

# A High-Density Single Nucleotide Polymorphism Map for *Neurospora crassa*

Randy Lambrechts,\* Mi Shi,\* William J. Belden,\* David deCaprio,<sup>†</sup> Danny Park,<sup>†</sup>  
Matthew R. Henn,<sup>†</sup> James E. Galagan,<sup>†</sup> Meray Baştürkmen,<sup>‡,1</sup> Bruce W. Birren,<sup>†</sup>  
Matthew S. Sachs,<sup>‡,1</sup> Jay C. Dunlap\* and Jennifer J. Loros\*<sup>§,2</sup>

\*Department of Genetics and <sup>§</sup>Department of Biochemistry, Dartmouth Medical School, Hanover, New Hampshire 03755,

<sup>†</sup>Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02142 and <sup>‡</sup>Department of Environmental and Biomolecular Systems, Oregon Health & Science University, Beaverton, Oregon 97006

Manuscript received March 14, 2008

Accepted for publication November 17, 2008

## ABSTRACT

We report the discovery and validation of a set of single nucleotide polymorphisms (SNPs) between the reference *Neurospora crassa* strain Oak Ridge and the Mauriceville strain (FGSC 2555), of sufficient density to allow fine mapping of most loci. Sequencing of Mauriceville cDNAs and alignment to the completed genomic sequence of the Oak Ridge strain identified 19,087 putative SNPs. Of these, a subset was validated by cleaved amplified polymorphic sequence (CAPS), a simple and robust PCR-based assay that reliably distinguishes between SNP alleles. Experimental confirmation resulted in the development of 250 CAPS markers distributed evenly over the genome. To demonstrate the applicability of this map, we used bulked segregant analysis followed by interval mapping to locate the *csp-1* mutation to a narrow region on LGI. Subsequently, we refined mapping resolution to 74 kbp by developing additional markers, resequenced the candidate gene, NCU02713.3, in the mutant background, and phenocopied the mutation by gene replacement in the WT strain. Together, these techniques demonstrate a generally applicable and straightforward approach for the isolation of novel genes from existing mutants. Data on both putative and validated SNPs are deposited in a customized public database at the Broad Institute, which encourages augmentation by community users.

NEUTRAL polymorphisms have been extensively used to elucidate the genetic causes of phenotypic aberrations, including human diseases (WANG *et al.* 1998; DETERA-WADLEIGH and McMAHON 2004) as well as mutations of interest in model organisms (WINZELER *et al.* 1998; BERGER *et al.* 2001; WICKS *et al.* 2001). The genetic location of any mutation or quantitative trait locus can be narrowed down to any degree of precision by following the cosegregation of the phenotype with a range of increasingly closer markers at known locations, provided that sufficient markers are available.

Recombination-based mapping depends on the existence of an alternative strain that bears genetic markers distinct from the reference but that remains interfertile with the reference strain. Fungal Genetic Stock Center (FGSC) 2225, an isolate of wild-type (WT) *Neurospora crassa* from Mauriceville (MV), Texas, had previously been used to establish a low density restriction fragment-length polymorphism (RFLP) map (METZENBERG and GROBELWESCHEN 1988) on the basis of the availability of phenotypically neutral but easily scorable polymorphisms relative to the standard Oak Ridge (OR) strain (74-OR23-1VA, FGSC 2489). With the completed ge-

nome sequence of the latter available (GALAGAN *et al.* 2003), the addition of more polymorphism data has become more straightforward as single nucleotide polymorphisms (SNPs) are easily assigned to specific physical locations within the genome.

By far the most prevalent type of intergenomic variation consists of SNPs. In transcribed regions, we found sequence divergence between Oak Ridge and Mauriceville to be ~0.1%, largely due to single nucleotide substitutions (DUNLAP *et al.* 2007). Hence, to detect a sufficiently large number of SNPs for mapping purposes, it would suffice to sequence a limited sample of the Mauriceville genome, as can be found in a Mauriceville-derived EST library. Since Mauriceville, on the other hand, is thought to be genetically close enough to the reference Oak Ridge strain to have highly similar gene structure, this effort would at the same time provide experimental validation of gene annotation and thus serve to improve gene calling (M. GALAGAN and B. BIRREN, unpublished results). Moreover, alignment of Mauriceville sequences to the finished genome sequence is straightforward given the low frequency of gene duplication in *Neurospora* (SELKER 1997), which allows most newly discovered SNPs to be placed easily on the physical map.

For SNPs to be useful in mapping, alternative alleles must be distinguishable by a simple assay that can be performed quickly and reliably on large numbers of

<sup>1</sup>Present address: Department of Biology, Texas A&M University, College Station, TX 77843

<sup>2</sup>Corresponding author: Department of Genetics, Dartmouth Medical School, Hanover, NH 03755. E-mail: jennifer.loros@dartmouth.edu

progeny. PCR based methods are commonly used. The single nucleotide amplified polymorphism (SNAP) (DRENKARD *et al.* 2000) assay employs two primer pairs, each of which is allele specific, to obtain differential amplification of a specific SNP. One advantage of this method is that it can potentially be applied to any given SNP; however, we found that individual SNPs often require optimization of primer sequences and PCR conditions, thereby making it a less desirable choice for high-throughput genotyping (DUNLAP *et al.* 2007). An alternative method, cleaved amplified polymorphic sequence (CAPS), relies on the 30–40% fraction of single nucleotide substitutions that result in the creation or deletion of a restriction enzyme recognition site (KONIECZNY and AUSUBEL 1993). Since in this assay both alleles are independently detectable they constitute codominant markers, allowing the researcher to perform the assay on multiple DNA samples in the same reaction, so-called bulked segregant analysis (BSA) (MICHELMORE *et al.* 1991).

In this work, we used data from high-throughput sequencing of two independent Mauriceville EST libraries to identify putative SNPs (pSNPs) and, for one of these libraries, developed experimental procedures on the basis of the CAPS methodology to distinguish Oak Ridge from Mauriceville alleles. By combining an accurate SNP detection algorithm (ALTSCHULER *et al.* 2000) with efficient CAPS design we obtained a near-saturating marker density (given the data set and the assay type), with, on average, one SNP every 7 cM. This is close to the minimum recombination that can be distinguished from perfect linkage using bulked segregant analysis (BSA) (JIN *et al.* 2007). We then mined the existing library of pSNPs and screened for additional CAPS polymorphisms to increase the local density of the SNP map to a level allowing high-precision mapping. We used these to map the *conidial separation 1* (*csp-1*) gene (SELITRENNIKOFF *et al.* 1974) to a 74-kbp region and employed a candidate gene approach to identify the causative mutation.

Strains containing the *csp-1*<sup>UCLA37</sup> allele develop superficially normal-looking conidia that fail to separate completely and remain tightly linked (SELITRENNIKOFF *et al.* 1974). Electron microscopy reveals that the process of conidial development is initiated normally, up until the formation of major constrictions in aerial hyphae, but that these constrictions do not fully close to release unicellular conidia (SPRINGER and YANOFSKY 1989). Macroscopically, the phenotype is easily scored by tapping a mature culture and observing if a cloud of airborne conidia is released, as is the case in wild-type (WT) cultures. Strains containing this mutation have seen wide utility in protocols that call for limited spread of conidia, *e.g.*, to contain contamination risks in teaching labs or in Petri-dish-based assays (MATTEEN and BRODY 1979), and are exempt from National Institutes of Health guidelines for recombinant DNA experiments. The *csp-1* gene had

previously been mapped to a region near the centromere of linkage group I (LGI) (SELITRENNIKOFF *et al.* 1974), but low recombination rates in this region have hindered its molecular identification. As mutations at most loci should be more tractable by genetic mapping, we reasoned *csp-1* would provide a good illustration of the general applicability of SNP mapping.

BSA is a method that can differentiate between linked and unlinked markers, provided the assay used allows independent visualization of both marker alleles, as with CAPS. BSA begins with obtaining progeny from a cross segregating two alleles of the gene of interest. In this case, *csp-1*<sup>UCLA37</sup> is the “mutant” allele and is in the OR parent, while the WT allele is in the MV parent. Individual progeny are sorted by phenotype and collected into mutant and WT pools. SNPs unlinked to the mutation will be randomly assorted between the two pools, while linked SNPs will have an OR/MV ratio >1:1 in the mutant pool and lower in the WT pool. Hence, genotyping of the two reciprocal pools for a number of well-spaced SNPs allows one to quickly home in on a genomic region containing the gene of interest. Subsequent genotyping of individual progeny for closely spaced SNPs will further delimit the location of the mutant gene.

The development of a global, high-density SNP map is part of a larger, interinstitutional program intended to make full use of the completed sequence (GALAGAN *et al.* 2003) and take the study of *Neurospora* to the functional genomics level (DUNLAP *et al.* 2007). Like other components of the project, such as gene annotation (M. GALAGAN and B. BIRREN, unpublished results) and targeted gene disruption (COLOT *et al.* 2006), this effort is aimed at community support: the program’s data, strains, and materials are made freely available, and in return users are encouraged to do the same with theirs. To facilitate this exchange, we have created a database allowing for easy browsing of and access to existing polymorphism data, including parameters for experimental verification where relevant, as well as a trove of suspected SNPs that can be selected and validated by the interested user. Such community-based efforts, in combination with future large-scale sequencing projects, have the additional potential to expand the SNP map of *N. crassa*. While positional cloning will remain an important initial step in the characterization of novel genes, it should no longer be rate limiting.

## MATERIALS AND METHODS

**Strains:** *N. crassa* strains used are: wild-type Oak Ridge 74-Oak Ridge23 *mat A* (FGSC 987), wild-type Mauriceville-1c *mat A* (FGSC 2225), *ras-1*<sup>bd</sup> *mat A* (FGSC 1858), *mus51::bar mat a* (FGSC 9718), *mus52::bar mat a* (FGSC 9719), *csp-1 mat a* (FGSC 2555) and *NCU02713*<sup>KO</sup> *mat a* (FGSC 11348). All strains were obtained from the Fungal Genetics Stock Center (Kansas City, MO) and maintained on YPD-supplemented (USBiological) Vogel’s (complete) medium (DAVIS and DESERRES 1970).

Crosses were executed on synthetic crossing medium and ascospores collected and activated by heat shock (DAVIS and DESERRES 1970). Progeny were picked under a dissecting microscope, grown on slants containing complete or Vogel's N minimal medium (DAVIS and DESERRES 1970) at 30° in constant light (LL) for 3 days and scored for conidial separation by tapping slants and observing release of mature conidia, or by examining if the conidial mass could be suspended in water (SELITRENNIKOFF *et al.* 1974).

**Isolation of genomic DNA:** Conidia were inoculated in 14-ml round-bottom tubes (Falcon) containing 3 ml of minimal medium containing 2% glucose and incubated for 36–48 hr at 30° and 200-rpm shaking. Mycelial plugs were collected by filtration, washed, collected in a 96-well plate preloaded with metal beads, and stored at –80°. Frozen tissues were cooled in liquid nitrogen and homogenized by two rounds of shaking (1 min at 30 Hz) using the QIAGEN TissueLyser bead-beater system, and genomic DNA (gDNA) extracted using the QIAGEN MagAttract 96 plant kit as described (COLOT *et al.* 2006). Total yield was 45 µl of solution containing 200–1000 µg/ml gDNA as estimated by OD260.

**cDNA library construction:** Two independent cDNA libraries were constructed. The first was used to derive the set of pSNPs for genomewide validation and made using the following procedure: Mauriceville conidia were resuspended in Vogel's medium + 2% glucose and cultured for 4 hr at 30° in LL. Total RNA was extracted with phenol-chloroform and mRNA purified using the PolyAttract SYSI kit (Promega). One microgram of mRNA was converted into double-stranded cDNA by combining first strand cDNA synthesis as described (CARNINCI and HAYASHIZAKI 1999) with second-strand cDNA synthesis and separated over a CL-2B sepharose-size fractionation column (Sigma). Two fractions with average sizes of ~1.5 and 1 kbp were combined and ligated into *Xho*I/*Eco*RI-cut Uni-Zap XR. λ-Phage was packaged *in vitro* using the Gigapack III kit (Stratagene). The library was amplified by infection of XL1-blue MRF' bacterial strain and mass excised into a phagemid form with a helper phage as described by Stratagene. The phagemids were stably transformed into the SOLR bacterial strain.

Additional SNPs were derived from a second library obtained as follows: Mauriceville conidia were resuspended in Vogel's medium + 2% sucrose and cultured for 7 hr at 34° in LL. Total RNA was extracted using the Trizol method (Invitrogen) and mRNA purified using oligo(dT) cellulose chromatography. Five micrograms of mRNA were converted into cDNA, purified, and ligated into the vector as described above. λ-Phage was packaged *in vitro* using Lambda Packaging Extract (Epicentre) and the library amplified, mass excised, and stably transformed as described above.

**Sequencing and SNP calling:** Sequences were aligned to the Oak Ridge-derived reference sequence (GALAGAN *et al.* 2003) using BLAT v33 (KENT 2002). ESTs that aligned with similar affinity to multiple locations on the reference were discarded. Alignments were retained only if (i) they were the highest scoring alignment for a given read and had a score of at least 50, (ii) the alignment covered at least half of the read length, and (iii) the alignment contained <20% gaps on the read (>20% gaps were allowed on the reference to allow for ESTs that spanned large introns). When both of the paired ends aligned, the sequences were retained on the conditions that they were (i) on opposite strands, (ii) of opposite orientations, and (iii) <100 kbp apart.

Mismatches between aligned EST and genomic sequences were retained when simultaneously fulfilling all criteria of the neighborhood quality standard (NQS) (ALTSHULER *et al.* 2000), defined as: (i) the mismatched base has a Phred score (EWING *et al.* 1998) of ≥25 in the EST sequencing read, (ii) a

window of five bases on either side of the mismatch is perfectly aligned, and (iii) all bases in said window have a Phred score of ≥20. For the library used for genomewide validation, putative SNPs were retained only if all Mauriceville ESTs containing the SNP unanimously agreed. In a modification of the algorithm applied to the second library, putative SNPs were also retained provided there was no more than one dissenting read among three or more reads.

**Selection of CAPS markers:** Putative SNPs were selected that either created or deleted a recognition site for any commercially available four-cutter or degenerate five-cutter (<http://www.neb.com/nebecomm/products/category1.asp>). For each of these, a digestion pattern was created *in silico* on the basis of the positions of the recognition site in a fragment containing 250 bp on either side of the pSNP. CAPS "quality" was defined as the maximum of the size differences between any allele-specific and all non-specific fragments. "Clusters" were defined as follows: for each contig, pSNPs were evaluated in their physical order so that (i) the first pSNP was assigned to the first cluster, and (ii) each subsequent pSNP was assigned to the current cluster if it was closer than 10 kbp from the previous pSNP, and assigned to a new cluster if it was not. The highest-quality CAPS-amenable pSNP from each cluster was identified and experimentally validated; where the validation failed, the process was reiterated with the nonconfirmed pSNP removed from the cluster. Primers were designed using Primer3 (<http://frodo.wi.mit.edu>, default parameters) to reflect the quality prediction: a 500-bp window centered on the pSNP was defined as the target region, and 50 additional bp on either side were included to create the region used for selecting primers. If the program failed to find a suitable primer pair the target region was reduced manually in small increments until one was found. If, upon experimental validation, the primer pair failed to give a product of the desired length a different pair was designed for either the same or a nearby pSNP.

**CAPS assay:** All markers were validated using identical PCR and digestion conditions. One hundred nanograms of gDNA (for Oak Ridge and Mauriceville standards) or 1 µl of 10-fold diluted gDNA solution regardless of concentration (for parallel-extracted progeny) were combined with 1 unit of Taq (Roche), 0.25 mM dNTPs, 0.25 µM of each primer, 1× Roche buffer and water to a final volume of 20 µl. Cycling conditions were as follows: denaturation at 94° for 5 min, 35 cycles of 30 sec at 94°, 30 sec at 54°, and 1 min at 72°, followed by 10 min at 72°. The resulting amplified DNA was used without purification. For the restriction digest, 10 µl of the PCR product was mixed with 1 unit of the appropriate enzyme (all from New England Biolabs), 1× enzyme buffer, 0.1 mg/ml bovine serum albumin (if required), and water to a final volume of 20 µl, and incubated for 3 hr at the optimal temperature. An 8-µl aliquot was electrophoresed on 1.8% agarose (in TAE, containing 0.2 mg/liter ethidium bromide) at 100 V and visualized under UV. When the assay was performed on progeny samples, we found it useful to include one well each of Oak Ridge and Mauriceville gDNA for each gel row as internal controls and size standards.

**Bulked segregant analysis:** Randomly chosen gDNA samples ( $n = 24$ ) representing either WT or mutant progeny were quantified by OD260 and separately pooled into equimolar mixes with a total DNA concentration of 20 ng/µl. CAPS assay was performed as above, but the PCR reaction was limited to 30 cycles and restriction digestion extended to 16 hr.

**SNAP assay:** Primer pairs were designed using the Web-Snaper program (DRENKARD *et al.* 2000), available at <http://pga.mgh.harvard.edu/cgi-bin/snap3/websnaper3.cgi>, with default parameters. Composition and cycling conditions for the PCR reaction were performed as suggested by the authors. For



each pSNP a common primer was used, and for each allele of that pSNP two or three allele-specific primers were tested using as template 100 ng of either Oak Ridge or Mauriceville gDNA. PCR reactions were continued for either 25 or 35 cycles. If two primer pairs could be found that distinguished between the alleles at both 25 and 35 cycles, these were chosen and experimental data collected using 30 cycles. Otherwise, two primer pairs that worked at 25 cycles were selected; genotyping was performed at 25 cycles and, if necessary, repeated at 30 or 35 cycles for those DNA samples that failed to give a product for either primer pair at 25 cycles.

**Phenocopying:** An upstream flank containing the entire NCU02713 ORF (1550/1750 bp, with the putative *osp-1* locus 760/960 bp from the 5' end) was amplified from *osp-1* gDNA and annealed with an *hph* cassette and a downstream flank (1390 bp) derived from WT gDNA (summarized in Figure 6A) using yeast transformation (COLOT *et al.* 2006). The construct was subcloned in *Escherichia coli*, digested with *SbfI*, and transformed into *mus-51* or *mus52::bar*. Hygromycin-resistant primary transformants were isolated and homokaryonized by backcrossing to the WT strain *ras-1<sup>hd</sup> mat A*. Hygromycin-resistant progeny were transferred to Vogel's minimal medium and the conidial phenotype determined by tapping and suspension tests (Figure 6B).

**Northern blotting:** Total RNA was isolated from wild-type (74A) and *ras-1<sup>hd</sup>* (BELDEN *et al.* 2007) cultures grown under standard circadian conditions using hot phenol extraction and 15 µg were separated on a 1.3% formaldehyde gel (LOROS and DUNLAP 1991). RNA was transferred to a Hybond-N+ nylon membrane (Amersham) and a region contained by NCU02713 visualized using a digoxigenin-labeled DNA probe (Roche).

**Data availability:** All pSNPs identified by single-pass sequencing of Mauriceville cDNA are accessible at <http://www.broad.mit.edu/annotation/genome/neurospora> and are searchable by location. The identifier used, NCS.<linkage group>.<seven-digit number>, orders the currently available pSNPs by their physical position along the linkage group and contigs of release 7 of the Neurospora genome (<http://www.broad.mit.edu/annotation/genome/neurospora/Home.html>); however, future additional SNPs, from either further genome-wide efforts or user-submitted targeted sequencing, will be assigned the next available number, regardless of position. Full experimental details for experimentally validated SNPs (the genomewide CAPS map described here, as well as a smaller number of previously unpublished SNAP markers) can be accessed through the same site. Registered users are able and encouraged to submit additional validated SNPs.

## RESULTS

**Sequencing and SNP selection:** Polyadenylated RNA from germinating Mauriceville conidia was isolated and converted into a cDNA library. Single-pass sequencing, starting from both ends of the insert, yielded evidence for 5487 unique clones, and sequences were aligned to the Oak Ridge genome using BLAT (KENT 2002) and checked for internal consistency (see MATERIALS AND METHODS).

EST sequences that could be successfully aligned contained 742 bp on average for a total of ~4 Mb in which 38,400 mismatches were detected. This number is consistent with previous estimates of 0.2–2% sequence divergence in coding regions between Oak Ridge and Mauriceville (DILLON and STADLER 1994; NELSON *et al.*

1997; COULTER and MARZLUF 1998). Potential mismatches were assessed using the NQS algorithm (ALTSCHULER *et al.* 2000), which takes into account the sequencing quality of both the mismatched and neighboring bases, and 4338 high-confidence putative SNPs (pSNPs) were retained (Figure 1A, supplemental online database <http://www.broad.mit.edu/annotation/genome/neurospora/>).

Using the second, independently derived cDNA library, an additional 17,394 pSNPs were subsequently identified. As the original set of 4338 pSNPs proved more than sufficient to obtain a CAPS marker set of the desired density, this set was not included in the systematic validation efforts described below. They are included in the supplemental database so that researchers can use them as a source for pSNPs in their regions of interest. To illustrate this approach, we validated and developed CAPS assays for a number of these while refining the location of our chosen mutation (see below and Figure 1B). Among all pSNPs, transitions outnumbered transversions by ~3:2 (Figure 1C).

The Neurospora genetic map is estimated to span 1000 cM across its seven linkage groups (PERKINS and BARRY 1976), so we estimated ~200 evenly spaced SNPs would be sufficient to establish linkage (<5 cM) with any locus on the genome. Since many pSNPs were close to others and would thus be functionally redundant in the context of recombination analysis, we performed a simple clustering of pSNPs by physical location and then sought to validate a single SNP that would be representative of the cluster within which it resided. Using an arbitrary maximal distance of 10 kbp (for subsequent pSNPs to be considered to belong to the same cluster, see MATERIALS AND METHODS), 424 clusters were defined. Some of these contained only 1 or a few SNPs, which weren't necessarily CAPS amenable (Figure 1B).

Furthermore, given the advantages of CAPS described above, we initially restricted ourselves to confirming pSNPs that could be detected by this assay under standardized conditions. The recognition site for 1 of 10 common four-base cutters (*AluI*, *BstUI*, *HaeIII*, *HhaI*, *MboI*, *MseI*, *NlaIII*, *RsaI*, *TaqI*, or *Tsp509I*) was altered in 1331 pSNPs. Since these enzymes will cut in nearby sites as well, not all SNPs will result in restriction pattern differences that can easily be visualized by standard gel electrophoresis. While programs to identify usable CAPS markers from sequencing data were available (TAYLOR and PROVART 2006), we chose to implement a customized script that selects the "best" CAPS marker (as defined by maximal size difference between allele-specific and any common restriction fragments) from each cluster (Figure 2A). Rounds of pSNP selection were alternated with rounds of experimental validation, and clusters were resampled only when the predicted optimal pSNP representing the cluster failed to be confirmed. Failure to confirm a pSNP does not necessarily reflect inaccurate base

calling; instead, it might be due to insufficient amplification of the surrounding fragment by the chosen primer set, or by a failure of the enzyme to completely digest a specific sequence even when the formal recognition site is present (observed in  $\sim 10\%$  of cases). The majority of CAPS markers in the finished set utilize 1 of the set of 10 reliable and inexpensive enzymes, and result in maximal size difference ( $\sim 560$  bp for the uncut *vs.* two  $\sim 280$ -bp fragments for the cut allele). Additionally, each allele produces a single band of roughly equal intensity given equal amounts of PCR fragment, simplifying quantification of allele frequencies in complex mixtures (see below). To maximize coverage, a limited number of CAPS markers were defined that use addi-

tional enzymes or result in more complex restriction patterns (supplemental Table S1).

All the putative and validated SNPs in this work pertain to the Mauriceville-1c (*mat A*, FGSC 2225) strain. The Mauriceville-1d (*mat a*, FGSC 2226) strain was obtained in the same geographical location but is an independent isolate that we have found is *not* isogenic with FGSC 2225. Hence, to use the SNP data presented here it is necessary to start from a *mat a* version of the mutant strain and to cross it to FGSC 2225. Preliminary data on a subset of validated SNPs has shown that only  $\sim 30\%$  are of the Mauriceville-1c type in the FGSC 2226 strain (C. SCHWERDTFEGER, J. C. DUNLAP and J. J. LOROS, unpublished observation). Likewise, the presence of OR-type alleles at all sites is only guaranteed for the 74-OR23-1VA strain (FGSC 2489), which was the source of the archival genome sequence. Mutants derived from this strain need to be outcrossed to a *mat a* strain first to obtain a mutant strain which can be crossed to Mauriceville *mat A*. While we did not specifically assay our SNP marker set in any additional strains, the 74-ORS-6a strain (FGSC 4200) was derived from a long series of recurrent backcrosses to 74-OR23-1VA and is generally assumed to be highly isogenic to the latter (PERKINS 2004). However, since loss of markers remains possible, we recommend including both parental controls when genotyping progeny for each individual SNP.

**SNP validation and discovery:** pSNPs were validated using the CAPS protocol described in MATERIALS AND METHODS (Figure 2A). A total of 515 primer pairs were tried on Oak Ridge and Mauriceville gDNA during alternating rounds of validating and updating the list of candidate CAPS markers. Occasionally, we observed significantly lower amplification efficiency on the Mauriceville-1c template, possibly due to additional mismatches in the genomic region binding the primer. Of the successfully amplified fragments,  $\sim 80\%$  yielded the

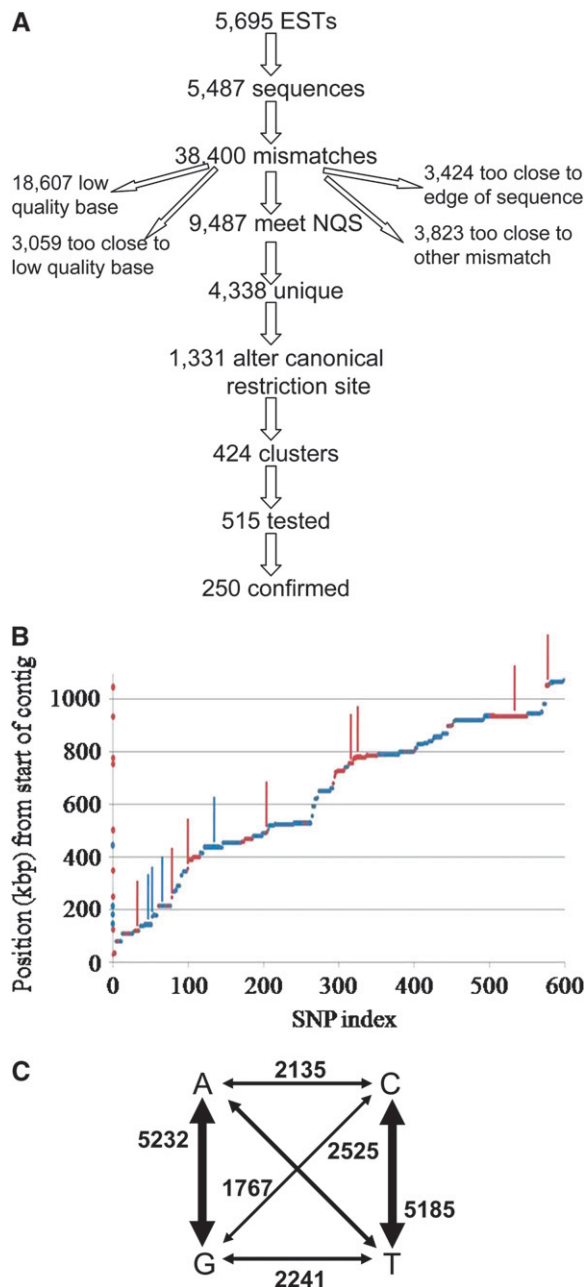


FIGURE 1.—SNP discovery. (A) Of all sequenced MV ESTs, 5487 aligned consistently with the OR genome, and these contained 38,400 mismatches. Of these, 9487 met NQS criteria (see MATERIALS AND METHODS); the remainder was excluded for the reasons shown. Because of redundancy in the library, these corresponded to 4338 unique putative SNPs (pSNPs), of which 1331 alter 1 of 10 common four-base restriction sites. These were divided over 424 clusters (see MATERIALS AND METHODS), and at least one pSNP in each cluster was tested until 250 were unambiguously confirmed. (B) Distribution of 599 putative SNPs along contig 6. Position from the start of the contig (in kilobase pairs) is plotted as a function of pSNP index (rank order of the pSNP's physical position along the contig). Note the long horizontal stretches showing groups of closely linked and thus functionally redundant SNPs. Red dots denote pSNPs included in genomewide validation; blue dots denote other pSNPs; red bars (projected on y-axis for clarity) mark confirmed SNPs from genomewide validation; blue bars mark confirmed SNPs from regional validation. (C) Frequency of transition and transversion events (based on complete set of 17,394 putative SNPs).

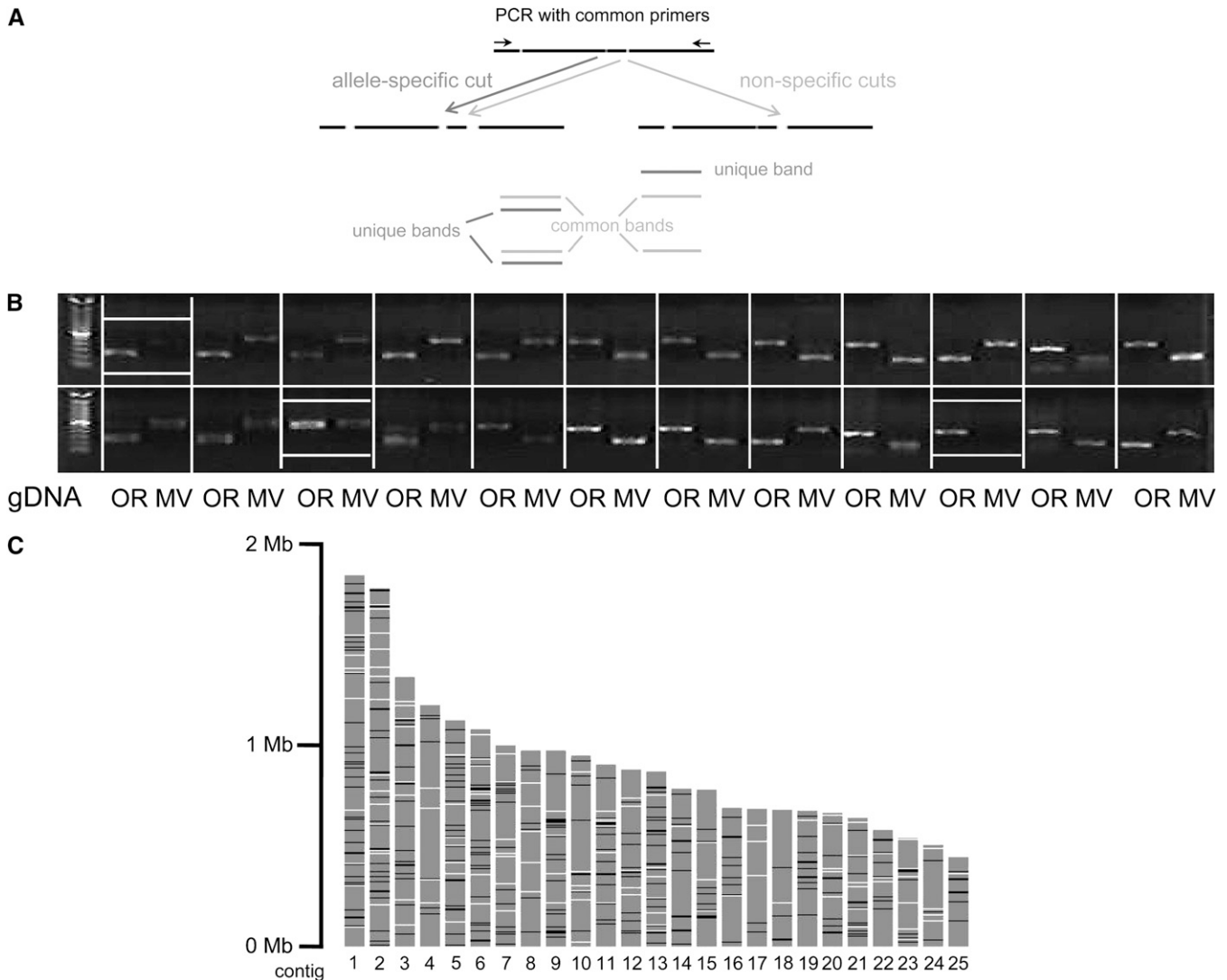


FIGURE 2.—SNP validation. (A) Principle of cleaved amplified polymorphic sequence (CAPS) assay. A genomic region containing an allele-specific restriction site (dark gray, present in the left and absent in the right allele) and possibly common restriction sites (light gray) is amplified by PCR with a common pair of primers. Subsequent digestion produces several fragments, some of which are unique to one allele. These can be visualized by gel electrophoresis. (B) Representative gel for SNP validation. Digested PCR fragments obtained from Oak Ridge (OR) and Mauriceville (MV) are loaded side by side for 24 pSNPs. Common negative results include failure to efficiently amplify from MV gDNA and absence of the expected extra site in MV (boxed). (C) SNP coverage in the 25 largest contigs (shaded bars). Available EST sequence (used for genome-wide validation) is denoted by open and validated CAPS markers by solid bars (resolution 5 kbp).

expected pattern when subjected to digestion (Figure 2B). Assuming that the Oak Ridge archival genomic sequence is accurate and ectopic digestion does not occur, deviation from the expected pattern can be explained from either enzyme dysfunction or erroneous base calling in the EST sequencing. The frequency of the latter was estimated to be up to 5% (although no further effort was made to examine pSNPs that were not confirmed in the first round). In contrast to the archival genome sequence of the reference strain, many of these pSNPs were predicted on the basis of single-pass sequencing, so even when steps are taken to eliminate some sequencing artifacts, independent experimental validation remains necessary.

The finished CAPS map contains 250 markers (details in supplemental Table S1), distributed roughly equally over the physical map. Given estimates of the total genetic length of the *N. crassa* genome ranging from 500 (JIN *et al.* 2007) to 1000 (PERKINS and BARRY 1976) MU, the SNP map can be expected to contain several markers linked to any given genetic locus. This number is close to the saturation level for this cDNA library, for the chosen assay (Figure 2C). A graphical summary of both EST coverage and location of confirmed SNPs, relative to physical and genetic locations of known and unknown genes and markers, is provided in Figure 3 for the left arms of linkage groups I and II, and in supplemental Figure S3 for the whole genome.



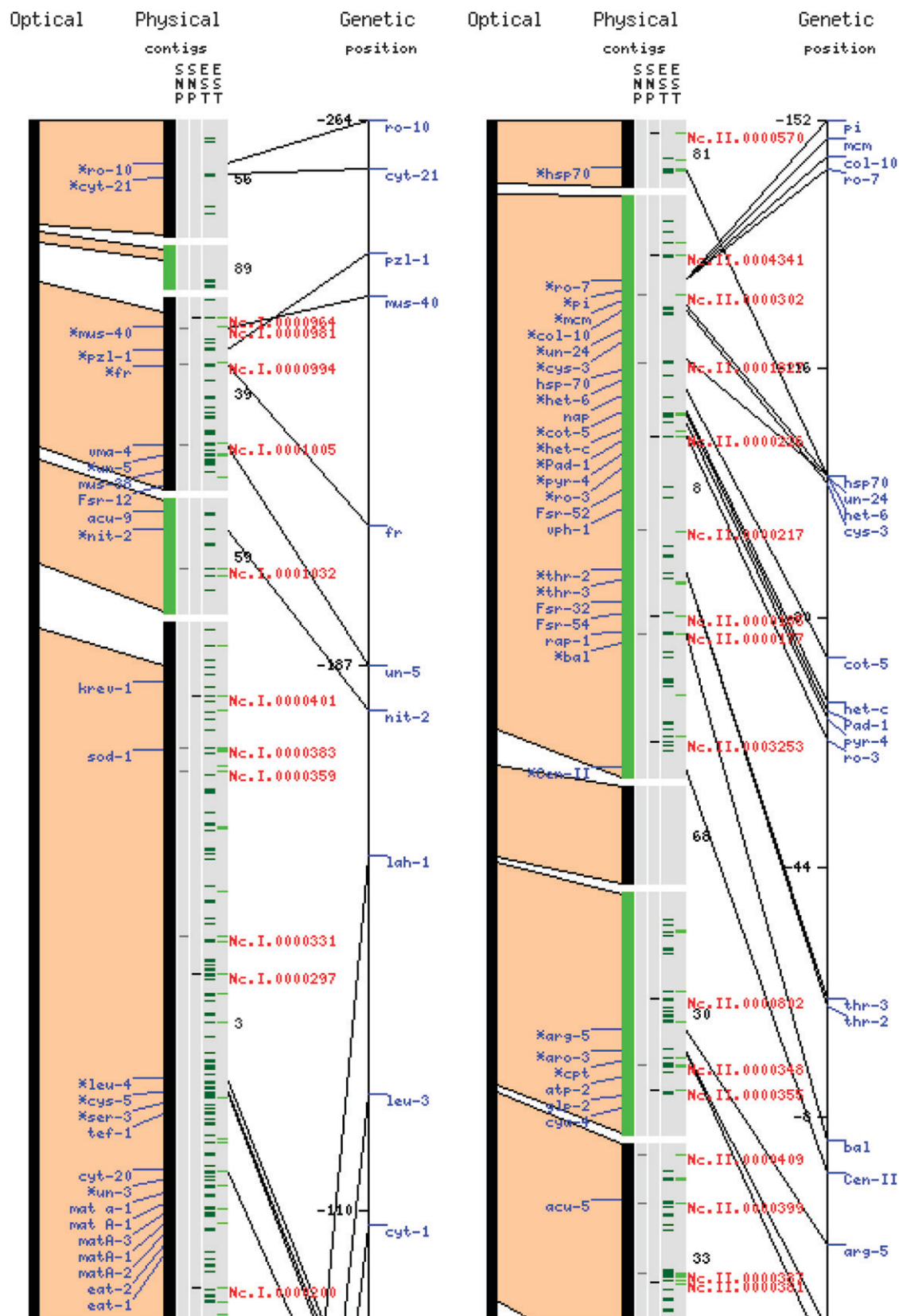


FIGURE 3.—Distribution of sequenced ESTs and CAPS markers. Optical, physical, and genetic maps of the left arms of *N. crassa* linkage groups I and II. In the gray bars marked “EST,” bright green dashes represent 3-kbp bins in which EST sequences used for genomewide validation were obtained; dark green dashes, idem for additional EST data. In the bar marked “SNP” red dashes represent confirmed CAPS markers, black dashes confirmed CAPS markers that were selected for inclusion in the master plates. Exact location, primer sequences, and restriction enzymes are given in supplemental Table S1.

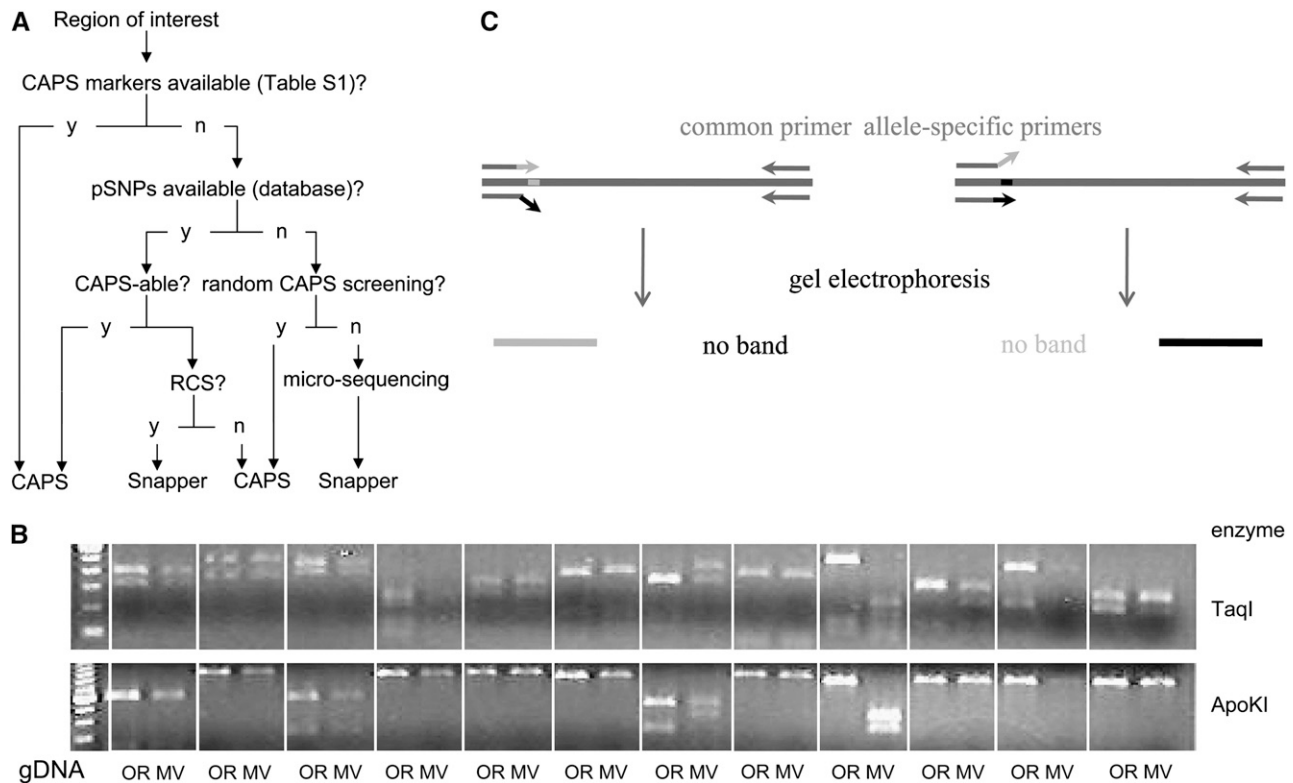


FIGURE 4.—Development of local SNP markers. (A) General approach for efficient marker development. The first choice for markers in the region of interest consists of the set of validated CAPS markers described in supplemental Table S1. When these are exhausted but pSNPs are available, these can be developed as either CAPS or single nucleotide amplified polymorphism (SNAP) markers. Finally, additional polymorphisms can be found by random CAPS screening (RCS) or sequencing of randomly chosen PCR fragments. (B) Representative gel for random CAPS screening. Amplicons were prepared from 12 intergenic regions, using Oak Ridge (OR) or Mauriceville (MV) gDNA and digested with different enzymes. Note more irregular nature of digestion patterns compared to CAPS markers found by genomewide validation (Figure 2B). (C) Principle of SNAP assay. Two PCR reactions are carried out, using a common primer and either of two allele-specific primers. The latter are designed to be perfectly complementary to one allele but not the other, so that genomic DNA containing a specific allele is amplified only in the reaction employing its respective primer. Presence or absence of PCR products is visualized by gel electrophoresis.

We ranked marker types into tiers (summarized in Figure 4A) to yield a strategy that would increase local SNP coverage as efficiently as possible, moving to a lower tier only when the higher one was considered exhausted for the region of interest, as described below:

- i. Identification of CAPS markers on the basis of genomewide sequencing: these include the 250 markers discussed above, which allow assignment of the locus being mapped to the 10- to 20-cM gap between the two closest markers, as well as identification of progeny with informative crossover events. By using the expanded set of 17,394 pSNPs as well as relaxing requirements for digestion patterns and enzyme choice, additional markers could be defined that are nonredundant on a regional scale.
- ii. Identification of CAPS markers on the basis of random screening: additional CAPS markers could be designed from pSNPs picked up by local sequencing of MV gDNA. However, we found it more convenient to use random CAPS screening (RCS) to identify SNPs in a defined genomic region. We

amplified ~800-bp fragments of intergenic MV gDNA using PCR (as described in MATERIALS AND METHODS), subjected them to digestion with a battery of the 10 most common restriction enzymes, and compared the digestion pattern to that obtained from OR gDNA. While the exact nature and position of the causative SNP is not determined in this procedure, establishing the location of the marker at a 1-kbp resolution is sufficient for mapping work. We found that about half (8/16) of such amplified fragments provide usable CAPS markers (Figure 4B). Established CAPS markers are easily assayed in a robust fashion. Since the PCR primers are perfectly complementary to the template sequence, amplification is quite insensitive to the quality of individual gDNA samples. In contrast, in our hands, SNAP markers gave inconclusive results for some progeny under standard conditions, even when they seemed robust when initially assayed on bulk-prepared reference Oak Ridge and Mauriceville gDNA (Figure 4C). (It should be noted,



however, that only two or three primer pairs per allele were tested, as opposed to up to eight in the original publication; DRENKARD *et al.* 2000). Although these could be resolved by repeating the reaction with different cycle numbers, the additional workload and ambiguity compelled us to use SNAP markers only when CAPS-amenable SNPs could not be detected. In this case the following options can be explored:

- iii. Identification of SNAP markers on the basis of genomewide sequencing: pSNPs picked up by cDNA sequencing but not suitable to CAPS can be developed into SNAP markers. As described in MATERIALS AND METHODS, a number of Oak Ridge- and Mauriceville-specific primers must be tried out at different cycle numbers to determine the ideal set of conditions (Figure 4D).
- iv. Identification of SNAP markers on the basis of microsequencing: finally, to cover the gaps between EST reads that did not provide CAPS markers using approach ii, the same 800-bp fragment and matching primers can be recycled for sequencing and pSNPs developed into SNAP markers.

We found that for the region containing *csp-1* and for our desired mapping resolution the first two, CAPS-based, approaches provided a sufficient number of markers when applied to saturation. The pSNPs derived from the second, expanded library (containing 17,394 pSNPs) were also confirmed and developed into markers in the same way and with similar success rate as described above; however, to maximize the number of CAPS-amenable SNPs we allowed every enzyme commercially available from New England Biolabs (listed at <http://www.neb.com/nebecomm/products/category1.asp>) to be considered. After the search for CAPS markers using this approach was exhausted, we filled in the remaining gaps by random screening (option ii above). If several enzymes resulted in restriction pattern polymorphism, the one containing the most distinctive differences was chosen for subsequent genotyping. Incidentally, two of the tested PCR fragments (both near the centromere) showed a noticeable size difference (~50 bp) prior to any digestion, due to insertions in MV relative to OR. In general, the nature and distribution of regional markers will be dictated by the required resolution and depend on the available local genomic variation.

**Proof of principle: recombination-based mapping of *csp-1*:** We used the SNP approach outlined above to map and subsequently clone the causative mutation leading to the conidial separation defect that defines the *csp-1* gene. First, we confirmed the location of this gene on a specific region of LGI (SELITRENNIKOFF *et al.* 1974) using BSA. We then genotyped individual progeny containing a recombination in this region for increasingly closer spaced SNPs until a sufficiently small candidate region could be delineated. For the BSA experiment,

we crossed FGSC 2555 (*csp-1 mat a*) to FGSC 2225 (Mauriceville-1c *mat A*) and picked 443 progeny. These were scored using the tapping assay and each individual progeny was used as a source of gDNA. We combined the independently extracted gDNA samples into equimolar mixtures composed of DNA from either all WT or all mutant progeny, and subjected these pools to the same assay as used on gDNA obtained from individual progeny. While this method reliably distinguished SNPs that were linked from those that were unlinked, some caveats are in order. First, as noted above, some primer pairs amplify the MV allele less efficiently than the OR allele although similar concentrations of each template are used in the PCR assay. While these primer pairs can be used for genotyping individual progeny, they should be avoided when performing BSA experiments. We also noted a consistent bias toward the allele refractory to restriction enzyme digestion (the “undigested allele”) for closely linked SNPs (*e.g.*, 6.773, 6.357, and 72.66 in Figure 5A): while the undigested allele was visible in the pool in which it was the minority, this was not the case for the digested allele in the reciprocal pool. One possible explanation for this phenomenon is that in the last stages of PCR amplification, when supplies of primers and/or dNTPs become limiting, some of the amplified fragments start to undergo cycles of melting and reannealing, effectively sequestering the low levels of digestible allele into indigestible heteroduplexes. By optimizing conditions for a few SNPs using mixes of parental gDNA we found that this bias can be reduced, but not completely eliminated, by lowering the number of elongation cycles and template concentration (data not shown). More importantly, one should perform BSA using both mutant and WT pools. Since the direction of experimental bias will be opposite in the reciprocal pools, a semiquantitative measure for linkage to any given SNP can then be obtained in the form of the map ratio, defined as  $[\text{OR, WT}] \times [\text{MV, mutant}] / [\text{OR, mutant}] \times [\text{MV, WT}]$  (based on relative intensities of allele-specific bands in the respective pools), which will vary between 0.5 for unlinked to 0 for perfectly linked markers (WICKS *et al.* 2001) (Figure 5B).

Of eight markers evenly spaced over LGI, those localized at the end of contig 6 and on contig 72 are the most tightly linked to the phenotype (Figure 5A). Since these two SNPs surround the known physical locations of *arg-3* and *nuc-1* (Figure 5C), which in turn are known to be on opposite sides of *csp-1* genetically, we chose the enclosed region as the smallest physical interval which, with certainty, contained the *csp-1* locus. Of 255 WT and 188 mutant progeny obtained from a single *csp-1*  $\times$  Mauriceville cross, 25 showed a recombination event between markers at 6.564 and 72.66 (SNPs 1 and 10, respectively, in Figure 5C), corresponding to a genetic distance of 6 cM for a physical distance of >1 Mb. This ratio is relatively low, and is likely due to the presence of the centromere suppressing recombination

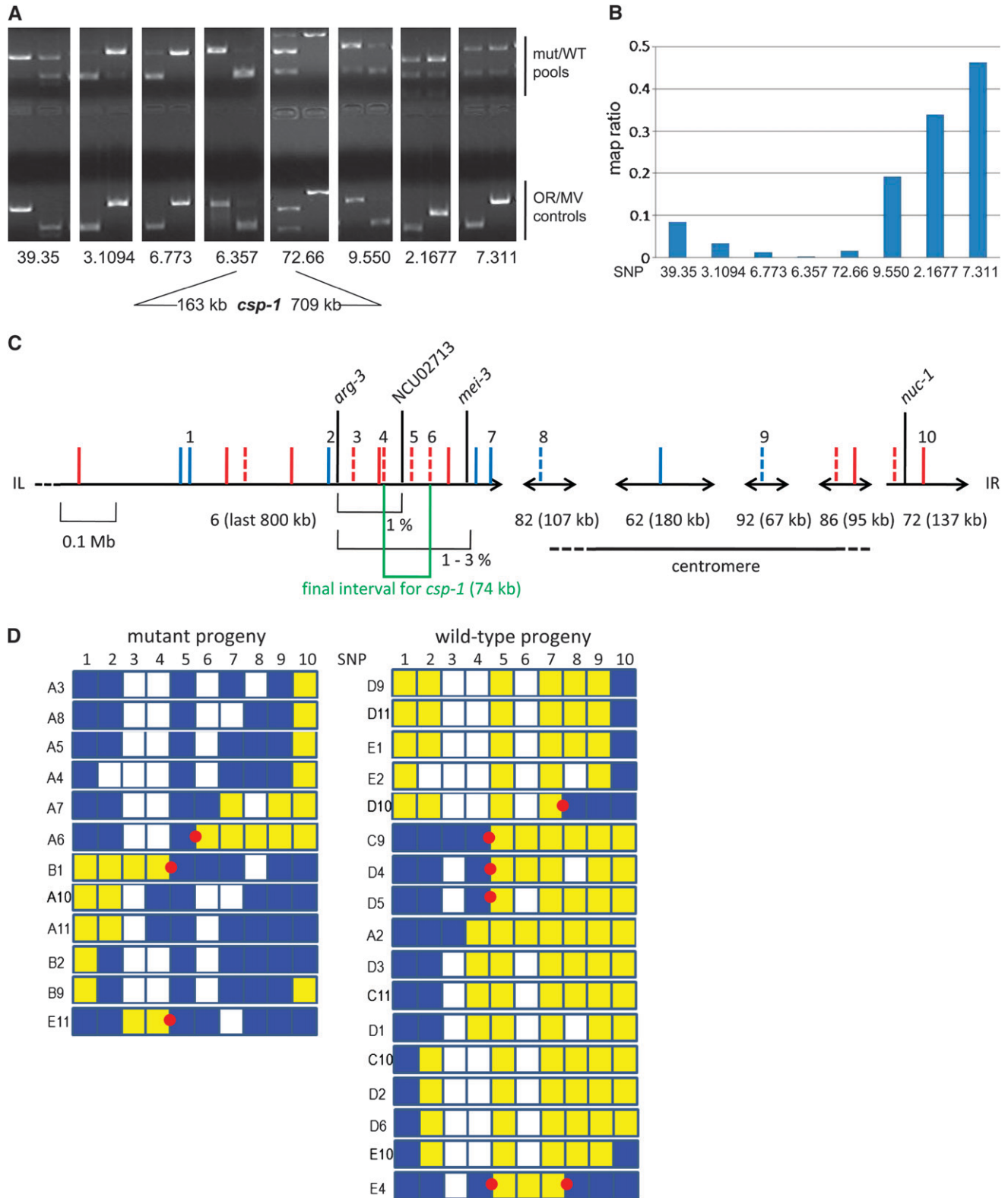


FIGURE 5.—Mapping of *csp-1*. (A) Local bulked segregant assay. CAPS assay performed on 8 markers equally spaced on LGI (physical location given as (contig).(distance from start of contig in kb)). Odd lanes, mutant pools, and OR control, even lanes, WT pool, and MV control. (B) Map ratios (see text) estimated from band intensities for SNPs in A. (C) Summary of markers used for interval mapping of *csp-1*. Solid red bars, SNPs from genome-wide validation; dashed red bars, SNPs from regional validation; solid blue bars, SNPs from random CAPS screening; dashed blue bars, AFLPs discovered during random CAPS screening. SNPs used are numbered 1 (6.564), 2 (6.322), 3 (6.269), 4 (6.212), 5 (6.173), 6 (6.139), 7 (6.17), 8 (82.9), 9 (92.50), and 10 (72.66). (D) Interval mapping. Genotype for selected mutant (left) and WT (right) progeny for SNPs numbered as in C. Blue, Oak Ridge; yellow, Mauriceville; white, n/d; red dots, most informative recombination events.

in between the two. Alternatively, naturally occurring polymorphisms might reduce recombination frequency in specific chromosomal regions (CATCHESIDE 1981; BOWRING 2005). Using this subset of informative recombinants, we alternated between developing new markers (as described above) in the region of interest and scoring the recombinant progeny for these markers, thereby reducing the region of interest to the interval contained by the two markers that remain the most tightly linked to the phenotype (Figure 5C). This approach led to the identification of the 74-kbp region contained between SNPs at 6.139 and 6.213 as the smallest interval to which the mutation could be localized using this set of progeny.

This region contains a total of 20 predicted genes, a sufficiently small number to identify a candidate gene among them. From microarray work (C. H. CHEN, C. S. RINGELBERG, R. H. GROSS, J. D. DUNLAP and J. L. LOROS, unpublished results), the putative Zn-finger transcription factor NCU02713.3 had been identified as a strongly light-inducible gene. Because conidiation is induced by light, we examined the possibility that NCU02713.3 encodes *csp-1*. Sequencing of the *csp-1* strain revealed a G-to-A mutation at position 194,611 relative to the start of contig 6, resulting in the substitution of Cys139 by tyrosine.

To exclude the possibility that the *csp-1* phenotype is due to a different, nearby mutation, we sought to generate strains that would differ only at this locus. Since the inability of the *csp-1* strain to form individual conidia made transformation into this strain difficult, we used an alternative but equivalent approach aimed at replicating the mutant phenotype in a WT background. Gene replacement in the endogenous locus (summarized in Figure 6A) followed by homokaryonization provided recombinant strains bearing either the WT or the mutant phenotype (Figure 6B), which correlated with the presence of respectively the WT and the mutant allele by CAPS assay (Figure 6B).

Since the mutated cysteine is part of one of two signature CX1-5CX12HX3H sequences required for zinc binding and proper function of the DNA-binding domain in this class of transcription factors (IUCHI 2001), the mutation can be expected to result in a complete loss of function. Indeed, we found the *Neurospora* Functional Genomics Consortium (COLOT *et al.* 2006) knock-out strain *NCU02713<sup>ko</sup>* to fully mimic the conidial separation phenotype and to be visually indistinguishable from the original *csp-1* strain, a previously unreported observation (Figure 6B).

Mapping a mutation to a physical location allows the deployment of molecular tools to further explore the phenotype. We asked whether the well-known circadian regulation of conidiation (DAVIS and PERKINS 2002) would extend to transcript levels of this gene when assayed in liquid culture. By Northern blotting of wild-type RNA we found that the *csp-1* message is robustly rhythmic, with peak levels around subjective dawn, co-

inciding with maximal formation of new conidia (Figure 6C). In addition, both baseline and peak levels are significantly elevated in the *ras-1bd* strain (BELDEN *et al.* 2007), consistent with this strain's ability to conidiate rhythmically under a much wider range of conditions than WT. Thus, SNP mapping mutations is a rapid and reliable method to identify a relevant gene and opens the way to further exploration of a phenotype of interest.

## DISCUSSION

*N. crassa* has a long and distinguished history as a model organism (reviewed in DAVIS and PERKINS 2002), a status it owes in large part to the early commitment made by its researchers to genetic tractability. The groundbreaking work of Beadle and Tatum on nutritional mutants (BEADLE and TATUM 1941) was one of the first studies to bridge the gap between genetics and biochemistry, initiating the integrative approach that has since become the standard in experimental biology. A variety of important biological phenomena have been either described for the first time or further elaborated using *Neurospora*, and it has become the best understood of the filamentous fungi, a large group of organisms including several of tremendous economical and medical importance. Much of this progress has been initiated by the use of forward genetics. For example, after many years of mainly descriptive work on circadian rhythmicity, the positional cloning of the *frequency* gene (MCCLUNG *et al.* 1989) initiated a cascade of molecular work which, over a relatively short span of time, led to a real, mechanistic understanding of the *Neurospora* circadian clock and paved the way for understanding rhythms in other organisms, including humans.

The availability of full-genome sequences for an increasing number of model organisms has allowed for new and powerful approaches that could only be dreamt of a decade ago. *Neurospora* entered the genomic age in 2003 (GALAGAN *et al.* 2003) and this landmark development has initiated several large-scale functional genomics projects (reviewed in DUNLAP *et al.* 2007) that have rapidly contributed to our understanding of *Neurospora* biology. However, a large role remains for classical forward genetics in the functional classification of novel genes. *N. crassa* contains a total of 9826 predicted genes (<http://www.broad.mit.edu/annotation/genome/neurospora>). Of these, 41% do not have significant homology to any known gene, illustrating the potential for uncovering new and interesting biology specific to filamentous fungi. Moreover, over the course of more than half a century of genetic screens, a total of 3441 single mutants, representing over 1700 loci, have been accumulated (MCCLUSKEY 2003), many of which have not yet been identified on the gene level nor correlated with a molecular function. The availability of the reference genome sequence and



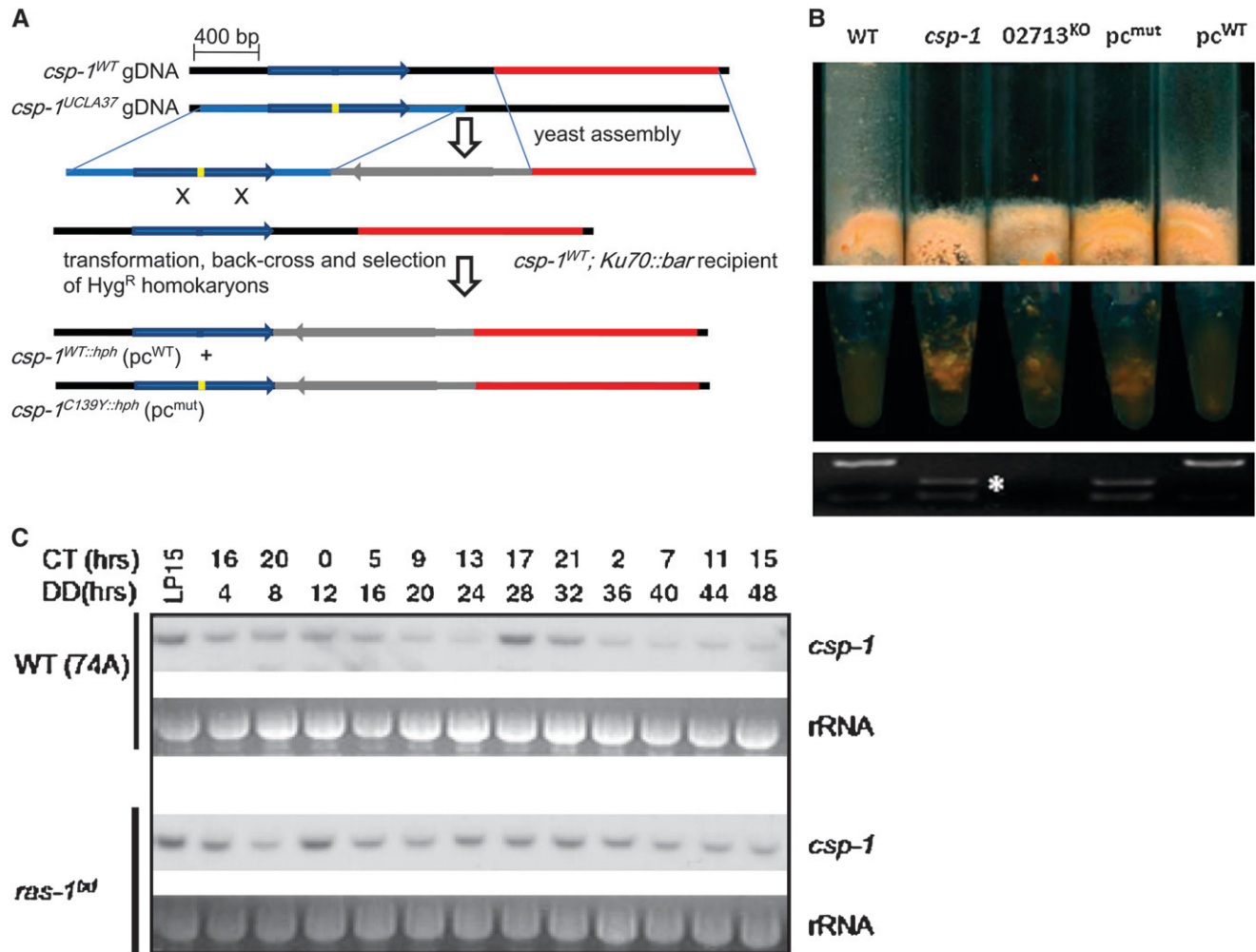


FIGURE 6.—Identification and characterization of *csp-1*. (A) Schematic of gene replacement at the endogenous locus. Blue arrow, NCU02713 ORF; blue line, left flank; red line, right flank; gray line and arrow, hygromycin-resistance cassette containing *hph* ORF; blue and yellow bars, WT and C139Y allele, respectively, at 6.194611. The resulting strains are *csp-1*<sup>WT::hph</sup> (*pc*<sup>WT</sup>) and *csp-1*<sup>C139Y::hph</sup> (*pc*<sup>mut</sup>), which differ only at the Cys139 locus of the NCU02713 ORF. (B) Phenocopying of *csp-1* by NCU02713<sup>KO</sup> and *pc*<sup>mut</sup> but not *pc*<sup>WT</sup>. Strains, including parental controls, were grown for 3 days at 30 LL on minimal medium and tapped sharply (top) or had their conidia transferred into water (middle). (Bottom) Genotype at the *csp-1* locus. The C139Y mutation creates an additional *RsaI* restriction site (asterisk). (C) Representative Northern blots of *csp-1* mRNA over a 48-hr time course, in WT (top) and *ras-1*<sup>hd</sup> (bottom) background. Both sets were performed on the same membrane and with identical exposure times.

homology-based annotation models greatly facilitates the identification of candidate genes; however, the initial step in identifying the molecular alteration underlying a phenotype of interest will remain, for some time to come, the determination of the approximate physical location of the mutant locus by genetic mapping.

Recombination-based mapping is the process of following an unknown genetic locus, via its phenotype, as it cosegregates with markers of known physical location. Traditionally, these markers were mutant genes with their own phenotypic effects (*e.g.*, specific auxotrophies) (PERKINS *et al.* 1982), requiring a qualitatively different assay for each tested marker. As the number of phenotypic markers in any given strain (including specially developed linkage tester strains) is limited, a large number of crosses were often required

to find closely linked markers. For these reasons, the use of phenotypically neutral polymorphisms is preferred, as these are much more abundant and occur all in a single strain. PCR-based marker sets have been developed (KOTIERK and SMITH 2004; JIN *et al.* 2007); these are sufficiently dense to establish linkage of a mutant locus to a chromosome arm but not to limit the search to a manageable number of candidate genes. These sets were developed either by trial and error or from a limited amount of sequencing data and thus cannot be easily expanded to include a higher density of confirmed markers in a region of interest. Recently, a microarray-based method for mapping in *Neurospora* (restriction site associated DNA or RAD) has been described (LEWIS *et al.* 2007). This technique allows the mapping of an unknown gene to a relatively small

region with unprecedented ease and is independent of the presence of OR-type markers in the mutant strain. However, development of additional RFLP markers is often still necessary to define a region small enough to identify candidate genes. The use of this and similar high-throughput techniques is certain to increase in the future, but a role will remain for simple genotyping methods using inexpensive and ubiquitously available reagents and equipment.

In this work we have used MV-derived EST libraries for limited high-throughput sequencing of a nonreference strain, followed by a computational filtering method designed to eliminate the majority of false positives arising from sequencing errors. Using two independently constructed libraries, we identified nearly 20,000 unique putative SNPs. We then applied experimental validation to the smaller of the two sets, yielding CAPS-based assays for 250 confirmed SNPs, more than sufficient to contain several markers that are sufficiently closely linked to any given locus in the genome. We selected putative SNPs and designed the validation assay in such a way that a single set of parameters can be used for genotyping any given marker, allowing for the rapid genotyping of multiple marker/progeny combinations in a single experiment. Assays were designed to optimize the visual readout of the experiment (size difference on agarose gels), facilitating the interpretation of experiments using pooled progeny sets. The general approach of limited sequencing of a crossing strain followed by experimental validation of carefully selected pSNPs is applicable to any organism with a complete reference genome.

Both the set of validated CAPS markers and the trove of unconfirmed SNPs were used to localize the *csp-1* mutation to a 74-kbp region on LGI. Since CAPS markers are codominant, the use of BSA (MICHELMORE *et al.* 1991) is a rapid way to determine a rough first approximation of a gene's location. To encourage the use of this technique, we have selected a subset of robust CAPS markers (boldface type in supplemental Table S1 and black in Figure 3 and supplemental Figure S3) roughly evenly spaced over the genome, and grouped their respective primer pairs in a convenient 96-well format (supplemental Table S2); these "master plates" are available through the Fungal Genetics Stock Center (<http://www.fgsc.net>) at cost. Assaying WT and mutant pools for all of these SNPs will, for any genetic location of the mutation, identify at least two contiguous markers significantly linked to the locus and thus allow for its assignment to a chromosome arm in a single experiment. Since earlier work using phenotypic markers (SELITRENNIKOFF *et al.* 1974) had localized *csp-1* to the left arm of LGI, we restricted ourselves to analyzing a few markers on each chromosome arm, and as expected found only those on LGIL tightly linked (data not shown). We scanned additional CAPS markers distributed over LGI and found two markers on either side

of the centromere that were tightly linked to the mutation and circumscribed the genetic markers *arg-3* and *nuc-1*, on either side of *csp-1* on the genetic map (SELITRENNIKOFF *et al.* 1974; PERKINS *et al.* 2001). Hence using as little as 24 WT and mutant progeny the mutant locus was delineated to a 5- to 10-cM region, bounded by markers that can subsequently be used to genotype larger sets of individual progeny and identify informative recombinants.

The *csp-1* mutation is located near a centromere; these regions are typically characterized by low recombination rate and low gene abundance. This necessitated the isolation of a larger-than-usual number of progeny, arguably the rate-limiting step in the mapping procedure. Once progeny were phenotyped and gDNA obtained, genotyping the delimiting CAPS markers quickly allowed for the selection of a more manageable number of progeny with informative recombination events. The low density of genes in this region meant markers developed from ESTs were of limited use; however, this limitation gave us the chance to explore alternative routes toward increasing local SNP density. Briefly, we started by exhausting putative SNPs discovered from the larger cDNA library and then proceeded to screening of intergenic fragments derived from this region for polymorphisms altering restriction by a battery of common four cutters. We also stumbled upon two instances of amplified fragment length polymorphisms (AFLPs) due to large (>50-bp) insertion/deletion events; while we expect these to be rare outside highly repetitive regions, genotyping for these markers requires only a single amplification step and is highly robust. Finally, while we did not make use of SNPs for the identification of *csp-1*, we have successfully used this technique in other mapping efforts (BELDEN *et al.* 2007) and consider it a useful addition to the arsenal of genotyping techniques, especially in regions or organisms with low sequence divergence where SNPs that alter restriction sites might not be obtainable. In summary, using a combination of validated SNPs with *de novo* marker development, any region on the *Neurospora* genome can be rapidly populated with SNP markers of sufficient density to allow the localization of a mutation of interest with a precision that is only limited by the local recombination rate and the amount of progeny one is willing to isolate and genotype.

We report the discovery and validation of a new set of SNPs for *N. crassa*, and outline an approach for constructing polymorphism maps that can be adapted to other organisms for which the reference genome sequence and a sufficiently divergent tester strain is available. We also report the identification and initial characterization of the developmental regulator, *csp-1*. Asexual development in *Neurospora* and other filamentous fungi is a complex and incompletely understood process, integrating signals from the circadian clock, metabolic state, and various types of external and

internal stress (TURIAN and BIANCHI 1971). This developmental process involves multiple intracellular signaling events including MAPK (PANDEY *et al.* 2004), cAMP (BANNO *et al.* 2005), and Ras-dependent pathways (BELDEN *et al.* 2007). We identified a binuclear zinc-finger transcription factor as a downstream regulator of conidiation (similar to another conidiation gene, *fluffy*; BAILEY and EBBOLE 1998), which is regulated at the transcriptional level in a time-of-day and Ras-dependent manner.

We thank Philip Montgomery, Reinhard Engels, and Mike Koehrsen for their valuable contributions. This work was supported primarily by National Institutes of Health (NIH) grant GM068087. J.C.D. and J.J.L. were also supported by GM083336 and J.C.D. by GM034985. W.J.B. is funded in part by a Ruth L. Kirschstein Postdoctoral Fellowship (GM071223) from NIH. We also acknowledge the support of the Norris Cotton Cancer Center at Dartmouth Medical School and the Fungal Genetics Stock Center, University of Missouri, Kansas City.

#### LITERATURE CITED

- ALTSHULER, D., V. J. POLLARA, C. R. COWLES, W. J. VAN ETTEN, J. BALDWIN *et al.*, 2000 An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* **407**: 513–516.
- BAILEY, L. A., and D. J. EBBOLE, 1998 The *fluffy* gene of *Neurospora crassa* encodes a Gal4p-type C6 zinc cluster protein required for conidial development. *Genetics* **148**: 1813–1820.
- BANNO, S., N. OCHIAI, R. NOGUCHI, M. KIMURA, I. YAMAGUCHI *et al.*, 2005 A catalytic subunit of cyclic AMP-dependent protein kinase, PKAC-1, regulates asexual differentiation in *Neurospora crassa*. *Genes Genet. Syst.* **80**: 25–34.
- BEADLE, G. W., and E. L. TATUM, 1941 Genetic control of biochemical reactions in *Neurospora*. *Proc. Natl. Acad. Sci. USA* **27**: 499–506.
- BELDEN, W. J., L. F. LARRONDO, A. C. FROELICH, M. SHI, C. H. CHEN *et al.*, 2007 The *band* mutation in *Neurospora crassa* is a dominant allele of *ras-1* implicating RAS signaling in circadian output. *Genes Dev.* **21**: 1494–1505.
- BERGER, J., T. SUZUKI, K. A. SENTI, J. STUBBS, G. SCHAFFNER *et al.*, 2001 Genetic mapping with SNP markers in *Drosophila*. *Nat. Genet.* **29**: 475–481.
- BOWRING, F. J., 2005 Recombination in filamentous fungi, pp. 1–32 in *Applied Mycology and Biotechnology: Genes and Genomics*, Vol. 5, edited by D. K. ARORA and R. BERKA. Elsevier, New York.
- CARNINGI, P., and H. HAYASHIZAKI, 1999 High-efficiency full-length cDNA cloning. *Methods Enzymol.* **303**: 19–44.
- CATCHESIDE, D. E., 1981 Genes in *Neurospora* that suppress recombination when they are heterozygous. *Genetics* **98**: 55–76.
- COLOT, H. V., G. PARK, G. E. TURNER, C. RINGELBERG, C. M. CREW *et al.*, 2006 A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. *Proc. Natl. Acad. Sci. USA* **103**: 10352–10357.
- COULTER, K. R., and G. A. MARZLUF, 1998 Functional analysis of different regions of the positive-acting CYS3 regulatory protein of *Neurospora crassa*. *Curr. Genet.* **33**: 395–405.
- DAVIS, R. H., and D. D. PERKINS, 2002 Timeline: *Neurospora*: a model of model microbes. *Nat. Rev. Genet.* **3**: 397–403.
- DAVIS, R. H., and F. J. DESERRRES, 1970 Genetic and microbiological research techniques for *Neurospora crassa*. *Methods Enzymol.* **71**: 79–143.
- DETERA-WADLEIGH, S. D., and F. J. MCMAHON, 2004 Genetic association studies in mood disorders: issues and promise. *Int. Rev. Psychiatry* **16**: 301–310.
- DILLON, D., and D. STADLER, 1994 Spontaneous mutation at the *mtr* locus in *Neurospora*: the molecular spectrum in wild-type and a mutator strain. *Genetics* **138**: 61–74.
- DRENKARD, E., B. G. RICHTER, S. ROZEN, L. M. STUTIUS, N. A. ANGELL *et al.*, 2000 A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in *Arabidopsis*. *Plant Physiol.* **124**: 1483–1492.
- DUNLAP, J. C., K. A. BORKOVICH, M. R. HENN, G. E. TURNER, M. S. SACHS *et al.*, 2007 Enabling a community to dissect an organism: overview of the *Neurospora* functional genomics project. *Adv. Genet.* **57**: 49–96.
- EWING, B., L. HILLIER, M. C. WENDL and P. GREEN, 1998 Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**: 175–185.
- GALAGAN, J. E., S. E. CALVO, K. A. BORKOVICH, E. U. SELKER, N. D. READ *et al.*, 2003 The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422**: 859–868.
- IUCHI, S., 2001 Three classes of C2H2 zinc finger proteins. *Cell. Mol. Life Sci.* **58**: 625–635.
- JIN, Y., S. ALLAN, L. BABER, E. K. BHATTARAI, T. M. LAMB *et al.*, 2007 Rapid genetic mapping in *Neurospora crassa*. *Fungal Genet. Biol.* **44**: 455–465.
- KENT, W. J., 2002 BLAT—the BLAST-like alignment tool. *Genome Res.* **12**: 656–664.
- KONIECZNY, A., and F. M. AUSUBEL, 1993 A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* **4**: 403–410.
- KOTIERK, M., and M. L. SMITH, 2004 PCR-based markers for genetic mapping in *Neurospora crassa*. *Fungal Genet. Newsl.* **44**: 34–36.
- LEWIS, Z. A., A. L. SHIVER, N. STIFFLER, M. R. MILLER, E. A. JOHNSON *et al.*, 2007 High-density detection of restriction-site-associated DNA markers for rapid mapping of mutated loci in *Neurospora*. *Genetics* **177**: 1163–1171.
- LOROS, J. J., and J. C. DUNLAP, 1991 *Neurospora crassa* clock-controlled genes are regulated at the level of transcription. *Mol. Cell. Biol.* **11**: 558–563.
- MATTERN, D., and S. BRODY, 1979 Circadian rhythms in *Neurospora crassa*: effects of saturated fatty acids. *J. Bacteriol.* **139**: 977–983.
- MCCLUNG, C. R., B. A. FOX and J. C. DUNLAP, 1989 The *Neurospora* clock gene *frequency* shares a sequence element with the *Drosophila* clock gene *period*. *Nature* **339**: 558–562.
- MCCCLUSKEY, K., 2003 The Fungal Genetics Stock Center: from molds to molecules. *Adv. Appl. Microbiol.* **52**: 245–262.
- METZENBERG, R. L., and J. GROBELWESCHEN, 1988 Restriction polymorphism maps of *Neurospora crassa*: updates. *Fungal Genet. Newsl.* **35**: 30–35.
- 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. USA* **88**: 9828–9832.
- 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*. *Fung. Genet. Biol.* **3**: 348–363.
- PANDEY, A., M. G. ROCA, N. D. READ and N. L. GLASS, 2004 Role of a mitogen-activated protein kinase pathway during conidial germination and hyphal fusion in *Neurospora crassa*. *Eukaryot. Cell* **3**: 348–358.
- PERKINS, D. D., and E. J. BARRY, 1976 The cytogenetics of *Neurospora*. *Adv. Genet.* **19**: 133–285.
- PERKINS, D. D., A. RADFORD, D. NEWMAYER and M. BJORKMAN, 1982 Chromosomal loci of *Neurospora crassa*. *Microbiol. Rev.* **46**: 426–570.
- PERKINS, D. D., A. RADFORD and M. SACHS, 2001 *The Neurospora Compendium: Chromosomal Loci*. Academic Press, San Diego.
- PERKINS, D. D., 2004 Wild type *Neurospora crassa* strains preferred for use as standards. *Fungal Genet. Newsl.* **51**: 7–8.
- SELITRENNIKOFF, C. P., R. E. NELSON and R. W. SIEGEL, 1974 Phase-specific genes for macroconidiation in *Neurospora crassa*. *Genetics* **78**: 679–690.
- SELKER, E. U., 1997 Epigenetic phenomena in filamentous fungi: Useful paradigms or repeat-induced confusion? *Trends Genet.* **13**: 296–301.
- SPRINGER, M. L., and C. YANOFSKY, 1989 A morphological and genetic analysis of conidiophore development in *Neurospora crassa*. *Genes Dev.* **3**: 559–571.
- TAYLOR, J., and N. J. PROVART, 2006 CapsID: a web-based tool for developing parsimonious sets of CAPS molecular markers for genotyping. *BMC Genet.* **7**: 27.
- TURIAN, G., and D. E. BIANCHI, 1971 Conidiation in *Neurospora crassa*. *Arch. Microbiol.* **77**: 262–274.



- WANG, D. G., J. B. FAN, C. J. SIAO, A. BERNO, P. YOUNG *et al.*, 1998 Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**: 1077–1082.
- WICKS, S. R., R. T. YEH, W. R. GISH, R. H. WATERSTON and R. H. PLASTERK, 2001 Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. *Nat. Genet.* **28**: 160–164.
- WINZELER, E. A., D. R. RICHARDS, A. R. CONWAY, A. L. GOLDSTEIN, S. KALMAN *et al.*, 1998 Direct allelic variation scanning of the yeast genome. *Science* **281**: 1194–1197.

Communicating editor: F. W. STAHL