Aflatoxin biosynthesis gene clusters and flanking regions

K.C. Ehrlich¹, J. Yu¹ and P.J. Cotty¹,²
¹USDA, ARS, Southern Regional Research Center, New Orleans, LA, USA, and ²Division of Plant Pathology and Microbiology, Department of Plant Sciences, University of Arizona, Tucson, AZ, USA
2004/1370: received 29 November 2004, revised 1 March 2005 and accepted 2 March 2005

ABSTRACT


Methods and Results: Sequencing was on fosmid clones selected by homology to Aspergillus parasiticus sequence. Alignments revealed that gene order is conserved among AF gene clusters of Aspergillus nomius, A. parasiticus, two sclerotial morphotypes of Aspergillus flavus, and an unnamed Aspergillus sp. Phylogenetic relationships were established using the maximum likelihood method implemented in PAUP. Based on the Eurotiomycete/Sordariomycete divergence time, the A. flavus-type cluster has been maintained for at least 25 million years. Such conservation of the genes and gene order reflects strong selective constraints on rearrangement. Phylogenetic comparison of individual genes in the cluster indicated that ver-1, which has homology to a melanin biosynthesis gene, experienced selective forces distinct from the other pathway genes. Sequences upstream of the polyketide synthase-encoding gene vary among the species, but a four-gene sugar utilization cluster at the distal end is conserved, indicating a functional relationship between the two adjacent clusters.

Conclusions: The high conservation of cluster components needed for AF production suggests there is an adaptive value for AFs in character-shaping niches important to those taxa.

Significance and Impact of the Study: This is the first comparison of the complete nucleotide sequences of gene clusters harbouring the AF biosynthesis genes of the main AF-producing species. Such a comparison will aid in understanding how AF biosynthesis is regulated in experimental and natural environments.

Keywords: aflatoxin, Aspergillus flavus, Aspergillus nomius, biosynthesis, gene cluster, phylogenetics, transcription.

INTRODUCTION

Aspergillus spp. have been frequently used for food fermentations and as a source of industrially important enzymes (Saxena et al. 2001). They are found in soils in temperate regions worldwide (Cotty et al. 1994). Some species are a threat to commercially important commodities because of their ability to produce toxic and carcinogenic aflatoxins (AF) (Cotty et al. 1994). In Aspergillus parasiticus, enzymes and the regulatory proteins involved in AF biosynthesis are encoded by more than 25 contiguous genes in a 70-kb cluster (Yu et al. 2004a). Aspergillus nidulans which produces sterigmatocystin (ST), an intermediate in AF formation, and species related to A. nidulans, have a similar biosynthetic cluster containing homologues of many of these genes (Brown et al. 1996). Some species more closely related to A. nidulans than Aspergillus flavus also can produce AF (Klich et al. 2003).

Both the A. flavus and A. nidulans clusters contain a gene that encodes a positive-acting master transcriptional regulator, AflR (Payne et al. 1993; Chang et al. 1995). Co-regulation by AflR of many of the genes in these clusters has been demonstrated for A. nidulans (Brown et al. 1996;...
Fernandes et al. (1998) and A. parasiticus (Ehrlich et al. 1999; Cary et al. 2000). Other factors also affect AF biosynthesis including nitrogen source, pH and soil humidity (Moreno Romo et al. 1986; Cotty 1988; Guo et al. 1996; Keller et al. 1997; Ehrlich and Cotty 2002). Recently, we provided evidence that different species of AF-producing Aspergilli have differential regulation of toxin production (Ehrlich et al. 2003).

Aflatoxin does not appear to be essential to the growth and life of the fungus, and much speculation about its role has been published (Bennett and Christensen 1983; Ciegler 1983; Lilleyhoj 1991; Jarvis and Miller 1996; Demain and A. 2000). Some proposals for the function of AFs are: they provide a way to remove excess carbon when fungi grow on carbon-rich sources (Bu’lock 1965), they act as chemical signals between species (Lillehoj 1991), they are involved in processes of fungal development (Cotty 1988; Beppu 1992; Trail et al. 1995; Kale et al. 1996), they may be protective against soil microbial or insect competitors (Matsumura and Knight 1967; Reiss 1975; Wright et al. 1982; Jarvis et al. 1984; Kurtzman et al. 1987; Llewellyn et al. 1988; Drummond and Pinnock 1990; Dowd 1992). Because AFs are not particularly phytotoxic (McLean et al. 1995; Hasan 2001), they are not believed to be plant virulence factors and atoxigenic isolates are equally able to invade susceptible crop species (Cotty 1989). It has been proposed that AF production may be a vestigial trait that has survived because clustering provides a mechanism for efficient horizontal gene transfer (Geiser et al. 1998; Walton 2000).

Here we show that the sequences of the AF cluster from three other closely related AF-producing species, A. flavus, Aspergillus nomius, and an unnamed taxon from West Africa, have the same gene order as A. parasiticus. These species show differences in the degree of conservation of genes proximal to the ends of the cluster. These findings provide further insight into the evolution of AF biosynthesis in Aspergillus spp.

MATERIALS AND METHODS

Aspergillus strains

Fungal isolates include A. flavus AF13, a strain that produces sclerotia with average size >300 mm (strain L, ATCC 96044), A. flavus AF36, an L-strain isolate that produces no AF (ATCC 96045) and A. flavus AF70, a strain that produces sclerotia with average size <300 mm (strain S, ATCC MYA384) (Ehrlich et al. 2003), A. nomius NRRL13137 (ATCC15546) (Kurtzman et al. 1987), an unnamed taxon isolate from Benin, West Africa, BN008R that produces both B and G AFs and S-type sclerotia (ATCC MYA-379) (Cotty and Cardwell 1999), and A. parasiticus SRRC 143 (NRRL2999 and ATCC 56775).

Sequencing

Clones from fosmid libraries (Agencourt Bioscience Corporation, Beverly, MA, USA) were selected by hybridization to PCR amplicons from the ends and middle of the A. parasiticus AF biosynthesis pathway gene cluster (GenBank no. AY371490). Sequence analyses were carried out on subclones prepared by shearing fosmid DNA in an HPLC apparatus, separating 3000–3500-bp fragments on a 1% agarose gel, and ligating to BstXI-cut pGTC plasmid (Agencourt). DNA from subclones was sequenced (MegaBase 1000 Sequencing Apparatus; Amersham, Piscataway, NJ, USA) using dye-terminator chemistry (Applied Biosystems, Foster City, CA, USA). Sequencing was performed with at least sixfold redundancy to obtain a probable sequencing error rate <1 bp per 1000 bp.

Sequence alignments

Sequence alignments were made with both Vector NTI Suite v.6 software (Informax, Bethesda, MD, USA) and DNAMAN version 5 (Lyndon Corporation, Vandreuil, QC, Canada). BLASTP and TBLASTN searches were carried out against the Genbank (http://www.ncbi.nlm.nih.gov/BLAST/) and A. nidulans genomic DNA (WICGR, http://www.broad.mit.edu/annotation/fungi/Aspergillus/) databases. Searches were also against an A. flavus EST Database constructed by the Institute for Genome Research using RNA from A. flavus NRRL3357 grown on a variety of AF-inducing and AF-inhibiting media (Yu, J. personal communication). Homology E-values are based on BLAST searches and are only reported when coverage was >80%. Homology of sequences are based upon the full alignment method in DNAMAN (gap open penalty = 10, gap extension penalty = 1, DNA transition weight = 0.5, protein weight = GONNET). The GenBank accession numbers for the new AF cluster sequences are: A. flavus (L-type) AF13, AY510451; A. flavus AF36, AY510455; unnamed taxon BN008R, AY510452; A. flavus (S-type) AF70, AY510453; A. nomius NRRL13137, AY510454.

Phylogenetic analyses

Phylogenetic analyses used PAUP 4b010 (Sinauer Associates, Sunderland, MA, USA). Concatenated intron sequences (2695 total characters of which 159 were parsimony informative), intergenic region sequences (14 066 characters of which 909 were parsimony informative) and coding sequences (47 474 characters of which 1476 were parsimony informative) were analysed separately by heuristic search in PAUP. For partition homogeneity tests, partitions either consisted of each AF cluster intergenic region, the coding sequences of individual genes or their encoded protein
sequences (14,556 total characters of which 322 were parsimony informative). The number of nonsynonymous nucleotide substitutions for each gene was estimated from the total number of variable characters obtained by comparison of the three *A. flavus* cluster sequences (AF70, AF36 and AF13) to the *A. nomius* sequence (NRRL13137) minus the number of synonymous changes. The number of synonymous changes in gene coding sequence was obtained by subtracting the number of variable characters in intron and protein sequences from the total number of variable characters. Divergence times were estimated from phylogenetic trees generated by heuristic search of rDNA internal transcribed spacer sequence (ITS) using the maximum likelihood option in PAUP (Kasuga et al. 2002). Tree branch lengths were assumed to be proportional to time of divergence. The divergence of Sordariomycetes (represented by *Neurospora crassa*) from Eurotiomycetes (*Aspergillus* spp.) is assumed to be >310 and <670 Ma based on the known fossil record (Berbee and Taylor 2001; Heckman et al. 2001).

**RESULTS**

**Organization of gene clusters from different AF-producing species**

The gene order for the AF clusters from isolates of *A. flavus*, the unnamed taxon from West Africa (BN008R), and *A. nomius* is the same as that for the previously characterized cluster from *A. parasiticus* (Fig. 1a and Table 1) (Yu et al. 2004a). Clusters varied in length from 66-1 kb for the small sclerotial (S) morphotype and 66-5 kb for the large sclerotial (L) morphotype of *A. flavus* to 68-4 kb for *A. nomius*. The shorter length of the *A. flavus* AF clusters is mainly because of deletion of portions of the coding regions of *norB* and the adjacent gene *cypA*, and the entire *norB1/cypA* intergenic region in the *A. flavus* clusters (Ehrlich et al. 2004). For all of the AF clusters, the gene, *norB*, which is predicted to encode a dehydrogenase (Yu et al. 2004a), is at the proximal end of the cluster and a gene predicted to encode a protein with unknown function, *hypA* (*aflY*), is at the distal end. The genes in the newly characterized AF clusters in *A. flavus*, *A. nomius*, and the unnamed taxon from West Africa are predicted to encode the same proteins as those encoded by genes in the *A. parasiticus* cluster (Table 1) (Yu et al. 2004a). Twenty of these genes are homologous to genes in the *A. nidulans* ST cluster (Fig. 1b). However, gene order in the ST cluster differs markedly from that of the AF cluster (Brown et al. 1996). Both gene clusters encode proteins that carry out the same biosynthetic steps, except that in *A. nidulans* the final metabolite is ST rather than AF (Fig. 1c). Two of the other AF cluster genes (*ordA* and *aflT*) have possible homologues in the *A. nidulans* genome outside of the ST cluster. The genes with highest nucleotide identities in the *A. nidulans* database are listed in Table 1. Some of the open reading frames (ORFs) identified in the *A. nidulans* cluster (e.g. *stcC* and *stcD*) were not found in the AF clusters. In general, homologous genes in the AF and ST clusters have similar lengths, but in a few cases, differ in number of introns (Yu et al. 2004a).

Two closely related putative genes were discovered, which had not been previously described in the *A. parasiticus* cluster. One (*hypB1*) is in the *pksA/nor-l* intergenic region and the other (*hypB2*) is in the intergenic region of *avfA* and *verB* (Fig. 1a and Table 1). A BLAST search of GenBank revealed no recognizable functional domains in the predicted protein products of these genes. The gene, *hypB1*, is predicted to encode a 139 aa protein and *hypB2*, a 163 aa

**Fig. 1** Schematic representation of the (a) AF and (b) ST gene clusters and (c) the AF biosynthetic pathway showing the known stable intermediates. The bar at the top shows the length in kilob. The direction of transcription is indicated by arrows. In (b), the genes shown in brackets are the AF cluster homologues. NOR, norsolorinic acid; AVN, averantin; HAVN, hydroxyaverantin; OAVN, o xoaverantin; AVF, averufin; HVN, hydroxyversicolorone; VHA, versicolorone hemiacetal acetate; VERB, versicolorin B; VERA, versicolorin A; ST, sterigmatocystin; AF, aflatoxin.
The predicted proteins HypB1 and HypB2 have 42% aa identity to one another (85% identity of a 20 aa region near the C-terminus). A search of GenBank with either protein gave the predicted protein encoded by the A. nidulans ST cluster gene, stcM, as the closest match (Table 1). Amino acid identity of A. parasiticus HypB1 and HypB2 to StcM was 43 and 46% respectively. A BLASTN search of the A. flavus EST database revealed that both hypB1 and hypB2 are transcribed. The highest scoring EST sequence for hypB1 had an E-value = 2e-64 (100% aa identity) and for hypB2 an E-value = 4e-70 (90% aa identity).

The promoter regions of aflR, azfA, aflT and A. nomius aflR lack the binding site, 5'-TCGN3CG3'- (Ehrlich et al. 1999), for the pathway-specific transcriptional regulator AflR. All other genes in both clusters have one or two predicted AflR-binding sites. Only one AflR site is located within the promoter regions of StcM. All other genes in both clusters have one or two predicted AflR-binding sites. Only one AflR site is located within the promoter regions of StcM.
transcribed genes, *norB/cypA*. In the shared promoter region of *hexA* and *hexB*, two overlapping AflR-binding sites are located in the intergenic region, 500 bp from the *hexA* translational start and 210 bp from the *hexB* translational start. In the ST cluster the promoter regions of the genes, *stsF, stsG, A. nidulans aflR* and *aflF, stsP, stsA, stsB* and *stsO* lack an AflR-binding site. In most cases, the locations and the sequences of the AflR sites in the intergenic regions of the ST cluster genes were not identical to the AflR sites in their AF cluster homologues (results not shown).

**Comparison of gene and protein homologues in ST and AF clusters**

The two sclerotial morphotypes of *A. flavus* (AF13 and AF70) share an overall DNA homology of >99% in the AF gene cluster. Homology of these *A. flavus* clusters to *A. parasiticus* is 96%, to BN008R, 93%, and to *A. nomius*, 91%. Coding regions have 4–10% higher sequence identity than intergenic regions (results not shown). By comparison of variable sites in the alignment of *A. nomius* and *A. flavus* protein and coding region sequence we found that the ratio of nonsynonymous substitutions to synonymous substitutions varied for different cluster genes (Table 1), from a low of 0.05 for *ver-1* to a high of 1.04 for *hypA*. The ratio was >0.5 for 10 genes and <0.3 for nine genes. Differences were also observed in the percent identity of genes in the AF cluster to their homologues in the ST cluster. Four AF cluster genes (*omtB, vbs, verB* and *ver-1*) have ST homologues with nucleotide identities ≥70% while five genes (*aflR, aflF, hexA, hexB* and *estA*) have ST homologues with nucleotide identities ≤55%. For the genes *aflT, omtA* and *ordA* the sequences with highest identity in the *A. nidulans* database were genes outside of the ST cluster locus. For *cypA*, partially deleted from the cluster of *A. flavus*, but present in *A. parasiticus*, the most similar *A. nidulans* gene was also outside the ST cluster. Except for the *ordA* gene, the closest matching genes outside the ST cluster have <40% nucleotide identities. The proportion of nonsynonymous changes between *A. flavus* and *A. nomius* increases as percent identity of the *A. flavus* genes to the corresponding *A. nidulans* genes decreases (Fig. 2).

**Genes upstream and downstream from the AF or ST clusters**

Sequence analysis for three of the clusters (*A. nomius* NRRL13137, *A. flavus* AF13 and unnamed taxon isolate BN008R) extended >3 kb from the translational stop codon of *norB* at the proximal end of the cluster. Previously, no genes were identified within 2 kb of the *norB* stop codon in *A. parasiticus* (Yu et al. 2004a). A survey of potential upstream ORFs in the three sequences revealed the presence of four putative genes in BN008R (12.3 kb), two genes in the *A. flavus* L-morphotype AF13 (10.6 kb) and one gene in *A. nomius* NRRL13137 (8 kb) upstream region (Fig. 3). The promoter regions for these putative genes do not contain AflR-binding sites. A gene predicted to encode a protein with high sequence identity to a xylanase occurs within 2 kb of the end of the AF cluster of *A. nomius* NRRL13137, but is not found in the other upstream regions examined. In both BN008R and AF13, a gene encoding a possible major facilitator superfamily (MFS) efflux transporter is found within 5 kb of the cluster and a putative dehydrogenase occurs >8-6 kb upstream of the cluster in these species. Between these putative genes, in BN008R, but not in AF13, is a gene predicted to encode a P450 monoxygenase. This predicted protein has closest sequence similarity (37%) to trichodiene C-15 hydroxylase, encoded by a gene in the *Fusarium* trichothecene biosynthetic gene cluster.

Like the *A. parasiticus* cluster (Yu et al. 2000), AF pathway clusters from the three other studied *Aspergillus* spp. contain a four gene sugar utilization cluster distal to *hypA*. The intergenic region between *hypA* and the first gene in this cluster, *nadA*, is 865 bp in *A. parasiticus* but 465 bp in the other species. A similar sugar cluster is not found adjacent to either end of the *A. nidulans* ST cluster. The four immediate ORFs at the *stsA* end of the ST cluster are predicted to encode a possible transporter (AN7826, *E*-value = $4e^{-37}$), a glucuronyl hydrolase (COG4225), a large terminal repeats (LTR) retrotransposon (AN7830, *E*-value = $6e^{-14}$), and a chromodomain-containing protein.
At the other end of the ST cluster the closest putative ORF (7-5 kb downstream) is predicted to encode an MFS transporter (AN7800) which is not homologous to AflT (<22% aa similarity) nor to the MFS transporter (13% aa similarity for the AF13 protein) encoded by the gene upstream of the AF cluster in AF13 and BN008R.

Phylogenetic relationships

Identical relationships among the different AF-producing Aspergillus taxa were predicted from alignment datasets of the intergenic region, concatenated intron, and coding sequences (Fig. 4a–c). The two A. flavus sclerotal

Fig. 3 Schematic depiction of putative genes upstream of the aflatoxin cluster. The lengths of DNA sequenced upstream from the norB stop codon were: Aspergillus nomius, 6861 bp, Aspergillus flavus AF13, 10 589 bp, the unnamed taxon from Benin, West Africa BN008R, 12 295 bp. The coding regions were identified using DNAMAN software. The most highly conserved domains identified by BLAST searches of GenBank and highest homology matches of the predicted proteins to proteins in the WICGR Database are indicated on the arrows, which show the direction of transcription. Accession numbers and full domain designations for these genes are XynB, XynB CBD IV domain, EAA57549, E-value 1E-148; MFS, MFS efflux transporter domain, EAA63077, 1E-112; P450 monooxygenase, EAA48365, 1E-102; aa oxidase, L-amino acid oxidase domain, EAA61261, 4E-87

Fig. 4 Phylogenetic trees based on alignments of aflatoxin cluster sequence. The trees shown in panels (a–c) were obtained from alignments of combined intergenic region, intron and coding sequence, respectively, using the maximum likelihood option implemented in the PAUP. The tree shown in panel (d) was obtained from the alignment of the ver-1 coding region sequence; trees for other individual genes (not shown) gave essentially the same pattern as that for the combined coding region sequence. Distances were used to compute divergence times for Aspergillus parasiticus and Aspergillus flavus. The trees are rooted with Aspergillus nomius. The number of nucleotide substitutions per site is shown above and the bootstrap support values based on 1000 replicates are shown underneath the branches. AP, A. parasiticus ATCC56775; AN, Aspergillus nomius NRRL13137. The unnamed taxon isolate (BN008R, ATCCMYA-379) is from West Africa
mophotypes are not predictably separated into clades with high bootstrap support based on any of these analyses. Although the tree obtained for coding sequence predicted similar phylogeny when all genes were included in the dataset, the topology of this tree was significantly different ($P = 0.001$) from the topology of the ver-$1$ tree, where the *A. flavus* S-strain isolate AF70 and the unnamed taxon isolate BN008R are predicted to be in the same clade (Fig. 4d). The predicted cladal structure for most of the other coding sequences is the same as that of the combined gene dataset. To obtain an estimate of the time of divergence of the different taxa, a tree based on ITS sequence included the Sordariomycete, *Neurospora crassa*, as the outgroup taxon and the Eurotiumycete *Aspergillus* species. Assuming the divergence of the Eurotiumycetes and Sordariomycetes to be between 310 and 670 Ma (Berbee and Taylor 2001; Heckman et al. 2001), the divergence of *A. nidulans* and *A. nomius* can be estimated to be between 99 and 217 Ma and the divergence of *A. nomius* and *A. parasiticus* to be between 25 and 55 Ma. Using distance parameters from phylogenetic trees based on either the intron or intergenic region sequence (Fig. 4a,b) and assuming linearity of the nucleotide substitution rate, the divergence of *A. parasiticus* and the unnamed taxon isolate BN008 from *A. flavus* is estimated to be between 8 and 17 Ma and the divergence of AF70 from AF13 to be between 1 and 3 Ma.

**DISCUSSION**

This study compared, for the first time, the complete nucleotide sequences of the gene clusters harbouring the AF biosynthesis genes of several closely related AF-producing species, including *A. parasiticus*, two morphotypes of *A. flavus*, *A. nomius*, and an unnamed taxon from West Africa. The results show that gene order, location of AflR-binding sites, and intergenic distances were well conserved for more than 25 million years of divergence of species within *Aspergillus* section Flavi. This level of conservation is remarkable for a cluster directed at production of highly toxic and mutagenic metabolites with no clear adaptive value. Such conservation suggests that AF production has provided an important adaptive benefit over most of this time and that perhaps only recently has the loss of cluster components occurred, for example, the loss of genes for production of G AFs in *A. flavus* following divergence from *A. parasiticus* <17 million years ago. *A. flavus* is the most common member of section Flavi in agricultural and natural environments (Boyd and Cotty 1997; Cotty 1997). Many members of *A. flavus* have lost all ability to produce AFs. Thus, AF-producing ability appears unnecessary to account for the observed *A. flavus* success in occupying certain plant-associated niches (Cotty 1989, 1997). However, the high conservation of cluster components needed for B AF production along certain *A. flavus* lineages and conservation of components for B and G AF production in *A. nomius* and *A. parasiticus* suggests important adaptive value for AFs in character-shaping niches important to those taxa. Although, the data suggests that, over the last several million years, the average adaptive value of AFs to section Flavi may have decreased, this decrease may have been caused by the movement into new niches or loss of formative niches in the current environment.

Phylogenetic analyses of combined sets of AF cluster introns, intergenic spaces, and/or coding regions resulted in single most parsimonious trees with identical topologies and very strong cladal bootstrap support. Parsimony analysis of individual coding regions resulted in trees with similar topologies with one exception. The coding region of *ver-1* produced a single most parsimonious tree (Fig. 4d) with a 92% bootstrap-supported clade containing the unnamed taxon BN008R from West Africa and the *A. flavus* isolate AF70 (both having sclerotial morphotype S). The partition homogeneity test revealed significant incongruency in the topology of combined coding sequence dataset, which is resolved when the *ver-1* dataset is excluded from the test. This homoplasy may result from *ver-1* experiencing selective pressures distinct from other cluster genes. Furthermore, the proportion of nonsynonymous compared with synonymous changes was also much lower for *ver-1* (ratio = 0.05) than for any of the 25 other AF cluster genes examined (ratio range = 0.2–1.04) suggesting that the most intense purifying selection was at this locus. The unnamed taxon and the S strain of *A. flavus* have remarkable similarities in sclerotial morphology and habit of sclerotial production. Indeed, when first described, even though these two taxa produce different spectra of AFs, they were grouped as one putative species (Hesseltine et al. 1970) and they have been referred to as strains SBG and SB of *A. flavus* (Cotty and Cardwell 1999). The results of the current study suggest that conservation of the S sclerotial character may be associated with strict conservation of the *ver-1* protein, an enzyme required for one step of the conversion of versicolorin A to ST. The section Flavi proteins encoded by this gene have a 61–64% identity to a melanin biosynthesis protein T$_4$HN of *Magnaporthe* (Vidal-Cros et al. 1994) and interrelationships between sclerotial morphogenesis and AF biosynthesis have long been suggested (Cotty 1988; Chang et al. 2002; Calvo et al. 2004). The gene *ver-1* might be an important link between these two processes and conservation of *ver-1* among taxa with the S morphotypes might indicate its participation in sclerotial maturation and possibly melanization.

The ratio of nonsynonymous to synonymous nucleotide substitutions between *A. nomius* and *A. flavus* AF cluster genes varies in a gene-specific manner not related to position within the cluster (Table 1, Fig. 2). For the AF cluster
genes, the proportion of nonsynonymous changes between the three A. flava strains and A. nomius increases as per cent identity of the A. flavus genes to the corresponding A. nidulans genes decreases (Fig. 2). High identity between ST and AF cluster genes reflects purifying selection during divergence of these two clusters. The low proportion of nonsynonymous substitutions between A. nomius and A. flavus for the genes with highest similarities between the ST and AF clusters suggests continuation of purifying selection during divergence of AF producers in Aspergillus section Flav. Conversely, those genes with lower identity between A. flavus and A. nidulans display the greatest proportion of nonsynonymous substitutions suggesting either increased positive selection (adaptive change) on gene structure during divergence within section Flav or lower functional constraints on gene character either because of loss of selection or lower requirements for specific gene structure. Purifying selection and gene duplications have been invoked to explain the formation of gene families and gene clusters in other eucaryotes (Lynch and Conery 2000; Ober and Hartmann 2000; Wheeler et al. 2001; Kroken et al. 2003), for example, the formation of the globin and the olfactory receptor gene clusters in mammals (Mombaerts 1999; Wheeler et al. 2001) and the presence of both purification and selection are evident here.

Although gene order within the AF cluster is highly conserved among AF producers in section Flav, the gene order differs markedly from that of the ST cluster of A. nidulans (Fig. 1a,b). In addition, the AF cluster contains three pairs of sister genes (norB and norA; hypB1 and hypB2, and omtA and omiB) that occur only singly in the ST cluster. Differences between AF and ST clusters in both gene order and homologous gene pairs may have resulted from gene duplication and/or adaptive translocation events after divergence of the AF and ST clusters. Such events have been shown to occur frequently during adaptation to new environmental stimuli (Dunham et al. 2002; Zuniga et al. 2002; Schmidt et al. 2003). The current results show not only remarkable conservation of gene order within the AF cluster, but also uniform conservation of an adjacent sugar utilization cluster adjacent to the distal end. This strong conservation of sequence suggests that a higher-order chromatin structure embracing both clusters could be important for expression of the genes in the clusters. If indeed both clusters are part of one chromatin domain, this could explain coordinate induction of expression of both clusters when simple sugars are used as the carbon source, by fostering a region of ‘active chromatin’ (Forsberg and Bresnick 2001; Fu et al. 2002). Co-regulation may involve a cis-acting enhancer element or locus control region outside of both clusters. Such regions are well-known in other eucaryotes and are necessary in these organisms to maintain cluster organization (Molete et al. 2001; Fu et al. 2002), but have not been described in fungi. This explanation for the observed sequence conservation beyond the proximal end of the AF cluster also offers an attractive explanation for locus-specific expression of AF cluster genes demonstrated by reduced expression of genes inserted more than 500 bp from the cluster when compared with insertion within the cluster (Chiou et al. 2002; Yu et al. 2004b). This one-domain–two cluster hypothesis is also compatible with the apparent glucose independence of induction of the ST cluster in A. nidulans, which lacks an adjacent sugar cluster (Barnes et al. 1994; Feng and Leonard 1998; Guzman-dePena et al. 1998), and the lack of downregulation of ST genes when inserted at loci remote from the gene cluster (Keller et al. 1994; Keller and Adams 1995). In contrast to the conservation of regions adjacent to the distal end of the AF cluster, regions adjacent to the proximal end differ markedly in gene composition between the three taxa examined in section Flav. Sequence variation in genes outside of the proximal end of the AF cluster suggests that this region is not under the same evolutionary constraints as are the genes in the AF and sugar clusters.

ACKNOWLEDGEMENTS

We thank Mark Rubenfield and Erik Gustafsen, Agencourt Bioscience Corporation for preparation of the fosmid libraries and sequencing. We also thank Beverly Montalbano for her technical assistance.

REFERENCES


