determined the sequence of *spe-1* from the original mutant (NMF164) (Fig. 7), three *ndc-1* siblings, and one *ndc-1*$^+$ sibling from the cross with MV (data not shown). In all cases, the DNA sequence from the *spe-1* gene matched that expected from high-throughput sequencing.

Complementation of *ndc-1*. To determine if the T-to-C mutation resulting in the F132S substitution was the cause of the *ndc-1* phenotype, we carried out two experiments. First, we grew *ndc-1* strains at restrictive temperatures in the presence of 1 mM spermidine trihydrochloride, a metabolic product downstream of the reaction catalyzed by ornithine decarboxylase (Fig. 6B). Normal growth of *ndc-1* at 37°C after supplementation (Fig. 7) strongly suggested that *ndc-1* is indeed an allele of *spe-1*. Next, we integrated the selectable *hph* marker into pGS1, a plasmid that contains the *spe-1* region and had been used previously to complement *spe-1* mutants (45). We transformed NMF164 with this construct, pKP19, and selected for hygromycin resistance. Transformants were tested for

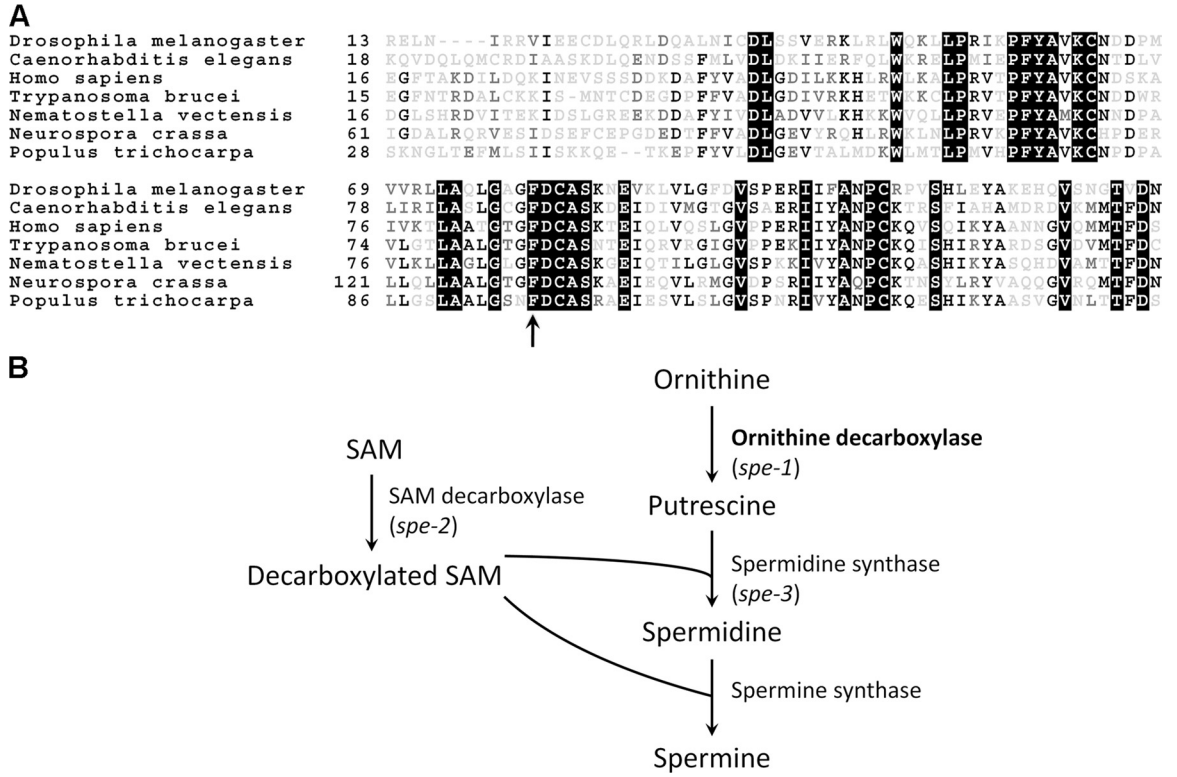

FIG. 6. The *ndc-1* mutation maps to the gene for ornithine decarboxylase, *spe-1*. (A) We identified a Phe-to-Ser mutation at residue 132 (F132S, indicated by an arrow) in the *N. crassa* ODC. Alignment of ornithine decarboxylase proteins from a variety of organisms, including wild-type *Neurospora*, indicates the mutated residue is highly conserved. Only a small region of ODC is shown here. (B) Production of polyamines in *N. crassa*. Mutant alleles are indicated in italics. SAM, *S*-adenosylmethionine.

growth at 37°C to determine if the ectopically integrated plasmid complemented the ts *ndc-1* phenotype. All (7/7) transformants grew at 37°C, strongly suggesting that *ndc-1* is an allele of *spe-l*. Integration of the plasmid and heterozygosity of the *spe-1* gene in the transformant heterokaryon were confirmed by Sanger sequencing (Fig. 7). Based on our results, we propose changing *ndc-1* to *spe-1*^{ndc}.

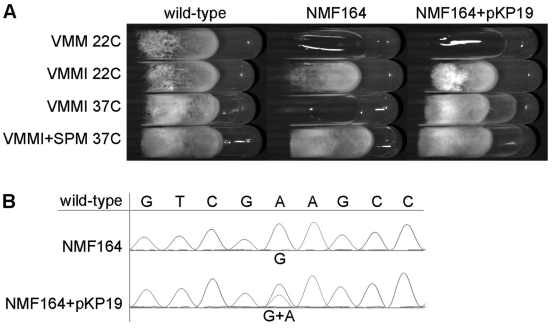

FIG. 7. The *ndc-1* ts phenotype can be reverted by supplementation with spermidine or integration of a wild-type copy of *spe-1*. (A) VMM slants with or without *myo*-inositol (VMMI) and spermidine (SPM) were inoculated with conidia from NMF398 (*ndc-1*$^{+}$ *inl*$^{+}$), NMF164 (*ndc-1 inl*), or NMF164 transformed with pKP19, a plasmid containing the wild-type *spe-1* gene (NMF401; *ndc-1*$^{+}$ *inl*). Images were captured after 3 days of incubation at either 22°C or 37°C. (B) Complementation of *ndc-1* was confirmed by Sanger sequencing. Sequencing peaks for both the wild type and the *ndc-1* allele indicate that both are present in the heterokaryotic transformant. The base(s) present at the SNP site is indicated.

Utility of *ndc-1* for blocking and releasing cells in the nuclear division cycle. We observed that *spe-1*^{ndc} strains recovered and resumed growth after flasks or slants were shifted from 37°C to 22°C. This paves the way for utilizing *spe-1*^{ndc} to synchronize the nuclear division cycle and DNA synthesis, which has been difficult in *Neurospora*. We grew *ndc-1* strains at 37°C in race tubes and shifted them to 22°C to measure the time lag before linear growth resumed. We found that cells stalled for 28.4 h took 15.8 h to recover (Fig. 8A), while cells stalled for 13.0 h took 10.5 h to recover (Fig. 8B). In both cases, recovery of three replicate strains occurred simultaneously.

DISCUSSION

Mauriceville genome assembly and SNP identification. We sequenced the *N. crassa* MV genome to ~20-fold median coverage and mapped reads to the current assembly of the OR genome to identify SNPs between these two commonly used mapping strains. The highest-quality SNP calls are available as a resource for the *Neurospora* community on the *Neurospora* model organism homepage at http://www.fgsc.net/Neurospora/SNPs/SNP_map.htm and are included here as Table S3 in the supplemental material. In addition, we carried out *de novo* assembly of MV reads to generate an MV genome that did not rely on a reference genome. *De novo* assembly statistics suggest

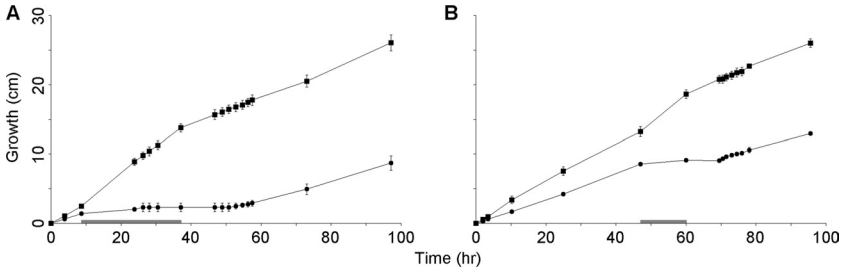

FIG. 8. Growth arrest and release of ts *spe-1*^{ndc} strain. Conidia from NMF398 (*ndc-1*$^{+}$ *inl*$^{+}$; squares) or NMF344 (*ndc-1 inl*; circles) were inoculated on VMMI in race tubes at 22°C to monitor linear growth. The conidia were allowed to germinate and establish linear growth prior to transfer to 37°C (indicated by a thick bar on the *x* axis). The strains were maintained at 37°C for at least 13 h and then transferred back to 22°C to determine the lag period before linear growth resumed. Error bars indicate the standard deviation between three replicates.

that we captured 34.7 Mb or 85% of the ~41-Mb genome in 10,434 contigs. Considering the relatively large amount of near-repetitive DNA (14), this compares favorably with the *de novo* assembly of the *Sordaria macrospora* genome (32), which was reference guided and included longer reads derived from Roche 454 sequencing.

We found that SNPs are distributed evenly over all seven chromosomes, with the exception of large blocks of AT-rich DNA present primarily at the centromeres or dispersed as transposon relics stemming from the action of repeat-induced point mutation (RIP) (37) (Fig. 2). We predicted that these regions are highly disparate in MV from those in OR and thus examined both read coverage and *de novo* assembly alignment in these regions. Several selected regions that had been previously shown to be absent from MV by Southern hybridization with OR probes (37) had few or no reads in MV (Fig. 3). The AT-rich reads remaining after alignment of MV reads to OR thus constitute dispersed subtelomeric and centromeric regions, and in further studies we will address the evolution of centromeric DNA. Here we show that these regions are highly variable between MV and OR.

Mapping *ndc-1* by bulk segregant analysis. Mapping a mutation by molecular means can be complicated under ideal conditions when the mutant was isolated in a genetic background for which a reference genome is available. When this is the case, mutants are backcrossed many times to isogenic or near-isogenic strains, and premutation and backcrossed mutant strains are compared with the idea of identifying the underlying lesion directly (for example, see reference 19). This can be a difficult and time-consuming undertaking, even with comparatively small, usually haploid genomes of fungi (C. Yam, K. R. Pomraning, M. Freitag, and S. Oliferenko, unpublished data). Direct comparisons of two or even several isogenic strains do not usually establish linkage, and to make matters worse, it is generally not known whether the mutation in question is a point mutation, indel, rearrangement, or the result of epigenetic changes. In contrast, doing even a single backcross to a strain of known and overall dissimilar haplotype followed by bulk segregant analysis with a dense SNP map is useful, since SNP maps immediately establish linkage of mutations to specific genomic regions even if the type of mutation is unknown. It is easiest to find point mutations, as we describe here, but with sufficient coverage or paired-end sequencing of bulk segregant pools, we predict it will be straightforward to identify indels or genetic rearrangements by high-throughput sequencing.

We have encountered two problems when using bulk segregant analysis. The first problem occurs when the genetic background of the mutant strain is unknown, as is often the case with classical mutants. One example of this case is discussed here, since *ndc-1* was obtained in a strain of mixed genetic background (38). We treated NMF164 as OR in the cross with MV, but we showed that OR was a poor surrogate parent, which left us to consider thousands of SNPs not present in the OR background as potential candidates for the *ndc-1* mutation. To alleviate this problem, we analyzed the ~1-Mb segment on LG V that we predicted to contain *ndc-1* in a diverse set of *Neurospora* wild-type strains with similarly mixed backgrounds (K. McCluskey and S. Baker, unpublished data), with the single criterion being a lack of the *ndc-1* phenotype. We reasoned that SNPs present in both wild-type and *ndc-1* strains cannot be responsible for the *ndc-1* phenotype and thus should be subtracted from the pool of candidate mutations. While not ideal, this strategy will become ever more effective as a wider palette of strains are sequenced and made publically available. If, however, *Neurospora* researchers embark on novel screens and selections for mutations, we recommend using the OR or MV strain as genetic background to accelerate and simplify future molecular mapping.

If the parental strain prior to mutagenesis is available but is not of OR or MV background, the analysis is simplified by setting mapping crosses with OR (because its assembly is superior to that of MV) and sequencing both the premutagenesis parent and the bulk mutant progeny. SNPs identified in the parent strain are subtracted from those identified by bulk segregant analysis from the pool of mutant progeny. This leaves the subset of SNPs introduced by the mutagenesis as candidates to screen.

The second problem encountered is specific to the mating type region on *N. crassa* LG I. This region is idiomorphic in the opposite mating types (15, 39). By sequencing additional *Neurospora* strains of various backgrounds, we found that the mating type region and adjacent large regions of LG I extending across the centromeric DNA are typically extremely polymorphic (K. R. Pomraning, K. M. Smith, F. J. Bowring, D. Bell-Pedersen, D. E. A. Catcheside, E. U. Selker, and M. Freitag, unpublished data). Lab strains of *Neurospora* are selected by most researchers to carry out genetic crosses with fertile indi-

viduals, with little regard as to neutral mutations associated with mating type, which are allowed to accumulate over several generations. This presumably resulted in many haplotypes flanking the mating type regions. There is also some evidence for rich polymorphisms in this region from different wild-type strains. In the Adiopodoume strain from West Africa, for example, a large number of silent mutations have been identified in a gene encoding a translesion DNA polymerase (NCU01951) near *mat* (40). This richness in SNPs makes mapping of mutations closely linked to the mating type region difficult, especially if the SNPs used are relevant only for a single mating type. If a mutation is linked to this region, progeny should be pooled by mating type and analyzed with a SNPome generated for either *mata* or *matA* of LG I.

It is tempting to sequence only the premutant parent and mutant strains, as has been done with success to identify single point mutations in fission yeast (19). In organisms with larger genomes, however, bulk segregant analysis is used (8, 25, 44). If the locations for recombination events are assumed to be random during meiosis, then the power to identify the region containing a mutation by bulk segregant analysis increases as $1/n$, where n is the number of progeny sequenced in bulk, and can be roughly modeled as

$$S = [C/(R + 1)]n^{-1} \quad (1)$$

where S is the size of the genomic region sought, C is the size of the mutated chromosome, and R is the number of recombination events that occur on the mutated chromosome during meiosis. For *N. crassa* LG V, we know the size is approximately 6.4 Mb and we can conservatively estimate that recombination occurs once on each chromosome. Thus, for *N. crassa* LG V,

$$S = [6.4 \text{ Mb}/(1 + 1)]n^{-1} \quad (2)$$

In this study, we sequenced two pools containing nine (S = 0.85 Mb; Fig. 5) and 54 (S = 0.15 Mb; data not shown) progeny (Fig. 9). We were able to narrow the mutation to a region larger than expected from equation 2 (S = 0.36 Mb for nine progeny, and S = 0.06 Mb for 54 progeny), suggesting that recombination is slightly suppressed in this region of LG V. By generating and carefully pooling DNA from additional progeny, one can narrow recombination blocks containing the mutation(s) in question to smaller sizes, but of course with diminishing rewards for every new progeny added.

Importance of ODC for the cell cycle. We identified the *ndc-1* mutation as F132S in ornithine decarboxylase (ODC), the enzyme that catalyzes the initial rate-limiting step of polyamine biosynthesis by converting ornithine into putrescine, which is then converted to spermidine and spermine (10) (Fig. 6). Putrescine production is important for the function of eukaryotic cells, and with few exceptions, e.g., within the *Brassicaceae*, which use arginine decarboxylase to synthesize putrescine (16), eukaryotes utilize a conserved ODC enzyme for putrescine production. Unraveling which of the many roles of polyamines contribute to the cell cycle stalling observed with the *ndc-1* allele will be the subject of future work. It is clear that polyamines affect a wide variety of cellular processes by interacting with DNA, RNA, and phospholipids (30, 43). Studies in fission yeast point to a spermidine-requiring modification of lysine to the unique amino acid hypusine in eukaryotic translation initiation factor 5A (33). This modification is essential for cell proliferation and utilizes much of the free spermidine in the cell under spermidine-limiting conditions (7). Polyamines interact with DNA as cations by binding to phosphates but also form nuclear aggregates of polyamines, which effectively protect DNA from endonucleases *in vitro* (9). These complexes stabilize DNA structure through condensation, depletion of which could be catastrophic for genomic DNA packaging within the nucleus. Finally, polyamines are required for normal microtubule assembly (36), which is presumably necessary for spindle formation after spindle pole body duplication, roughly the point at which the *ndc-1* allelic mutant appears arrested (38). Outside of the cell cycle proper, growing hyphae of filamentous fungi require microtubules for the transport of cargo to the Spitzenkörper at the growing hyphal tip (41). Disruption of this process may result in the stalled linear growth observed here.

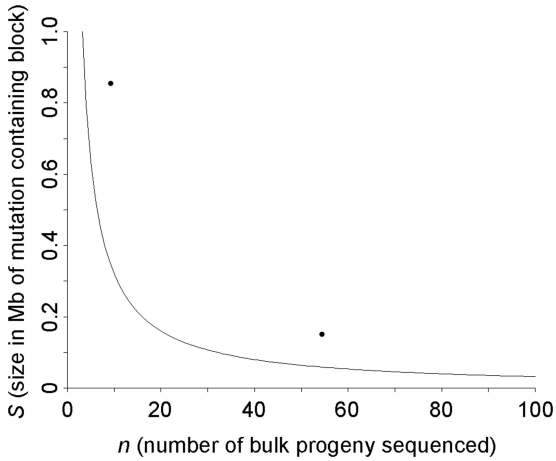

FIG. 9. Predicted relationship of number of progeny in pools to expected size of mutation-containing region. Equation 1 predicts the relationship between the number of progeny sequenced in bulk (n) and the size of the mutation-containing region found (S). It is plotted for *N. crassa* chromosome V assuming the size of the chromosome (C) equals 6.4 Mb and the number of recombination events on chromosome V during meiosis (R) equals 1. We experimentally narrowed S to a 0.85-Mb region using 54 pooled progeny and a 0.15-Mb region using 9 pooled progeny (data not shown), as indicated by two circles.

Application of the *spe-1^{ndc}* allele. At present, the *spe-1^{ndc}* mutant is the only ts nuclear division cycle mutant known in *N. crassa*. We observed that growth of the mutant is stalled by incubation for several hours at 37°C but can be restored by transfer back to 22°C (Fig. 8). Supplementation with 1 mM spermidine allows the strain to grow at 32°C and 37°C (Fig. 7A). While picolinic acid, which rapidly slows DNA accumulation (26), and deoxyadenosine, a DNA synthesis inhibitor (13), have been used in the past to synchronize the nuclear division cycle in *N. crassa*, we suggest that shifting *Neurospora* from high to low temperature or supplementation with spermidine is less invasive than the former treatments. To that end, we are currently defining the precise growth conditions to synchronize the cell cycle of *spe-1^{ndc}* strains with the goal of generating a system that will allow synchronized DNA synthesis upon addition of spermidine.

ACKNOWLEDGMENTS

We thank Kevin McCluskey (FGSC, University of Missouri, Kansas City, MO) and Scott Baker (Pacific Northwest National Lab, Richland, WA) for access to *N. crassa* raw sequence data and helpful comments on the manuscript. Mark Dasenko, Matthew Peterson, and Chris Sullivan at the OSU CGRB core facility assisted with Illumina sequencing. We are indebted to the *Neurospora* Functional Genomics Project for primers and the Fungal Genetics Stock Center for *Neurospora* strains.

This work was made possible by start-up funds from the OSU Computational and Genome Biology Initiative and grants from the American Cancer Society (RSG-08-030-01-CCG), DOE (DE-FG02-08ER64665), and NIH (P01 GM068087) to M.F.

REFERENCES

1. **Baird, N. A., et al.** 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One **3:**e3376.
2. **Beauchamp, P. M., E. W. Horn, and S. R. Gross.** 1977. Proposed involvement of an internal promoter in regulation and synthesis of mitochondrial and cytoplasmic leucyl-tRNA synthetases of Neurospora. Proc. Natl. Acad. Sci. U. S. A. **74:**1172–1176.
3. **Birkeland, S. R., et al.** 2010. Discovery of mutations in *Saccharomyces cerevisiae* by pooled linkage analysis and whole-genome sequencing. Genetics **186:**1127–1137.
4. **Blumenstiel, J. P., et al.** 2009. Identification of EMS-induced mutations in *Drosophila melanogaster* by whole-genome sequencing. Genetics **182:**25–32.
5. **Brauer, M. J., C. M. Christianson, D. A. Pai, and M. J. Dunham.** 2006. Mapping novel traits by array-assisted bulk segregant analysis in *Saccharomyces cerevisiae*. Genetics **173:**1813–1816.
6. **Case, M. E., H. E. Brockman, and F. J. de Serres.** 1965. Further information on the origin of the Yale and Oak Ridge wild-type strains of *Neurospora crassa.* Neurospora Newsl. **8:**25–26.
7. **Chattopadhyay, M. K., M. H. Park, and H. Tabor.** 2008. Hypusine modification for growth is the major function of spermidine in *Saccharomyces cerevisiae* polyamine auxotrophs grown in limiting spermidine. Proc. Natl. Acad. Sci. U. S. A. **105:**6554–6559.
8. **Cuperus, J. T., et al.** 2010. Identification of MIR390a precursor processing-defective mutants in Arabidopsis by direct genome sequencing. Proc. Natl. Acad. Sci. U. S. A. **107:**466–471.
9. **D'Agostino, L., M. di Pietro, and A. Di Luccia.** 2006. Nuclear aggregates of polyamines. IUBMB Life **58:**75–82.
10. **Davis, R. H.** 2000. Neurospora: contributions of a model organism. Oxford University Press, Oxford, United Kingdom.
11. **Ehrenreich, I. M., et al.** 2010. Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature **464:**1039–1042.
12. **Filichkin, S. A., et al.** 2010. Genome-wide mapping of alternative splicing in *Arabidopsis thaliana.* Genome Res. **20:**45–58.
13. **Fletcher, M. H., and A. P. Trinci.** 1980. Synchronization of DNA synthesis in *Neurospora crassa* by 2′-deoxyadenosine and spore selection. Arch. Microbiol. **128:**113–119.
14. **Galagan, J. E., et al.** 2003. The genome sequence of the filamentous fungus *Neurospora crassa.* Nature **422:**859–868.
15. **Glass, N. L., J. Grotelueschen, and R. L. Metzenberg.** 1990. *Neurospora crassa* A mating-type region. Proc. Natl. Acad. Sci. U. S. A. **87:**4912–4916.
16. **Hanfrey, C., S. Sommer, M. J. Mayer, D. Burtin, and A. J. Michael.** 2001. Arabidopsis polyamine biosynthesis: absence of ornithine decarboxylase and the mechanism of arginine decarboxylase activity. Plant J. **27:**551–560.
17. **Horowitz, N. H., M. Fling, H. L. Macleod, and N. Sueoka.** 1960. Genetic determination and enzymatic induction of tyrosinase in Neurospora. J. Mol. Biol. **2:**96–104.
18. **Igarashi, K., and K. Kashiwagi.** 2010. Modulation of cellular function by polyamines. Int. J. Biochem. Cell Biol. **42:**39–51.
19. **Irvine, D. V., et al.** 2009. Mapping epigenetic mutations in fission yeast using whole-genome next-generation sequencing. Genome Res. **19:**1077–1083.
20. **Jin, Y., et al.** 2007. Rapid genetic mapping in *Neurospora crassa.* Fungal Genet. Biol. **44:**455–465.
21. **Kurtz, S., et al.** 2004. Versatile and open software for comparing large genomes. Genome Biol. **5:**R12.
22. **Lambreghts, R., et al.** 2009. A high-density single nucleotide polymorphism map for *Neurospora crassa.* Genetics **181:**767–781.
23. **Lewis, Z. A., et al.** 2007. High-density detection of restriction-site-associated DNA markers for rapid mapping of mutated loci in Neurospora. Genetics **177:**1163–1171.
24. **Li, H., J. Ruan, and R. Durbin.** 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. **18:**1851–1858.
25. **Liu, S., et al.** 2010. High-throughput genetic mapping of mutants via quantitative single nucleotide polymorphism typing. Genetics **184:**19–26.
26. **Martegani, E., M. Levi, F. Trezzi, and L. Alberghina.** 1980. Nuclear division cycle in *Neurospora crassa* hyphae under different growth conditions. J. Bacteriol. **142:**268–275.
27. **McDougall, K. J., J. Deters, and J. Miskimen.** 1977. Isolation of putrescine-requiring mutants of *Neurospora crassa.* Antonie Van Leeuwenhoek **43:**143–151.
28. **Metzenberg, R. L.** 1984. A method for finding the genetic map position of cloned DNA fragments. Neurospora Newsl. **31:**35–39.
29. **Michelmore, R. W., I. Paran, and R. V. Kesseli.** 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Natl. Acad. Sci. U. S. A. **88:**9828–9832.
30. **Miyamoto, S., K. Kashiwagi, K. Ito, S. Watanabe, and K. Igarashi.** 1993. Estimation of polyamine distribution and polyamine stimulation of protein synthesis in *Escherichia coli.* Arch. Biochem. Biophys. **300:**63–68.
31. **Nelson, M. A., and D. D. Perkins.** 2000. Restriction polymorphism maps of *Neurospora crassa*: 2000 update. Fungal Genet. Newsl. **47:**25–39.
32. **Nowrousian, M., et al.** 2010. De novo assembly of a 40 Mb eukaryotic genome from short sequence reads: *Sordaria macrospora*, a model organism for fungal morphogenesis. PLoS Genet. **6:**e1000891.
33. **Park, M. H.** 2006. The post-translational synthesis of a polyamine-derived amino acid, hypusine, in the eukaryotic translation initiation factor 5A (eIF5A). J. Biochem. **139:**161–169.
34. **Paulus, T. J., P. Kiyono, and R. H. Davis.** 1982. Polyamine-deficient *Neurospora crassa* mutants and synthesis of cadaverine. J. Bacteriol. **152:**291–297.
35. **Pomraning, K. R., K. M. Smith, and M. Freitag.** 2009. Genome-wide high throughput analysis of DNA methylation in eukaryotes. Methods **47:**142–150.
36. **Savarin, P., et al.** 2010. A central role for polyamines in microtubule assembly in cells. Biochem. J. **430:**151–159.
37. **Selker, E. U., et al.** 2003. The methylated component of the *Neurospora crassa* genome. Nature **422:**893–897.
38. **Serna, L., and D. Stadler.** 1978. Nuclear division cycle in germinating conidia of *Neurospora crassa.* J. Bacteriol. **136:**341–351.
39. **Staben, C., and C. Yanofsky.** 1990. *Neurospora crassa* a mating-type region. Proc. Natl. Acad. Sci. U. S. A. **87:**4917–4921.
40. **Tamuli, R., and D. P. Kasbekar.** 2008. Dominant suppression of repeat-induced point mutation in *Neurospora crassa* by a variant catalytic subunit of DNA polymerase zeta. Genetics **178:**1169–1176.
41. **Uchida, M., R. R. Mourino-Perez, M. Freitag, S. Bartnicki-Garcia, and R. W. Roberson.** 2008. Microtubule dynamics and the role of molecular motors in *Neurospora crassa.* Fungal Genet. Biol. **45:**683–692.
42. **Vigfusson, N. V., and J. Weijer.** 1972. Sexuality in *Neurospora crassa.* II. Genes affecting the sexual development cycle. Genet. Res. **19:**205–211.
43. **Watanabe, S., K. Kusama-Eguchi, H. Kobayashi, and K. Igarashi.** 1991. Estimation of polyamine binding to macromolecules and ATP in bovine lymphocytes and rat liver. J. Biol. Chem. **266:**20803–20809.
44. **Wenger, J. W., K. Schwartz, and G. Sherlock.** 2010. Bulk segregant analysis by high-throughput sequencing reveals a novel xylose utilization gene from *Saccharomyces cerevisiae.* PLoS Genet. **6:**e1000942.
45. **Williams, L. J., et al.** 1992. Ornithine decarboxylase gene of *Neurospora crassa*: isolation, sequence, and polyamine-mediated regulation of its mRNA. Mol. Cell. Biol. **12:**347–359.
46. **Zerbino, D. R., and E. Birney.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18:**821–829.