Identification and Characterization of the Genes Encoding the Core Histones and Histone Variants of *Neurospora crassa*

Shan M. Hays,[1] Johanna Swanson and Eric U. Selker[2]

Department of Biology and Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403-1229

Manuscript received November 21, 2001
Accepted for publication December 26, 2001

ABSTRACT

We have identified and characterized the complete complement of genes encoding the core histones of *Neurospora crassa*. In addition to the previously identified pair of genes that encode histones H3 and H4 (*hH3* and *hH4-1*), we identified a second histone H4 gene (*hH4-2*), a divergently transcribed pair of genes that encode H2A and H2B (*hH2A* and *hH2B*), a homolog of the F/Z family of H2A variants (*hH2Az*), a homolog of the H3 variant CSE4 from *Saccharomyces cerevisiae* (*hH3v*), and a highly diverged H4 variant (*hH4v*) not described in other species. The *hH4-1* and *hH4-2* genes, which are 96% identical in their coding regions and encode identical proteins, were inactivated independently. Strains with inactivating mutations in either gene were phenotypically wild type, in terms of growth rates and fertility, but the double mutants were inviable. As expected, we were unable to isolate null alleles of *hH2A*, *hH2B*, or *hH3*. The genomic arrangement of the histone and histone variant genes was determined. *hH2Az* and the *hH3-hH4-1* gene pair are on LG IIR, with *hH2Az* centromere-proximal to *hH3-hH4-1* and *hH3* centromere-proximal to *hH4-1*. *hH3v* and *hH4-2* are on LG IIIR with *hH3v* centromere-proximal to *hH4-2*. *hH4v* is on LG IVR and the *hH2A-hH2B* pair is located immediately right of the LG VII centromere, with *hH2A* centromere-proximal to *hH2B*. Except for the centromere-distal gene in the pairs, all of the histone genes are transcribed toward the centromere. Phylogenetic analysis of the *N. crassa* histone genes places them in the Euascomycota lineage. In contrast to the general case in eukaryotes, histone genes in euascomycetes are few in number and contain introns. This may be a reflection of the evolution of the RIP (repeat-induced point mutation) and MIP (methylation induced premeiotically) processes that detect sizable duplications and silence associated genes.

EUKARYOTES employ an elaborate system to package and organize their extensive genetic material. The first order of this packaging system involves the incorporation of DNA into nucleosomes, the fundamental units of chromatin. Each nucleosome consists of 146 bp of DNA wrapped around an octameric protein complex composed of two proteins of each of the core histones H2A, H2B, H3, and H4 (McGhee and Felsenfeld 1980). Structural studies of the nucleosome have revealed that histones contain two distinct structural domains, a globular domain and N- and C-terminal "tails" (Arents *et al.* 1991). The globular domains of the core histones, which are similar to each other, allow the histones to interact with one another and with the surrounding DNA to form the nucleosome (Luger *et al.* 1997). The N- and C-terminal tails protrude from the nucleosome core. Gene expression can be regulated through post-translational modifications of the histone tails including acetylation, methylation, phosphorylation, and ADP ribosylation (Imhof and Becker 2001; Wang *et al.* 2001). The tails are known to play roles in some epigenetic processes, such as repression of the silent mating loci in yeast (Kayne *et al.* 1988). In *Neurospora crassa*, treatment with the histone deacetylase (HDAC) inhibitor, trichostatin A, results in selective loss of DNA methylation (Selker 1998). Studies in animal systems have found that methylated sequences can recruit HDACs via proteins that associate with methylated DNA (reviewed by Dobosy and Selker 2001). A recent study implicates methylation of the N-terminal tail of H3 in DNA methylation (Tamaru and Selker 2001). Finally, the suggestion has been made that histone acetylation may direct demethylation of associated DNA (Cervoni and Szyf 2001). To facilitate further studies on the role of histones in gene regulation and in epigenetic processes such as DNA methylation, we carried out a thorough characterization of the histone genes of *N. crassa*.

Available information suggests that Neurospora nucleosomes and histones are typical for eukaryotes (Noll 1976; Kornberg 1977). The full complement of histones has been purified and their amino acid profiles are similar to those of rabbit and pea histones (Goff

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY062171, AY062172, and AY062173.

[1]*Present address:* Department of Biology, California State University, Fresno, CA 93740-8034.

[2]*Corresponding author:* Institute of Molecular Biology, University of Oregon, Eugene, OR 97403-1229.
E-mail: selker@molbio.uoregon.edu

1976), consistent with the high conservation of histones generally (Thatcher and Gorovsky 1994). This conservation was confirmed in the three instances in which the sequence of a Neurospora histone was determined. A partial sequence of the H2B globular domain and C terminus showed near identity to H2B sequences from plants, animals, and fungi (Karpova *et al.* 1986). Single genes encoding H3 and H4 were cloned on a contiguous stretch of genomic DNA and sequenced (Woudt *et al.* 1983). A comparison of the conceptual translation products of these genes with yeast H3 and H4 revealed 95 and 92% identity, respectively. The two genes are transcribed divergently and separated by ~2 kbp. Southern analysis of Neurospora genomic DNA using these histone genes as probes revealed no other putative H3 or H4 genes, leading the authors to conclude that this gene pair is the only source of H3 and H4 in Neurospora. This contrasts with the situation in most eukaryotes examined, which have multiple, sometimes divergent, histone genes (Maxson *et al.* 1983; Old and Woodland 1984; Chaboute *et al.* 1993).

Whereas metazoan and plant genomes often contain tens or hundreds of genes that encode each histone, fungal genomes appear to have at most three genes for each histone (Wallis *et al.* 1980; Choe *et al.* 1982, 1985; Smith and Andresson 1983; Matsumoto and Yanagida 1985; May and Morris 1987; Ehinger *et al.* 1990). The genomic sequence of *Saccharomyces cerevisiae* (Goffeau *et al.* 1996) includes two genes for each histone, organized into two H2A-H2B and two H3-H4 gene pairs, in agreement with previous studies (Wallis *et al.* 1980; Choe *et al.* 1982; Smith and Andresson 1983). The two H3 genes and two H4 genes each encode identical proteins, but the two H2A and H2B genes each encode slightly different H2A and H2B proteins. In addition to *S. cerevisiae*, comprehensive surveys for histone genes have been carried out in two other fungi, *Schizosaccharomyces pombe* and *Emericella* (*Aspergillus*) *nidulans*. *S. pombe* has two H2A genes, each encoding slightly different proteins and a single H2B gene paired with one of the H2A genes (Choe *et al.* 1985). Three H3-H4 gene pairs have been identified in *S. pombe* (Matsumoto and Yanagida 1985). All three predicted H4 proteins are identical, but only two of the H3 genes encode identical H3 proteins; the third encodes a slightly different protein. *E. nidulans* has single genes encoding H2A, H2B, and H3 and two genes encoding slightly different H4 proteins (May and Morris 1987; Ehinger *et al.* 1990). Again, the H2A and H2B genes are physically paired, as are the H3 and H4 genes.

Given the variability in histone gene numbers in fungi, we were pleased to discover that *N. crassa* has a relatively simple set of histone genes, as described below. Molecular and genetic experiments were used to demonstrate that the genes we identified are responsible for producing all of the core histones of Neurospora.

MATERIALS AND METHODS

Informatics: Routine sequence analyses were carried out using the program MacVector (Oxford Molecular, Palo Alto, CA). Protein and nucleotide sequence database searches were performed using the BLAST program at the National Center for Biotechnology Information (Altschul *et al.* 1997). BLAST searches for histone gene ESTs were carried out with on-line databases from the *Aspergillus nidulans* and *Neurospora crassa* cDNA Sequencing Project (Roe *et al.* 2001b) and the New Mexico Neurospora Genome Project (http://www.unm.edu/~ngp/; Nelson *et al.* 1997). BLAST searches for histone genes in Neurospora used the MIPS *Neurospora crassa* database (MNCDB; Schulte *et al.* 2001) and the second assembly version of the Whitehead Institute Neurospora Sequencing Project (NSP; WICGR 2001).

Manipulation of *N. crassa*: Standard techniques for culturing Neurospora (Davis and De Serres 1970) were followed, except that a modified crossing medium (Russo *et al.* 1985) was used. Microconidiation was carried out as described by Pandit and Maheshwari (1993). Transformations for gene replacement at *his-3* were carried out by electroporation of conidia as previously described (Margolin *et al.* 1997) after linearization of pSH10, pJS94, and pJS95 with *Ase*I, pSH25 with *Nde*I, or pSH14 with *Not*I. Transformation for gene replacement at *hH4-2* was carried out similarly with pSH18 linearized with *Apa*I, selecting for inositol prototrophy. The genotypes and sources of Neurospora strains are listed in Table 1. Strains N1679 (FGSC no. 5888) and N1997 (FGSC no. 7267) were provided by the Fungal Genetics Stock Center (FGSC; University of Kansas Medical Center, Kansas City).

Southern analysis: DNA was purified from cultures grown to stationary phase in 5 ml liquid Vogel's medium with 1.5% sucrose, necessary supplements and hygromycin (200 μg/ml), as appropriate. Mycelial pads were pat dried and lyophilized and DNA was purified as described (Irelan *et al.* 1993). Restriction digests were carried out under the conditions recommended by the manufacturer (New England Biolabs, Beverly, MA). Typically, 0.6 μg of genomic DNA was digested overnight with at least 3 units of restriction endonuclease to ensure complete digestion. Digests were analyzed as previously described (Irelan and Selker 1997). Blots were exposed either to film (Kodak X-Omat Blue XB-1) or to a phosphorimager screen, which was read by a Storm 860 phosphorimager (Molecular Dynamics, Sunnyvale, CA) and visualized using ImageQuant software (version 1.11 for Macintosh; Molecular Dynamics). Complete digestion was verified by probing with DNA corresponding to a known unmethylated region of the genome (data not shown).

Restriction fragment length polymorphism mapping: Genomic DNA from either the large or the small standard set of progeny generated by Metzenberg *et al.* (1984) was digested with one of several restriction enzymes and analyzed by Southern hybridization as described above. Segregation data were compared with published data (Nelson and Perkins 2000).

Cosmid libraries: The Orbach-Sachs (O-S; Orbach 1994) and Vollmer-Yanofsky (V-Y; Vollmer and Yanofsky 1986) *N. crassa* genomic DNA cosmid libraries, provided by the FGSC, were screened and processed as previously described (Kouzminova and Selker 2001).

Plasmids: pBM60 and pBM61 have been described (Margolin *et al.* 1997). pOKE01 was a gift from J. Groteluesch en and R. Metzenberg. Plasmids were constructed by standard procedures (Sambrook *et al.* 1989). pJS94 was constructed by subcloning a 2.7-kb *Bam*HI fragment of Neurospora genomic DNA from O-S cosmid G25:G12 into pBM61 digested with *Bam*HI. The fragment includes the entire *hH2A* coding region, plus 1.4 and 0.7 kb of upstream and downstream sequences,

TABLE 1

Neurospora strains

Strain	Genotype	Source
N1089	*mata*; *trp-2*; *dim-3*; *mtr::hph*$^{+}$	Lab collection
N1674	*matA his-3*; *lys-1 am*$_{132}$ *inl*; *[(am/hph/am)*$^{ec42\ bJI2}$*]*^{RIP}	This study
N1679	*mata*; *mei-2*; *nic-3*	FGSC no. 5888
N1821	*matA his-3*$^{+}$*::hH2B*; *Bml*; *mei-2 inl inv*; *arg-10 frq::hph*$^{+}$	This study
N1822	*matA his-3*$^{+}$*::hH2A*; *Bml*; *mei-2 inl inv*; *arg-10 frq::hph*$^{+}$	This study
N1825	*matA his-3*; *Bml*; *mei-2 inl inv*; *arg-10 frq::hph*$^{+}$	This study
N1997	*matA*; *mei-2*; *am*$_{132}$ *inl inv*; *met-7*	FGSC no. 7267
N2014	*mata his-3 cyh-1*; *Bml pan-2*; *lys-1 am*$_{132}$ *inl*; *[(am/hph/am)*$^{ec42\ bJI2}$*]*^{RIP}	This study
N2015	*mata his-3 cyh-1*; *am*$_{132}$ *inl*; *[(am/hph/am)*$^{ec42\ bJI2}$*]*^{RIP}	This study
N2016	*mata his-3 cyh-1*; Δ*hH4-2::inl*$^{+}$; *Bml pan-2*; *lys-1 am*$_{132}$ *inl*; *[(am/hph/am)*$^{ec42\ bJI2}$*]*^{RIP}	This study
N2017	*mata his-3 cyh-1*; Δ*hH4-2::inl*$^{+}$; *Bml pan-2*; *lys-1 am*$_{132}$ *inl*; *[(am/hph/am)*$^{ec42\ bJI2}$*]*^{RIP}	This study
N2018	*mata his-3 cyh-1*; Δ*hH4-2::inl*$^{+}$; *am*$_{132}$ *inl*; *[(am/hph/am)*$^{ec42\ bJI2}$*]*^{RIP}	This study
N2019	*mata his-3 cyh-1*; *am*$_{132}$ *inl*; *[(am/hph/am)*$^{ec42\ bJI2}$*]*^{RIP}	This study
N2020	*mata his-3*$^{+}$*::hH4-1 cyh-1*; *am*$_{132}$ *inl*; *[(am/hph/am)*$^{ec42\ bJI2}$*]*^{RIP}	This study
N2021	*matA his-3*; *arg-12*	This study
N2022	*mata his-3*$^{+}$*::hH4-1*^{RIP} *cyh-1*; *hH4-1*^{RIP1}	This study
N2023	*matA his-3*; *hH4-1*^{RIP2}	This study
N2024	*matA his-3*; *hH4-1*^{RIP1}; *am*$_{132}$ *inl*	This study
N2025	*mata his-3 cyh-1*; *arg-12*; Δ*hH4-2::inl*$^{+}$; *am*$_{132}$ *inl*; *[(am/hph/am)*$^{ec42\ bJI2}$*]*^{RIP}	This study
N2026	*mata his-3*$^{+}$*::hH4-2*; *cyh-1*; *arg-12*; Δ*hH4-2::inl*$^{+}$; *am*$_{132}$ *inl*; *[(am/hph/am)*$^{ec42\ bJI2}$*]*^{RIP}	This study
N2027	*matA his-3*$^{+}$*::hH3*; *lys-1 am*$_{132}$ *inl*; *[(am/hph/am)*$^{ec42\ bJI2}$*]*^{RIP}	This study
N2028	(*mata*; *mei-2*; *Bml*; *nic-3*) + (*mata*; *mei-2*; *Bml*; *hH2A*^{RIP} *arg-10 frq::hph*$^{+}$)	This study
N2029	(*mata*; *mei-2*; *Bml*; *nic-3*) + (*mata*; *mei-2*; *Bml*; *hH2B*^{RIP1} *arg-10 frq::hph*$^{+}$)	This study
N2034	(*mata*; *mei-2*; *met-7*) + (*mata*; *mei-2*; *hH2A*^{RIP} *arg-10 frq::hph*$^{+}$)	This study
N2035	(*mata*; *mei-2*; *met-7*) + (*mata*; *mei-2*; *hH2A*^{RIP} *arg-10 frq::hph*$^{+}$)	This study

respectively. pJS95 was constructed by subcloning a 2-kb *Nde*I-*Pst*I fragment of Neurospora genomic DNA from O-S cosmid G25:G12 into pBM61 cut with *Sma*I. The fragment includes the entire *hH2B* coding region, plus 0.6 and 0.75 kb of upstream and downstream sequences, respectively.

pSH2 was identified by colony hybridization, using *hH4-1* as a probe, in a sublibrary of *Sau*3AI fragments cloned from O-S cosmid G11:G5 into *Bam*HI-digested pBluescript II SK(−). pSH2 contains the entire *hH4-2* coding region, plus 40 and 550 bp of upstream and downstream sequences, respectively. pSH10 was created by subcloning a 2.2-kb *Bgl*II-*Spe*I fragment of Neurospora genomic DNA from a genomic DNA clone into pBM60 digested with *Bam*HI and *Spe*I. The subcloned fragment includes the entire *hH4-1* coding region, plus 1.0 and 0.9 kb of upstream and downstream sequences, respectively. pSH14 was constructed by subcloning a 2-kb *Eae*I-*Spe*I fragment from a PCR-generated fragment of genomic DNA into pBM60 digested with *Apa*I and *Spe*I. The subcloned fragment contains a 2-kb *Bgl*II-*Hin*cII fragment of Neurospora genomic DNA that contains the entire *hH3* coding region, plus 1.2 kb and 370 bp of upstream and downstream sequences, respectively. pSH16 was constructed by subcloning a 9-kb *Nar*I-*Sac*I fragment of Neurospora genomic DNA containing *hH4-2* from a genomic clone into pBluescript II SK(−) digested with *Hin*cI and *Sac*I. pSH18 was created by replacing a 2.3-kb *Xba*I-*Bsp*106I fragment of pSH16, which contained all of the *hH4-2* coding sequence, plus 890 and 720 bp of upstream and downstream sequences, respectively, with a 3.6-kb *Xba*I-*Bst*BI fragment from pOKE01, which carries the *inl* gene. pSH18 retains 1.6 and 5.2 kb of upstream and downstream sequences, respectively, to facilitate homologous recombination at the *hH4-2* locus. pSH25 was constructed by subcloning a 3.4-kb *Eae*I-*Spe*I fragment of Neurospora DNA from pSH16 into pBM60 digested with *Not*I and *Spe*I. The subcloned fragment contains the entire *hH4-2* coding region, plus 1.3 and 1.45 kb of upstream and downstream flanking sequences, respectively.

Sequencing: The initial *hH4-2* sequence was generated by sequencing pSH2 with standard T3 and T7 primers at the Oregon State University Sequencing Center. All other sequencing was carried out at the University of Oregon Sequencing Facility using custom primers. To minimize the risk of sequencing a mutation created during polymerase chain reaction (Ausubel *et al.* 1998), products from at least three reactions were pooled for sequencing.

Phylogenetic analyses: The source and accession number of sequences obtained from GenBank is available on request. The following expressed sequence tagged (EST) sequences were obtained from the *Cryptococcus neoformans* cDNA Sequencing Project (Roe *et al.* 2001a) and conceptually translated to obtain their putative histone products: H2B from a5f05cn; H3 from b4d04cn.r1; H4.1 from b1f06cn, b1f10j2, and a7c02j2; and H4.2 from b5g08cn. The following EST sequences were obtained from the *Fusarium sporotrichioides* cDNA Sequencing Project (Roe *et al.* 2001c) and conceptually translated to obtain their putative histone products: H2A from b2d06fs, j4d05fs.f1, and d3h05fs.f1; H2B from l1g06fs and l3f03fs; and H4.1 from o1c08fs.r1 and f1f06fs.r1. The following is a partial list of EST sequences for the core histone genes of *N. crassa* from the *Aspergillus nidulans* and *Neurospora crassa* cDNA Sequencing Project (Roe *et al.* 2001b). *hH2A*: d5c07ne, e9d02ne, and g8g08nm.r1; *hH2B*: a9c10ne, a6c09ne.f1, and arf06nm.r1; *hH3*: b7c11ne, f7b11nm, and a4a05np; *hH4-1*: b2g12ne, b7b12ne, and a4b01ne.f1; *hH4-2*: d8e05ne, h4b06nm, and a1b03ne.

Protein and nucleic acid sequences were imported into Biology Workbench (http://workbench.sdsc.edu/), where they were aligned with CLUSTALW (Felsenstein 1989; Thompson *et*

al. 1994) and manually edited. Aligned sequences were used to determine the identity of other fungal histones relative to Neurospora using MVIEW Multiple Alignment Display (Brown *et al.* 1998). Phylogenetic trees were constructed from the aligned sequences with CLUSTALX and drawn by Neighbor-Joining Plot (Felsenstein 1989).

RESULTS

Identification and sequence analysis of genes encoding the core histones and histone variants of *N. crassa*: To identify all of the *hH3* and *hH4* homologs in the *N. crassa* genome, we screened two genomic cosmid libraries with the previously identified *hH3-hH4* gene pair (Woudt *et al.* 1983). Of the cosmids identified, only SV31:G9 from the V-Y cosmid library (Vollmer and Yanofsky 1986) was found to carry the *hH3-hH4* gene pair. The other cosmids, G11:G5, G22:D6, and X18:H1 from the O-S cosmid library (Orbach 1994), were found to carry a second gene encoding H4, which we named *hH4-2*. The previously identified *hH4* gene in the *hH3-hH4* gene pair has been renamed *hH4-1* (Perkins *et al.* 2001).

hH4-1 and *hH4-2* encode identical proteins and are 96% identical (300/312) at the nucleotide level in their coding regions. Both genes contain two introns, at precisely conserved locations, but no similarity was found in the introns or in the 5′ or 3′ untranslated regions (UTRs). The intron lengths are also different. Introns 1 and 2 of *hH4-1* are 69 and 68 bp, whereas the introns of *hH4-2* are 316 and 65 bp, respectively (Figure 1).

To identify the genes encoding histones H2A and H2B, we first searched the two publicly available Neurospora EST databases (Nelson *et al.* 1997; Roe *et al.* 2001b) by tBLASTn for corresponding cDNAs. We identified one set of ESTs for each and designated their corresponding genes *hH2A* and *hH2B*. Genomic copies of the genes were identified in O-S cosmid G25:G10, subcloned, and sequenced. We found that *hH2A* and *hH2B* lie 2.6 kb apart, are transcribed divergently, and are composed of three and four exons, respectively (Figure 1).

We failed to detect additional *hH2A*, *hH2B*, *hH3*, or *hH4* genes in the *N. crassa* genome. Southern analysis of genomic DNA digested with a variety of restriction enzymes using *hH2A*, *hH2B*, *hH3*, *hH4-1*, and *hH4-2* as probes detected only the bands expected for the known genes, as shown for *hH2A* (Figure 2). tBLASTn searches for genes encoding H2A, H2B, H3, and H4 in the EST databases (Nelson *et al.* 1997; Roe *et al.* 2001b) detected numerous ESTs for *hH2A*, *hH2B*, *hH3*, *hH4-1*, and *hH4-2*, indicating that each of these genes is expressed (Figure 1). No other histone ESTs were found.

tBLASTn searches were also carried out using the databases of the two *N. crassa* genome sequencing projects, MNCDB (Schulte *et al.* 2001) and NSP (WICGR 2001). No new genes encoding core histones were found, but three histone variants were discovered. One was an H2A variant of the H2A F/Z family (Carr *et al.* 1994), which we named *hH2Az* (Figure 3A). In the absence of ESTs for *hH2Az*, we used the high conservation within the *hH2A* F/Z family of variants to determine that the *hH2Az* coding region consists of three exons (Figure 1). We also found an H3 variant, which we named *hH3v*. It showed the greatest similarity to SpCENP-A (Figure 3B), a *S. pombe* homolog of CSE4, which is an essential gene of *S. cerevisiae* (Stoler *et al.* 1995). On the basis of similarity to other H3 variants, we suggest that the *hH3v* open reading frame is composed of two exons (Figure 1). The intron location in *hH3v* appears identical to the intron location in *hH3*. Without a cDNA sequence, we cannot be certain that *hH3v* does not have a second intron in the region of the expected N terminus, where there is low conservation among H3 variants (Figure 3B). We also identified a fragmented *hH4* homolog, *hH4v*, which appears composed of six exons and five introns at positions different from those in *hH4-1* and *hH4-2* (Figure 1). The canonical methionine codon (ATG) of the potential *hH4v* translation product was

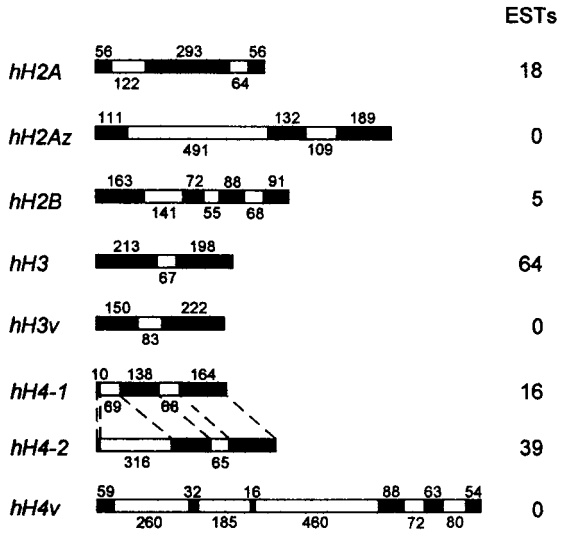

Figure 1.—Structure of histone genes and number of associated ESTs identified in publicly available databases (Nelson *et al.* 1997; Roe *et al.* 2001b). Solid boxes represent exons and open boxes represent introns. The gene structures were based on the similarity of the conceptually translated products with closely related proteins or were determined by comparing ESTs to the genomic sequences. Precise splice sites were assigned only if they matched the *N. crassa* splice site acceptor and donor consensus sequences (Radford and Parish 1997). Except for *hH4v*, numbering begins with the A in the initiating methionine codon and ends at the termination codon. For *hH4v*, numbering begins at a leucine codon that corresponds to the initiating methionine codon in the *hH4* genes and ends at a glutamic acid codon that similarly corresponds to the *hH4* termination codon (see Figure 3C). Due to the low similarity among H3 variants, the N terminus of *hH3v* is uncertain. Sizes of exons and introns (in base pairs) are shown above and below the diagrams, respectively. The exons are the same size in *hH4-1* and *hH4-2* because of conservation of intron positions, as indicated by the dashed lines.

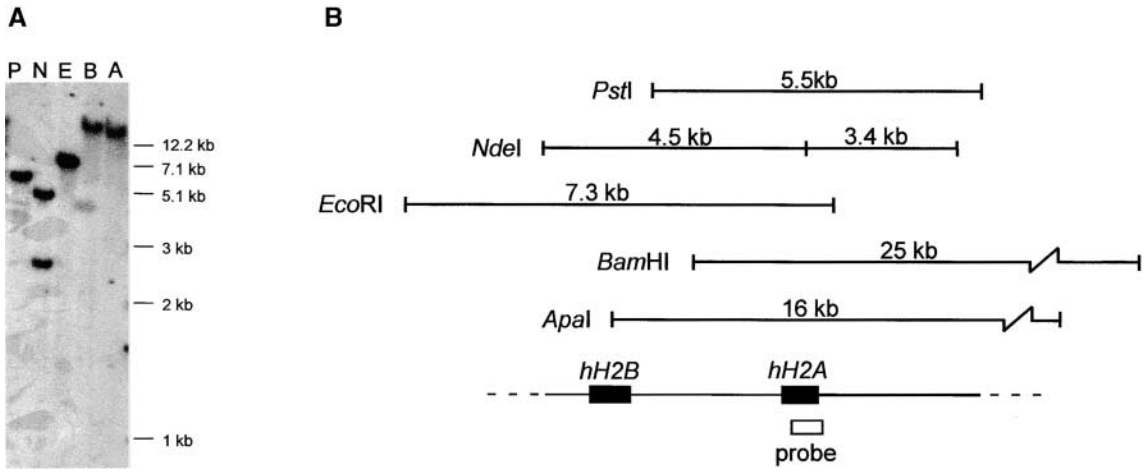

FIGURE 2.—Southern analysis for *hH2A* homologs. (A) Genomic DNA from strain N1089 was digested with the indicated restriction enzymes (P, *Pst*I; N, *Nde*I; E, *Eco*RI; B, *Bam*HI; A, *Apa*I) and probed for *hH2A*. The positions of size markers are shown on the right. No fragments <1 kb in size were detected. (B) Partial restriction map for the *hH2A-hH2B* genomic region. Fragment sizes resulting from digestion with each restriction enzyme are indicated. The probe used in A is represented by the open box below the map.

replaced by a leucine codon (TTG) and its predicted sequence is only 62% identical to that of *N. crassa* H4 (Figure 3C). This is significantly lower than the >91% homology among H4 proteins across the fungal kingdom (Table 3), indicating that *hH4v* either produces a highly divergent H4 or is a pseudogene.

Genomic organization of the genes encoding the core histones and histone variants: The genomic location of the *hH3-hH4-1* gene pair on the right arm of linkage group (LG) II was previously determined by restriction fragment length polymorphism (RFLP) mapping (METZENBERG and GROTELUESCHEN 1987). The presence of this gene pair in the V-Y cosmid SV31:G9 (VOLLMER and YANOFSKY 1986), which carries the LG IIR gene *aro-1* (CATCHESIDE *et al.* 1985; PERKINS *et al.* 2001), corroborates this assignment (Figure 4). The genomic locations of *hH4-2* and the *hH2A-hH2B* gene pair were also determined by RFLP mapping using the standard sets of RFLP progeny (METZENBERG *et al.* 1984). *hH4-2* cosegregated with *con-7* and *trp-1* on LG IIIR and *hH2A-hH2B* cosegregated with *ars-1* and the centromere of LG VII (*Cen-VII*; data not shown).

To map the histone and histone variant genes more precisely, we determined their location in the sequenced regions available from the NSP (Table 2) and MNCDB, searched for nearby genetic loci by BLASTn and BLASTx queries at NCBI and correlated matches with the Neurospora genetic map. The precise locations of all the histone genes were determined (Figure 4). The *hH3-hH4-1* pair lies 15 kb centromere-proximal to *aro-1* with *hH4-1* lying centromere-distal to *hH3*. *hH2Az* also lies on LG II, 158 kb centromere-proximal to *arg-5*. Since no centromere-related sequences (CENTOLA and CARBON 1994) were found between *arg-5* and *hH2Az*, we conclude that it lies on the right arm of LG II and is transcribed toward the centromere. *hH4-2* and *hH3v* were both found to lie on the right arm of LG III and are both transcribed toward the centromere. *hH4-2* lies 38 kb centromere-distal to *trp-1,* whereas *hH3v* lies ≥121 kb centromere-proximal to *ad-2*. *hH4v* is located on the right arm of LG IV ≥416 kb centromere-distal to *tol* and is transcribed toward the centromere.

The *hH2A-hH2B* gene pair lies on supercontig 31 from the NSP (WICGR 2001), which also contains *Cen-VII*, in agreement with our RFLP mapping data. When we compared conceptual restriction digests of the contigs from supercontig 31 to published restriction maps of *Cen-VII* (CENTOLA and CARBON 1994), we found that *hH2A-hH2B* lies to the right of the identified centromeric DNA with *hH2A* lying centromere-proximal to *hH2B*. CENTOLA and CARBON (1994) originally determined the extent of the A + T-rich *Cen-VII* by constructing a restriction map of the centromere and flanking regions with the restriction enzyme *Pac*I, which cleaves at an 8-bp recognition site composed solely of A:T pairs. The centromere was found to have a large concentration of *Pac*I sites, while the relatively A + T-poor flanking regions were found to be devoid of *Pac*I sites. To determine the distance between *hH2A-hH2B* and the centromere, we determined the concentration of *Pac*I sites around *hH2A-hH2B*. We found that *Pac*I sites were numerous both centromere-proximal and centromere-distal to it, suggesting that this gene pair may lie in centromeric heterochromatin (Figure 4).

Genetic identification of the histone gene complement: All of the core histones are essential in *S. cerevisiae* (RYKOWSKI *et al.* 1981; SCHUSTER *et al.* 1986; KAYNE *et al.* 1988; MANN and GRUNSTEIN 1992) and are presumably essential in all eukaryotes. Thus, as another test of whether we had identified the full complement of genes encod-

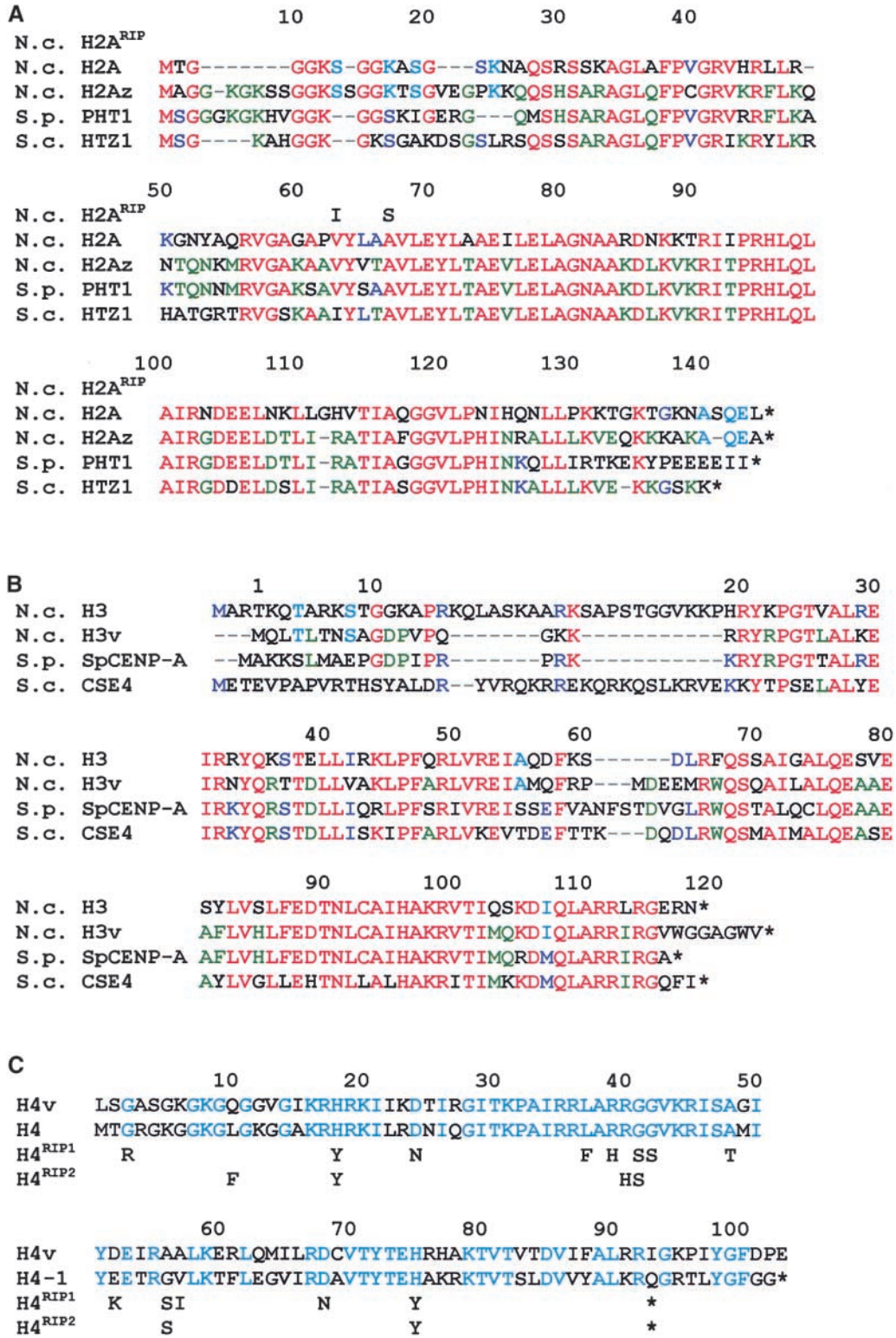

FIGURE 3.—Alignments of the conceptually translated products of *N. crassa* histone variant genes with corresponding *N. crassa* histones, other fungal histone variants, and predicted mutant proteins. (A) Alignment of H2A, H2A^{RIP}, and H2Az from *N. crassa* with PHT1 from *S. pombe* and HTZ1 from *S. cerevisiae*. Numbering is relative to the H2Az sequence. (B) Alignment of H3v and H3 from *N. crassa* with fungal H3 variants SpCENP-A from *S. pombe* and CSE4 from *S. cerevisiae*. Numbering is relative to the H3v sequence. Due to the lack of homology in the extreme N terminus of H3v and the lack of ESTs, the N terminus shown is tentative. (C) Alignment of H4 with the conceptually translated products of *hH4v* (H4v), *hH4-1^{RIP1}* (H4-1^{RIP1}), and *hH4-1^{RIP2}* (H4-1^{RIP2}). Numbering is relative to H4. Turquoise residues are those that are identical between the core histone and the variant in *N. crassa*. Green residues are those that are identical between the *N. crassa* variant and at least one of the other fungal histone variants. Red residues are those that are identical between the core histone, the *N. crassa* histone variant, and at least one of the other fungal histone variants. Dark blue residues are those that are identical between any of the sequences, not including the *N. crassa* histone variant. Only the missense and nonsense mutations are shown in the mutant sequences. Asterisks indicate the position of stop codons.

ing each of the core histones of Neurospora, we attempted to inactivate each of the identified histone genes. Viability of strains bearing mutations in all known genes encoding a given histone would suggest that we had missed a functional histone gene, whereas lethality would suggest that we had not. Alternatively, lethality could result if redundant genes exist, but the expression level of each lies below a threshold for viability.

In the case of H4, the test required us to inactivate both *hH4-1* and *hH4-2*. For *hH4-2*, this was accomplished by replacing the *hH4-2* gene with a functional *inl* gene (Δ*hH4-2*::*inl*$^+$), which encodes an enzyme in the inositol biosynthesis pathway (PERKINS *et al.* 2001). Two strains (N2014 and N2015) were transformed with pSH18, a plasmid carrying Δ*hH4-2*::*inl*$^+$, and Inl$^+$ strains were selected. Southern analysis of the transformants indicated

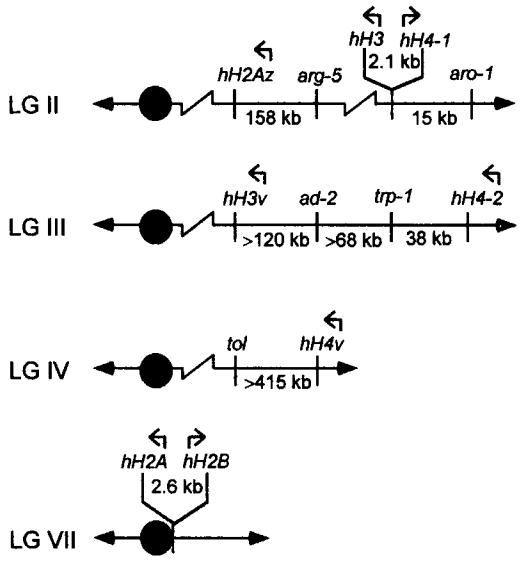

FIGURE 4.—Genomic locations of genes encoding core histones and histone variants. Solid circles represent centromeres. Distances between the genes in the two gene pairs were determined by sequencing the intervening regions. The distances to the other genetic loci were determined using available genome sequence data. Some distances are uncertain, due to gaps in the intervening sequenced regions. The *hH2A-hH2B* gene pair resides at the right edge of the A + T-rich DNA of the LG VII centromere.

that homologous replacement events occurred at a frequency higher than that normally reported for Neurospora (FINCHAM 1989; ARONSON *et al.* 1994): 24% (10/42) for N2014 and 14% (6/43) for N2015 (Figure 5). Two Δ*hH4-2*::*inl*$^+$ transformants of N2014 (N2016 and N2017) and one of N2015 (N2018) were rendered homokaryotic by microconidiation (PANDIT and MAHESHWARI 1993) and used in further tests. We compared the growth rates of N2018 and its host N2015 in race tubes and found no difference (data not shown). We also found that the transformants were fully fertile either as males or as females (data not shown). Apparently, either *hH4-1* fully complements the loss of *hH4-2* or *hH4-2* is not a major contributor of H4.

To determine the importance of *hH4-1* in the production of H4, we inactivated *hH4-1* with RIP, a process in which duplicated sequences are peppered with G:C to A:T transition mutations during the premeiotic stage of crosses (SELKER 1990). Many sequences altered by RIP are found methylated in vegetative cells (SELKER *et al.* 1993; SINGER *et al.* 1995). We targeted *hH4-1* to the *his-3* locus, crossed the resulting *hH4-1* duplication strains and analyzed the progeny for mutations and/or methylation at *hH4-1*. The nonduplication parent carried a nonfunctional *arg-12* allele, which lies <1 map unit from *aro-1* (PERKINS *et al.* 2001). Because the *hH3-hH4-1* gene pair lies between *arg-12* and *aro-1*, crossovers between *arg-12* and *hH4-1* should be rare. Therefore, Arg$^+$ progeny were selected, enriching for progeny whose native *hH4-1* allele came from the duplication parent and therefore had the opportunity to undergo RIP. Two of 12 progeny screened for RFLPs and methylation, N2022 and N2023, exhibited methylation at *hH4-1*. N2023 also exhibited an altered digestion pattern with *Dpn*II (data not shown). The *hH4-1* gene at the native locus of each of the strains was sequenced and both alleles showed extensive mutations by RIP (Figure 3C). The *hH4-1*^{RIP1} allele from N2022 exhibited 32 mutations in the 450 bp from the start codon through the stop codon, and allele *hH4-1*^{RIP2} from N2023 showed 22 mutations in this span. A conceptual translation of *hH4-1*^{RIP1} revealed 7 silent mutations and 13 missense mutations, whereas a conceptual translation of *hH4-1*^{RIP2} revealed 6 silent mutations and 6 missense mutations. Both alleles include a nonsense mutation at Q94, which should cause a deletion of 10 amino acids from the C terminus. In *S. cerevisiae*, deletion of the 4 C-terminal amino acids of H4 is not lethal, whereas deletion of 6 C-terminal amino acids is (KAYNE *et al.* 1988). Therefore, even disregard-

TABLE 2

Core histone and histone variant gene locations in the publicly available databases of the Neurospora sequencing projects and their corresponding linkage groups (LG)

Gene	MNCDB BAC[a]	Coordinates[b]	NSP supercontig:contig[c]	Coordinates[b]	LG
hH2Az	B15B24	42,207-41,176	21:2.36	52,515-53,546	II
hH3	B14H13	49,363-49,840	3:2.64	23,393-23,870	II
	B7K22	14,758-15,235			
hH4-1	B14H13	47,225-46,777	3:2.64	21,255-20,807	II
	B7K22	12,620-12,172			
hH3v	None		2:2.35	961-507	III
hH4-2	None		2:2.41	50,896-50,204	III
hH4v	None		16:2.301	59,659-58,291	IV
hH2A	None		31:2.465	10,885-11,475	VII
hH2B	None		31:2.465	8,277-7,600	VII

[a] MIPS *Neurospora crassa* database (SCHULTE *et al.* 2001).
[b] Reverse coordinates indicate that the gene is on the complementary strand.
[c] Neurospora Sequencing Project (WICGR 2001).

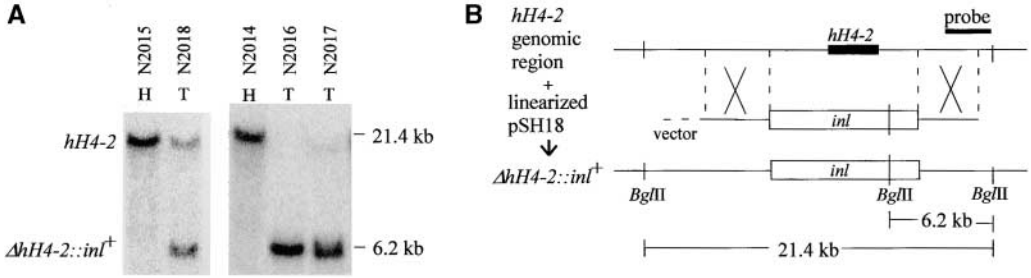

FIGURE 5.—Homologous replacement of *hH4-2* with *inl*. (A) Southern analysis of host (H) N2015 and its transformant (T) N2018 and of host N2014 and its transformants, N2017 and N2016. Genomic DNA was digested with *Bgl*II and probed with sequences flanking the replaced region. The native *hH4-2* region yields a 21.4-kb *Bgl*II fragment, while a proper replacement event yields a 6.2-kb fragment. (B) Partial restriction map of the *hH4-2* genomic region, the linearized pSH18 plasmid used in the transformations and the *hH4-2* genomic region after a clean replacement event. In the genomic sequence the numbering is relative to the A in the initiation codon of *hH4-2*. In pSH18, solid lines represent sequences in common with the *hH4-2* region, dashed lines represent vector DNA, and the boxed sequence represents the *inl*-bearing fragment from pOKE01. The precise location of *inl* in the fragment is unknown. Figure is not drawn to scale.

ing the missense mutations, it seems unlikely that our H4 mutant alleles are functional. The growth rates of the *hH4-1*^{RIP1} strains (N2022, N2023) and their parents were found to be equivalent, suggesting that *hH4-2* supplied adequate H4. N2022 and N2023 were crossed as a male and female and with each other and in all cases were found to be fertile. Apparently, either *hH4-1* or *hH4-2* is sufficient for normal growth.

To confirm this conclusion and to verify that histone H4 is essential in Neurospora, we crossed strains carrying each *hH4* mutant allele to determine whether or not *hH4-1*^{RIP1};Δ*hH4-2::inl*$^{+}$ progeny could be generated. Two crosses were carried out. In the first, N2024 × N2025, both parents carried defective *inl* alleles, so that the *inl*$^{+}$ gene that replaced *hH4-2* was the only functional *inl* gene in the cross. Furthermore, the Δ*hH4-2::inl*$^{+}$ parent (N2025) carried *arg-12*, which is tightly linked to *hH4-1*. Ascospores from the cross were activated and plated directly on medium lacking inositol and arginine, thus selecting for Δ*hH4-2::inl*$^{+}$ and the *arg-12*$^{+}$ allele near *hH4-1*^{RIP1}. If strains lacking functional *hH4-1* and *hH4-2* genes could be viable, for example, due to a hypothetical third *hH4* gene in the genome, the vast majority of progeny (>99%) should have been *hH4-1*^{RIP1};Δ*hH4-2::inl*$^{+}$. Instead, analysis of 10 progeny by Southern hybridization revealed that all carried the wild-type *hH4-1* allele, suggesting that the *hH4-1*^{RIP1};Δ*hH4-2::inl*$^{+}$ double mutant is not viable.

The second cross was equivalent to the first, except that a functional copy of *hH4-2* was inserted at the *his-3* locus of N2025 by transformation with pSH25, creating N2026. This artificially provided a third *hH4* gene as a control. Progeny were selected as before, except that the functional *hH4-2* at *his-3* was selected by requiring histidine prototrophy. The majority of progeny (7/9) were *hH4-1*^{RIP1};Δ*hH4-2::inl*$^{+}$. Therefore, *hH4-1*^{RIP1} and Δ*hH4-2::inl*$^{+}$ are null alleles and are synthetically lethal. On the basis of these results and our inability to detect other genes encoding H4, we conclude that *hH4-1* and *hH4-2* are the only functional *hH4* genes in *N. crassa*.

We also attempted to recover an inactive *hH3* allele by RIP. To provide a second copy of *hH3* to activate RIP, we targeted *hH3* to *his-3* by transforming N1674 with pSH14, creating N2027, which was subsequently crossed with N2025. Progeny were selected using the same selection regime used in the *hH4-1* RIP cross, described above, and 16 progeny were analyzed for hallmarks of RIP at *hH3*. No methylation or altered digestion patterns were observed (data not shown). Compared to the success rate of recovering null *hH4-1*^{RIP} alleles (2/12), the inability to recover *hH3*^{RIP} alleles (0/16) suggests that the loss of *hH3* is lethal. On the basis of these results and our inability to identify other genes encoding H3, we conclude that *hH3* is the only source of H3 in *N. crassa*.

To confirm genetically that *hH2A* and *hH2B* are unique in the genome, we employed the sheltered RIP strategy of METZENBERG and GROTELUESCHEN (1992). This strategy relies on the inclusion of a *mei-2* mutation in each parent to cause chromosome nondisjunction and thus production of aneuploid progeny. Disomic progeny break down rapidly to give heterokaryons with euploid nuclei that should be genetically identical, except for the chromosome that was subject to nondisjunction. One nuclear type should inherit this chromosome from one parent, while the other nuclear type should inherit this chromosome from the other parent. In this way a strain carrying an essential gene that has been disrupted by RIP can survive due to complementation by an extra wild-type allele.

In the sheltered RIP cross that we carried out, the two parents carried different mutations on the chromosome carrying *hH2A-hH2B* (LG VII). *hH2A* and *hH2B* were targeted to the *his-3* locus of N1825 by transformation with pJS94 and pJS95 to create duplication strains N1822 and N1821, respectively. These transformants were crossed with N1679 and the resulting ascospores were germinated on medium lacking nicotinamide and arginine. This selected for the *nic-3*$^{+}$ allele from N1821 and N1822 and the *arg-10*$^{+}$ allele from N1679, resulting solely in the recovery of progeny disomic for LG VII. Eight progeny from the N1822 (*hH2A*-duplication) cross and 13 progeny from the N1821 (*hH2B*-duplication) cross were an-

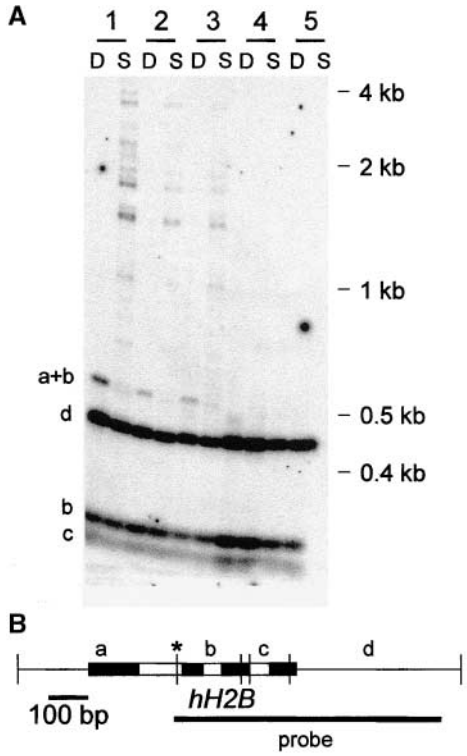

FIGURE 6.—Generation and propagation of *hH2B*^{RIP1} allele. (A) Southern analysis of genomic DNA of N2035 (1), N2034 (2), N2029 (3), N1821 (4), and N1679 (5) digested with *Sau*3AI (S) or *Dpn*II (D) and probed with *hH2B*. The higher molecular weight bands in the *Sau*3AI lanes are indicative of methylation. N2029 is the dikaryotic progeny of N1821 × N1679 and N2034 and N2035 are the dikaryotic progeny of N2029 × N1997. The letters to the left of the autoradiogram indicate the identity of the fragments as indicated in B. Positions of size markers are indicated on the right. No fragments >4 kb in size were detected. (B) *Dpn*II/*Sau*3AI restriction map of *hH2B*. Vertical lines represent restriction sites; solid and open boxes represent the exons and introns of *hH2B*, respectively. The asterisk refers to the *Sau*3AI/*Dpn*II site lost in *hH2B*^{RIP1}, resulting in the a+b fragment evident in A. The probe used in Southern blot in A is shown below the restriction map.

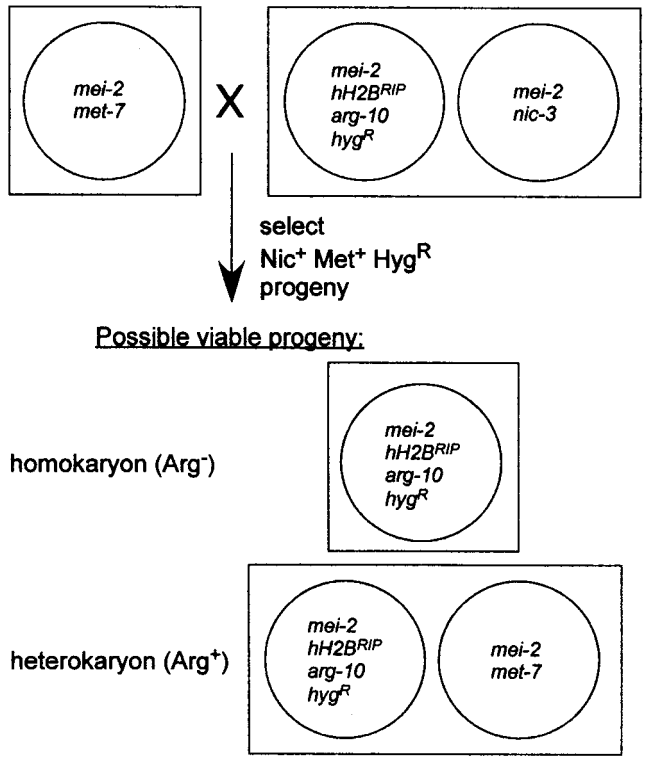

FIGURE 7.—Cross to determine whether the *hH2A*^{RIP} and *hH2B*^{RIP} (represented here only by *hH2B*^{RIP}) alleles are lethal. The markers shown are all carried on LG VII and the cross is homokaryotic for *mei-2*. The selection strategy allows only the growth of progeny carrying the LG VII with the mutant allele. Recovery of homokaryotic progeny indicates the mutant allele is not lethal.

alyzed by Southern hybridization for hallmarks of RIP at *hH2A* and *hH2B*. One strain from the *hH2A* RIP cross, N2028, exhibited methylation at *hH2A*. Two of the progeny from the *hH2B* RIP cross exhibited methylation at *hH2B* and one of these (N2029) also exhibited an altered *Dpn*II digestion pattern and was selected for further analysis (Figure 6).

To test whether the *hH2A*^{RIP} and *hH2B*^{RIP} alleles were null and essential, we crossed N2028 and N2029 with N1997, a *mei-2* strain with a genetically distinct LG VII (Figure 7). Ascospores from this second cross were germinated on medium containing hygromycin and arginine, which allows only two classes of progeny to survive: homokaryotic euploid progeny carrying the mutant histone allele and progeny disomic for LG VII, arising from a cross between N1997 and the nucleus bearing the mutant histone allele. Since the homokaryotic progeny would be Arg^{-} and the heterokaryotic progeny would be Arg^{+}, we distinguished between these two expected classes by spot-testing on plates lacking arginine. In the cross involving the *hH2B*^{RIP1} allele, all eight progeny were Arg^{+}. The apparent requirement for a wild-type allele of *hH2B* in any strain carrying *hH2B*^{RIP1} strongly argues that *hH2B*^{RIP1} is null and that *hH2B* is essential, indicating that it is the only source of H2B in Neurospora.

In the case of the cross involving the *hH2A*^{RIP} allele, three of the seven progeny tested were Arg^{-}, indicating either that the *hH2A*^{RIP} allele is not fully defective or that it is a null and there is a second *hH2A* gene in the genome that makes up for this deficiency. To distinguish between these two possibilities, we amplified the *hH2A*^{RIP} allele from the Arg^{+} progeny by PCR and sequenced it. Mutations from RIP were found, but it was not obvious that they would result in a null allele. In the 594-bp coding region and introns, only five mutations were found, three of which are silent mutations (Figure 3A). Since we were unable to generate a definite null allele of *hH2A*, we could not determine genetically the number of genes encoding H2A in the *N. crassa*. Nevertheless, since only *hH2A* was detected when genomic DNA blots or genomic cosmid libraries were screened with *hH2A* and since all of the H2A ESTs and genomic se-

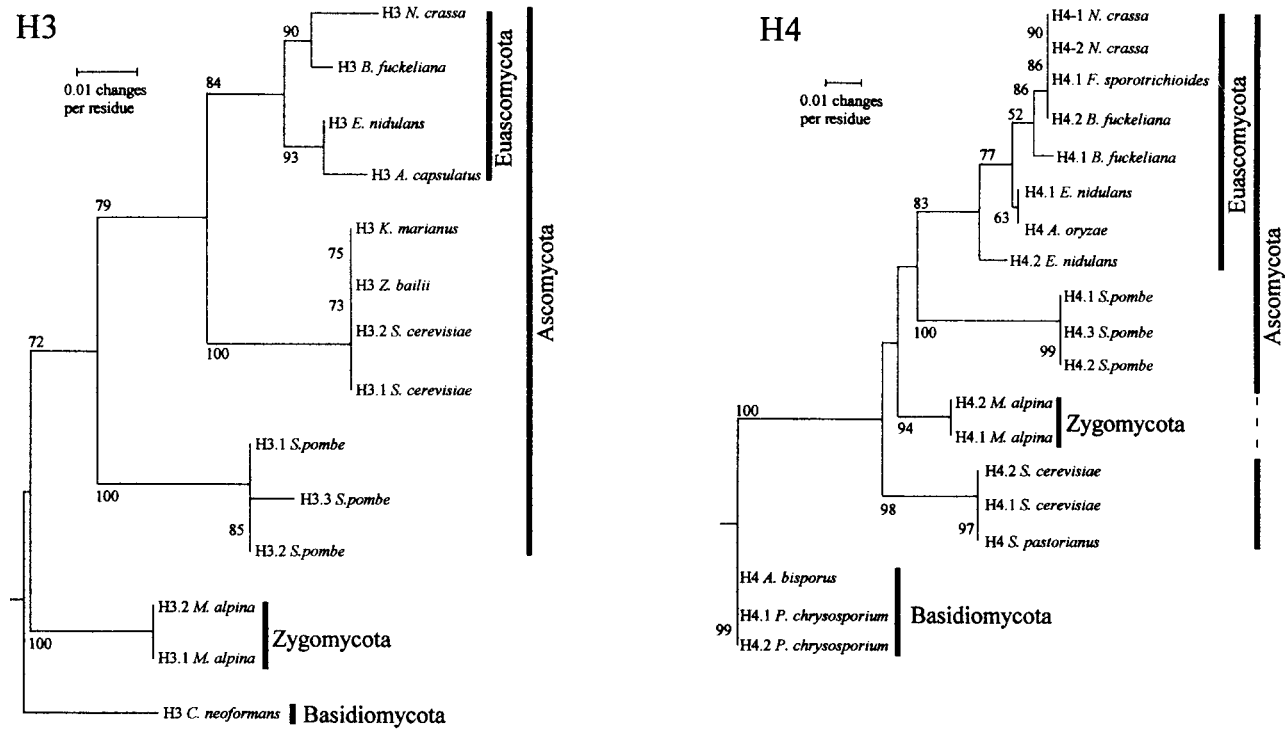

FIGURE 8.—Most parsimonious tree of fungal histone H3 and H4 proteins. Branch lengths are proportional to amino acid substitutions. Percentages shown are from 1000 bootstrap replicates. Percentages <50 are not shown. Ascomycota, Zygomycota, and Basidomycota are sister phyla, whereas Euascomyctes are an order of Ascomycota. Basidiomycetes were used as the outgroup. The dotted portion of the Ascomycota line indicates that Zygomycota are not part of the Ascomycota.

quences that we found belong to *hH2A*, we conclude that *hH2A* is unique in the genome.

Evolutionary relationship of the core histones of *N. crassa* to those of other fungi: Previous phylogenetic analyses of histones included few fungal representatives (THATCHER and GOROVSKY 1994). To place the Neurospora histone genes in a fungal phylogeny, we searched the nonredundant databases at NCBI and several publicly available fungal EST databases to identify as many fungal core histone genes as possible. The amino acid sequences of the identified histone genes were then aligned and phylogenetic trees were constructed. All of the histones of Neurospora were found to group consistently with those of other Euascomycetes, as expected from previous phylogenetic studies based on morphological and molecular data (SPATAFORA 1995; LIU *et al.* 1999; BERBEE *et al.* 2000; Figure 8). As known for eukaryotes as a whole (THATCHER and GOROVSKY 1994), H2A and H2B are the least highly conserved histones in fungi (Table 3) and most of their divergence is found in the N- and C-terminal tails. In contrast, H3 and H4 are more highly conserved (Table 3) and most of their divergence occurred in the structured globular domains.

Histone H4 of *N. crassa* is identical to one of the two H4 proteins of *Botrytis fuckeliana* and to the one H4 protein from *F. sporotrichioides* for which the entire amino acid sequence is available (Table 3). To estimate the time of divergence between *hH4-1* and *hH4-2* relative to their divergence from these counterparts, we constructed a phylogenetic tree of the coding sequences of the H4-coding genes from the euascomycetes. Although most of the tree is at low resolution, *hH4-1* and *hH4-2* cluster together with high confidence (data not shown). Thus, it appears that the two *hH4* genes of Neurospora arose from a relatively recent gene duplication event. This is supported by an examination of intron locations between the two *hH4* genes of Neurospora and the two H4-coding genes from *E. nidulans*, the only other euascomycete for which the genomic sequence of the H4 genes are known (EHINGER *et al.* 1990). Although the location of the second intron in *hH4-1* and *hH4-2* is not shared by the two H4 genes from *E. nidulans*, the first intron is located at the same position in all four genes.

DISCUSSION

We set out to identify all of the genes encoding the core histones in *N. crassa*. Previously, a single gene pair (*hH3* and *hH4-1*), coding for H3 and H4, had been identified (WOUDT *et al.* 1983). We identified another gene encoding H4 (*hH4-2*) and single closely linked genes encoding H2A and H2B (*hH2A* and *hH2B*). To be certain that we had identified the entire complement of histone genes, we assessed the dispensability of each histone gene genetically. In the cases of *hH2B* and *hH3*, we found that we could not generate viable strains car-

TABLE 3

Percentage identity of fungal histones to *N. crassa* histones

	H2A[a] (%)	H2B[a] (%)	H3[a] (%)	H4[a] (%)
Euascomycota				
A. capsulatus	NA	NA	97.1	NA
A. niger	95.6	NA	NA	NA
A. oryzae	NA	NA	NA	99.0
B. fuckeliana	90.0	86.5	98.5	99.0 (1) 100 (2)
E. nidulans	96.3	87.4	97.8	99.0 (1) 97.1 (2)
F. sporotrichioides	97.8	96.4	NA	100[b]
P. anserina	96.3	97.8	NA	NA
Ascomycota				
K. marzianus	NA	NA	94.9	NA
S. pastorianus	NA	NA	NA	92.2
S. cerevisiae	83.9 (1) 84.7 (2)	77.9 (1) 76.4 (2)	94.9 (1,2)	92.2 (1,2)
S. pombe	86.8 (1,2)	74.8	92.6 (1,2) 91.9 (3)	92.2 (1,2, 3)
Z. bailii	NA	NA	94.9	NA
Fungi				
A. bisporus	74.3	69.5	NA	91.3
C. neoformans	NA	71.4	89.9	NA
M. alpina	NA	NA	91.9 (1,2)	94.2 (1,2)
P. chrysosporium	NA	NA	NA	91.3 (1,2)

NA, not available.

[a] In those organisms in which it is known that a given histone is encoded by multiple genes, the particular gene product being compared is indicated by its number designation in parentheses.

[b] A second gene encoding H4 is omitted, because it has been only partially sequenced (Roe *et al.* 2001c).

rying null alleles at these loci. Strains carrying null alleles of *hH4-1* or *hH4-2* were viable and phenotypically wild type, however, suggesting that neither of these genes is solely responsible for producing all the histone H4 in Neurospora. We were unable to generate a double mutant without providing an extra functional *hH4* gene. We were unable to generate a null allele of *hH2A*, but our inability to detect homologs, as discussed above, leads us to conclude that *hH2A* is unique in the genome. We identified variant forms of histones H2A (*hH2Az*), H3 (*hH3v*), and H4 (*hH4v*), but do not know whether they are involved in the same processes as their homologs in other species.

RFLP mapping and database mining were used to map *hH2Az* and the *hH3-hH4-1* gene pair to LG IIR, *hH3v* and *hH4-2* to LG IIIR, hH4v to LG IVR, and the *hH2A-hH2B* gene pair to LG VII near the right edge of the centromere. The number and arrangement of the core histone genes in *N. crassa* is identical to that in the closest relative for which a comprehensive search of histone genes has been undertaken, *E. nidulans* (May and Morris 1987; Ehinger *et al.* 1990). Furthermore, an analysis of numerous cDNA sequences available for the euascomycetes *F. sporotrichioides* (Roe *et al.* 2001c) and *B. fuckeliana* suggests that they also have single genes for H2A, H2B, and H3 and two genes coding for H4. Why do euascomycetes appear to maintain two H4 genes, when single genes suffice for the other core histones? We found that under laboratory conditions either gene is sufficient for wild-type viability, fertility, and rate of growth. Interestingly, in *E. nidulans* the abundance of *hhfB* mRNA, which codes for H4.2, is regulated differentially from that of *hhfA*, whose product is 98% identical to H4.2, suggesting that these H4 genes play different roles (Ehinger *et al.* 1990). Perhaps different euascomycetes maintain two H4 genes for different reasons.

Although the single-copy histone genes from each euascomycete are presumably orthologous, the *hH4* genes appear exceptional. For instance, the H4 phylogenetic trees show that H4-1 and H4-2 group together, while H4.1 of *E. nidulans* forms a close sister group and H4.2 of *E. nidulans* forms the outgroup (Figure 8B). This indicates that *hH4-1* and *hH4-2* are not orthologs of *hhfA* (H4.1) and *hhfB* (H4.2), respectively, but instead are paralogs of *hhfA*, having arisen from a duplication of an ancestral ortholog of *hhfA*.

The progenitor of Neurospora that duplicated the ancestral *hH4* gene may have been competent at RIP or an equivalent duplication-sensing mechanism, since the fungus *Ascobolus immersus*, which is more distantly related to Neurospora than *E. nidulans*, has an apparently related mechanism, MIP (Goyon and Faugeron 1989). The duplication of an *hH4* gene may have escaped detection due to its small size. The *hH4* coding

region requires only 312 bp, below the apparent minimum (~400 bp) for a duplication to be subject to RIP or MIP (Goyon *et al.* 1996; Watters *et al.* 1999). This suggests that RIP/MIP may not be an impediment to the evolutionary development of new genes through gene duplication, so long as the genes are small.

RIP/MIP may be a factor in the unique genomic organization of euascomycete histone genes, including the minimal histone gene set and the existence of introns in the histone genes. Although introns are extremely uncommon in histone genes, generally (Wallis *et al.* 1980; Choe *et al.* 1982, 1985; Maxson *et al.* 1983; Smith and Andresson 1983; Old and Woodland 1984; Matsumoto and Yanagida 1985; Horowitz *et al.* 1987; Stark and Milner 1989; Wefes and Lipps 1990; Chaboute *et al.* 1993; Puerta *et al.* 1994; Mackenzie *et al.* 2000; Mollapour and Piper 2001), each report of a histone gene from a euascomyte has noted the existence of introns (Woudt *et al.* 1983; May and Morris 1987; Ehinger *et al.* 1990). In fact, introns are found in all 14 euascomycete histone genes, for which the genomic sequence is known, which includes all core histone genes from *N. crassa*, all core histone genes from *E. nidulans*, H2A from *B. fuckeliana*, H2A and H2B from *Podospora anserina*, and H3 from *Ajellomycetes capsulatus*. Thus introns in histone genes are the rule, not the exception in euascomycetes, setting them apart from yeasts (Wallis *et al.* 1980; Choe *et al.* 1982, 1985; Smith and Andresson 1983; Matsumoto and Yanagida 1985; May and Morris 1987; Ehinger *et al.* 1990) and most other eukaryotes (Maxson *et al.* 1983; Old and Woodland 1984; Chaboute *et al.* 1993). The common ancestor of euascomycetes likely carried multiple histone genes. If so, the evolution of RIP/MIP would have made this situation untenable, since the coding regions for histones H2A, H2B, and H3 are >400 bp in length, making uninterrupted duplicate copies of these genes substrates for these genome defense systems. Consequently, only histone genes that were altered, by the insertion of nonhomologous introns to break up the contiguous stretches of homology and/or by a reduction in the number of histone gene sets to one, would be able to survive repeated sexual crosses intact.

We thank Michael Freitag and Lenna Kuzminova for materials used in some experiments, Robert Metzenberg and Greg Kothe for discussions, and Michael Freitag for comments on the manuscript. S.M.H. was supported in part by U.S. Public Health Services training grant GM-07759. This work was supported by U.S. Public Health Services grant GM-35690 to E.U.S.

LITERATURE CITED

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Arents, G., R. W. Burlingame, B. C. Wang, W. E. Love and E. N. Moudrianakis, 1991 The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. Proc. Natl. Acad. Sci. USA **88:** 10148–10152.

Aronson, B. D., K. M. Lindgren, J. C. Dunlap and J. J. Loros, 1994 An efficient method for gene disruption in *Neurospora crassa*. Mol. Gen. Genet. **242:** 490–494.

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman *et al.* (Editors), 1998 *Current Protocols in Molecular Biology*. John Wiley & Sons, New York.

Berbee, M. L., D. A. Carmean and K. Winka, 2000 Ribosomal DNA and resolution of branching order among the ascomycota: how many nucleotides are enough? Mol. Phylogenet. Evol. **17:** 337–344.

Brown, N. P., C. Leroy and C. Sander, 1998 MView: a web-compatible database search or multiple alignment viewer. Bioinformatics **14:** 380–381.

Carr, A. M., S. M. Dorrington, J. Hindley, G. A. Phear, S. J. Aves *et al.*, 1994 Analysis of a histone H2A variant from fission yeast: evidence for a role in chromosome stability. Mol. Gen. Genet. **245:** 628–635.

Catcheside, D. E. A., P. J. Storer and B. Klein, 1985 Cloning of the ARO cluster gene of *Neurospora crassa* and its expression in *Escherichia coli*. Mol. Gen. Genet. **199:** 446–451.

Centola, M., and J. Carbon, 1994 Cloning and characterization of centromeric DNA from *Neurospora crassa*. Mol. Cell. Biol. **14:** 1510–1519.

Cervoni, N., and M. Szyf, 2001 Demethylase activity is directed by histone acetylation. J. Biol. Chem. **276:** 40778–40787.

Chaboute, M. E., N. Chaubet, C. Gigot and G. Philipps, 1993 Histones and histone genes in higher plants: structure and genomic organization. Biochimie **75:** 523–531.

Choe, J., D. Kolodrubetz and M. Grunstein, 1982 The two yeast histone H2A genes encode similar protein subtypes. Proc. Natl. Acad. Sci. USA **79:** 1484–1487.

Choe, J., T. Schuster and M. Grunstein, 1985 Organization, primary structure, and evolution of histone H2A and H2B genes of the fission yeast Schizosaccharomyces pombe. Mol. Cell. Biol. **5:** 3261–3269.

Davis, R. H., and F. J. De Serres, 1970 Genetic and microbiological research techniques for *Neurospora crassa*. Methods Enzymol. **17A:** 47–143.

Dobosy, J. R., and E. U. Selker, 2001 Emerging connections between DNA methylation and histone acetylation. Cell. Mol. Life Sci. **58:** 721–727.

Ehinger, A., S. H. Denison and G. S. May, 1990 Sequence, organization and expression of the core histone genes of *Aspergillus nidulans*. Mol. Gen. Genet. **222:** 416–424.

Felsenstein, J., 1989 PHYLIP (Phylogeny Inference Package), version 3.2. Cladistics **5:** 164–166.

Fincham, J. R. S., 1989 Transformation in fungi. Microbiol. Rev. **53:** 148–170.

Goff, C. G., 1976 Histones of *Neurospora crassa*. J. Biol. Chem. **251:** 4131–4138.

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

Goyon, C., and G. Faugeron, 1989 Targeted transformation of *Ascobolus immersus* and *de novo* methylation of the resulting duplicated DNA sequences. Mol. Cell. Biol. **9:** 2818–2827.

Goyon, C., C. Barry, A. Grégoire, G. Faugeron and J. L. Rossignol, 1996 Methylation of DNA repeats of decreasing sizes in Ascobolus immersus. Mol. Cell. Biol. **16:** 3054–3065.

Horowitz, S., J. K. Bowen, G. A. Bannon and M. A. Gorovsky, 1987 Unusual features of transcribed and translated regions of the histone H4 gene family of Tetrahymena thermophila. Nucleic Acids Res. **15:** 141–160.

Imhof, A., and P. B. Becker, 2001 Modifications of the histone N-terminal domains. Evidence for an "epigenetic code?" Mol. Biotechnol. **17:** 1–13.

Irelan, J., V. Miao and E. U. Selker, 1993 Small scale DNA preps for *Neurospora crassa*. Fungal Genet. Newsl. **40:** 24.

Irelan, J. T., and E. U. Selker, 1997 Cytosine methylation associated with repeat-induced point mutation causes epigenetic gene silencing in *Neurospora crassa*. Genetics **146:** 509–523.

Karpova, O. I., N. M. Ananyeva, T. M. Ermokhina and I. A. Krasheninnikov, 1986 Some peculiarities of primary structure of histone H2b of the mould fungus Neurospora crassa. Biokhimiia **51:** 788–800.

Kayne, P. S., U. J. Kim, M. Han, J. R. Mullen, F. Yoshizaki *et al.*, 1988 Extremely conserved histone H4 N terminus is dispensable

for growth but essential for repressing the silent mating loci in yeast. Cell **55:** 27–39.

Kornberg, R. D., 1977 Structure of chromatin. Annu. Rev. Biochem. **46:** 931–954.

Kouzminova, E. A., and E. U. Selker, 2001 *Dim-2* encodes a DNA-methyltransferase responsible for all known cytosine methylation in Neurospora. EMBO J. **20:** 4309–4323.

Liu, Y. J., S. Whelen and B. D. Hall, 1999 Phylogenetic relationships among ascomycetes: evidence from an RNA polymerse II subunit. Mol. Biol. Evol. **16:** 1799–1808.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond, 1997 Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature **389:** 251–260.

Mackenzie, D. A., P. Wongwathanarat, A. T. Carter and D. B. Archer, 2000 Isolation and use of a homologous histone H4 promoter and a ribosomal DNA region in a transformation vector for the oil-producing fungus Mortierella alpina. Appl. Environ. Microbiol. **66:** 4655–4661.

Mann, R. K., and M. Grunstein, 1992 Histone H3 N-terminal mutations allow hyperactivation of the yeast *GAL1* gene *in vivo*. EMBO J. **11:** 3297–3306.

Margolin, B. S., M. Freitag and E. U. Selker, 1997 Improved plasmids for gene targeting at the *his-3* locus of *Neurospora crassa* by electroporation. Fungal Genet. Newsl. **44:** 34–36.

Matsumoto, S., and M. Yanagida, 1985 Histone gene organization of fission yeast: a common upstream sequence. EMBO J. **4:** 3531–3538.

Maxson, R., R. Cohn, L. Kedes and T. Mohun, 1983 Expression and organization of histone genes. Annu. Rev. Genet. **17:** 239–277.

May, G. S., and N. R. Morris, 1987 The unique histone H2A gene of Aspergillus nidulans contains three introns. Gene **58:** 59–66.

McGhee, J. D., and G. Felsenfeld, 1980 Nucleosome structure. Annu. Rev. Biochem. **49:** 1115–1156.

Metzenberg, R. L., and J. Grotelueschen, 1987 A restriction polymorphism map of *Neurospora crassa:* more data. Fungal Genet. Newsl. **34:** 39–44.

Metzenberg, R. L., and J. S. Grotelueschen, 1992 Disruption of essential genes in Neurospora by RIP. Fungal Genet. Newsl. **39:** 37.

Metzenberg, R. L., J. N. Stevens, E. U. Selker and E. Morzycka-Wroblewska, 1984 A method for finding the genetic map position of cloned DNA fragments. Neurospora Newsl. **31:** 35–39.

Mollapour, M., and P. Piper, 2001 Targeted gene deletion in Zygosaccharomyces bailii. Yeast **18:** 173–186.

Nelson, M., and D. D. Perkins, 2000 Restriction polymorphism maps of Neurospora crassa: 2000 update. Fungal Genet. Newsl. **47:** 25–39.

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

Noll, M., 1976 Differences and similarities in chromatin structure of *Neurospora crassa* and higher eucaryotes. Cell **8:** 349–355.

Old, R. W., and H. R. Woodland, 1984 Histone genes: not so simple after all. Cell **38:** 624–626.

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

Pandit, A., and R. Maheshwari, 1993 A simple method of obtaining pure microconidia in *Neurospora crassa*. Fungal Genet. Newsl. **40:** 64–65.

Perkins, D. D., A. Radford and M. S. Sachs, 2001 *The Neurospora Compendium; Chromosomal Loci*. Academic Press, San Diego.

Puerta, C., J. Martin, C. Alonso and M. C. Lopez, 1994 Isolation and characterization of the gene encoding histone H2A from Trypanosoma cruzi. Mol. Biochem. Parasitol. **64:** 1–10.

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

Roe, B. A., D. Kupfer, J. Lewis, S. Yu, K. Buchanan *et al.*, 2001a The *Cryptococcus neoformans* cDNA Sequencing Project (available from http://www.genome.ou.edu/cneo.html).

Roe, B. A., D. Kupfer, H. Zhu, J. Gray, S. Clifton *et al.*, 2001b The *Aspergillus nidulans* and the *Neurospora crassa* cDNA Sequencing Project (available from http://www.genome.ou.edu/fungal.html).

Roe, B. A., Q. Ren, D. Kupfer, H. Lai, M. Beremand *et al.*, 2001c The *Fusarium sporotrichioides* cDNA Sequencing Project (available from http://www.genome.ou.edu/fsporo.html).

Russo, V. E. A., T. Sommer and J. A. A. Chambers, 1985 A modified Vogel's medium for crossings, mating-type tests, and the isolation of female-sterile mutants of *Neurospora crassa*. Neurospora Newsl. **32:** 10–11.

Rykowski, M. C., J. W. Wallis, J. Choe and M. Grunstein, 1981 Histone H2B subtypes are dispensable during the yeast cell cycle. Cell **25:** 477–487.

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

Schulte, R., V. Aign, J. Hoheisel, P. Brandt, B. Fartmann *et al.*, 2001 German Neurospora genome project (www.mips.biochem.mpg.de/proj/neurospora).

Schuster, T., M. Han and M. Grunstein, 1986 Yeast histone H2A and H2B amino termini have interchangeable functions. Cell **45:** 445–451.

Selker, E. U., 1990 Premeiotic instability of repeated sequences in *Neurospora crassa*. Annu. Rev. Genet. **24:** 579–613.

Selker, E. U., 1998 Trichostatin A causes selective loss of DNA methylation in Neurospora. Proc. Natl. Acad. Sci. USA **95:** 9430–9435.

Selker, E. U., D. Y. Fritz and M. J. Singer, 1993 Dense non-symmetrical DNA methylation resulting from repeat-induced point mutation (RIP) in Neurospora. Science **262:** 1724–1728.

Singer, M. J., B. A. Marcotte and E. U. Selker, 1995 DNA methylation associated with repeat-induced point mutation in *Neurospora crassa*. Mol. Cell. Biol. **15:** 5586–5597.

Smith, M. M., and O. S. Andresson, 1983 DNA sequences of yeast H3 and H4 histone genes from two non-allelic gene sets encode identical H3 and H4 proteins. J. Mol. Biol. **169:** 663–690.

Spatafora, J. W., 1995 Ascomal evolution of filamentous ascomycetes: evidence from molecular data. Can. J. Bot. **73**(Suppl. 1): S811–S815.

Stark, M. J., and J. S. Milner, 1989 Cloning and analysis of the Kluyveromyces lactis TRP1 gene: a chromosomal locus flanked by genes encoding inorganic pyrophosphatase and histone H3. Yeast **5:** 35–50.

Stoler, S., K. C. Keith, K. E. Curnick and M. Fitzgerald-Hayes, 1995 A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. Genes Dev. **9:** 573–586.

Tamaru, H., and E. U. Selker, 2001 A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. Nature **414:** 277–283.

Thatcher, T. H., and M. A. Gorovsky, 1994 Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. Nucleic Acids Res. **22:** 174–179.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Vollmer, S. J., and C. Yanofsky, 1986 Efficient cloning of genes of *Neurospora crassa*. Proc. Natl. Acad. Sci. USA **83:** 4869–4873.

Wallis, J. W., L. Hereford and M. Grunstein, 1980 Histone H2B genes of yeast encode two different proteins. Cell **22:** 799–805.

Wang, H., Z. Q. Huang, L. Xia, Q. Feng, H. Erdjument-Bromage *et al.*, 2001 Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. Science **293:** 853–857.

Watters, M. K., T. A. Randall, B. S. Margolin, E. U. Selker and D. R. Stadler, 1999 Action of repeat-induced point mutation on both strands of a duplex and on tandem duplications of various sizes in Neurospora. Genetics **153:** 705–714.

Wefes, I., and H. J. Lipps, 1990 The two macronuclear histone H4 genes of the hypotrichous ciliate Stylonychia lemnae. DNA Seq. **1:** 25–32.

WICGR, 2001 Neurospora Sequencing Project, assembly version 2. Whitehead Institute/MIT Center for Genome Research (available from www-genomewimitedu).

Woudt, L. P., A. Pastink, A. E. Kempers-Veenstra, A. E. M. Jansen, W. H. Mager *et al.*, 1983 The genes coding for histone H3 and H4 in *Neurospora crassa* are unique and contain intervening sequences. Nucleic Acids Res. **11:** 5347–5361.

Communicating editor: R. H. Davis