The genome and the genes of *Epichloë festucae*

U. HESSE[1], P. MAYNARD[2], S. MACMIL[3], G. WILEY[3], K. ANDREEVA[1], W.E. BEECH[4], E. ARNAOUDOVA[2], B.T. WILLEY[2], V.-G.PURAM[5], J. WISEMAN[5], J. WEBB[1], L. GILL[1], M.L. FARMAN[1], J.W. JAROMCZYK[2], B.A. ROE[3] and C.L. SCHARDL[5]
[1]*University of Kentucky, Department of Plant Pathology, 201 PSB, 1405 Veterans Dr., Lexington, KY 40546-0312, U.S.A.*
[2]*University of Kentucky, Department of Computer Science, 775 F. Paul Anderson Tower,Lexington, KY 40506-0046, U.S.A.*
[3]*University of Oklahoma, Department of Chemistry and Biochemistry, 20 Parrington Oval, Rm 208, Norman, OK 73019-3051, U.S.A.*
[4]*University of Kentucky, Academic Technology, 210 McVey Hall, Lexington, KY 40506-0045, U.S.A.*
[5]*University of Kentucky Advanced Genetic Technologies Ctr, 113 PSB, 1405 Veterans Dr., Lexington, KY 40546-0312, U.S.A.*

Abstract

The ascomycete *Epichloë festucae* is a model endophyte that 1) switches between mutualistic and antagonistic states, 2) is seed transmissible, 3) has a sexual state amenable to genetic analysis, and 4) is rich in bioprotective alkaloids. This fungus grows systemically and intercellularly throughout the life of its host plant. On each reproductive tiller the fungus either infects benignly and transmits clonally in seeds, or produces its sexual state (stroma) and chokes inflorescence development. The *E. festucae* genome was estimated at 29 Mb in six chromosomes. The genome sequence was assembled from cloned insert end reads (4.2 x coverage) and preassembled pyrosequencing reads (454-sequencing: 20 x raw, 1.7 x assembled), giving 3967 supercontigs, of which 1004 were larger than 2 kb and covered 92% of the genome. Gene prediction with FGENESH identified ~10,000 putative genes. We also sequenced 25,000 ESTs from each of two normalised libraries — one of choked inflorescences, the other of benignly infected inflorescences — yielding 5077 *E. festucae* unigenes, annotated by BLAST and InterPro. Sequence data and annotations are stored in a database for visualisation and inspection with the GBrowse browser. The genomic sequences can be queried by BLAST at http://www.genome.ou.edu/blast/ef_blastall.html.

Keywords: bioinformatics, DNA sequence, *Epichloë festucae,* expressed sequence tags, *Festucae pratensis*, fungal genomics, *Lolium pratense*

Introduction

The ascomycete, *Epichloë festucae* (Leuchtmann *et al*. 1994), is an endophyte of cool season grasses which grows systemically and intercellularly within its host for most of its life. With the development of inflorescences it displays a dual transmission mode: in some reproductive tillers it retains the benign growth and is transmitted vertically through seeds, while on other reproductive tillers of the same plant it completes its sexual life cycle by enveloping and ultimately choking the inflorescence with a mycelial layer (stroma) which serves as a cradle for ascogenous perithecia (reviewed in Schardl & Leuchtmann 2005). The fungus is rich in bioprotective alkaloids (Leuchtmann *et al*. 2000) that protect the host against insects (Funk & White 1997) and mammals (Bazely *et al*. 1997), and was shown to increase the drought stress tolerance of its host grass, *Festuca rubra* (Bazely *et al*. 1997). It therefore represents an intermediate between the seed transmitted, often mutualistic, asexual endophytes (e.g. *N. coenophialum* and *N. lolii*, two species used commercially to improve sustainability of their respective host grasses, *Lolium arundinaceum* and *Lolium perenne*), and the more parasitic *Epichloë* species that rarely infect seeds, but virtually terminate host reproduction. In fact, phylogenetically, *E. festucae* is next of kin to *N. lolii* (Schardl *et al*. 1994), and a parent to several other asexual endophytes, including *N. coenophialum*, which evolved through hybridisation of two or more epichloë endophyte species (Kuldau *et al*. 1999, Moon *et al*. 2004). *Epichloë festcuae* has become a very useful model for genetic and molecular genetic analysis of grass endophytes (Kutil *et al*. 2004; Schardl 2001; Tanaka *et al*. 2005; Wilkinson *et al*. 2000; Young *et al*. 2005). For these reasons, *Epichloë festucae* is a suitable model for studying the genomic background of mutualistic and antagonistic fungal-plant symbioses, and was selected for genome and EST sequencing.

Material and Methods

Fungal isolate for genome analysis

The isolate E2368 was selected for genome sequencing due to its reliable production of stromata and a seed transmission rate of 100% in meadow fescue (*Lolium* pratense = *Festuca pratensis*). Furthermore, this isolate has the gene clusters for loline alkaloid and ergot alkaloid production, and belongs to the mating type mat1.

Symbiotic plant material for EST sequencing

Asymptomatic inflorescences and stromata were obtained from a population of *L. pratense–E. festucae* symbiota. Meadow fescue was chosen as a model host because of its diploid nature, its interfertility with other *Lolium* species (Darbyshire & Warwick 1992), and its role as a genome contributor to *Lolium arundinaceum* (=*Festuca arundinacea*, tall fescue) (Humphreys *et al*. 1995; Xu & Sleper 1994), the most widely planted forage, pasture, turf, and conservation grass in the United States (Ball *et al*. 1993).

Genomic library

For library construction, nuclear DNA was enriched by bisbenzimide-CsCl isopycnic ultracentrifugation, randomly-sheared, twice gel-fractionated to select DNA fragments of 3.5-4.5 kb, and cloned into pBCKS+ (Stratagene Cloning Systems, La Jolla, California). The library consisted of 5–6 million clones, of which 2.5 million cfu were stored at -80°C as aliquots of transformed T1-phage resistant cells (Electromax DH10B; Invitrogen Corp., Carlsbad, CA, USA), and the remainder as ligation mixture. Sequence analysis indicated that 99.76% of the clones contained inserts, with the remaining 0.24% having *E. coli* sequence or (more commonly) no insert.

Normalised cDNA libraries

Asymptomatic fully emerged inflorescences were collected from symbiotic meadow fescue plants at the developmental stages, preanthesis, postanthesis, and early dough stage of the seeds. Similarly, stromata at stages ranging from the earliest visible manifestation of choke to mature stromata with perithecia, were selected for the stromata library. From each of the tissue samples, total RNA was extracted, and the expression of the fungal genes *dmaW* (involved in ergot alkaloid production (Wang *et al*. 2004)) and *lolC* (necessary for loline production (Spiering *et al*. 2005)) were determined by PCR-analysis of the cDNA. To account for allelic differences, RNA from four genetically different symbiota (both plant and endophyte were distinct geneotypes) were mixed at similar concentrations. Normalised

cDNA library construction was conducted by Clontech Laboratories, Inc. (Mountain View, California).

Clone-end sequencing

Transformant colonies were picked by the QPix robot (Genetix, Hampshire, UK) into 96-deep-well plates with 2x YT medium (1.5 ml per well), and grown overnight in a HiGro (GeneMachines, San Carlos, California) oxygenated shaking incubator for microtiter plates. The plasmids were purified robotically (Biomek FX, Beckman Coulter Inc, Fullerton, California) with the Perfect-Prep Plasmid 96 kit (Eppendorf AG, Hamburg Germany). Sequence reactions and capillary electrophoresis were conducted in 96-well plates with two wells reserved for control reactions (pUC18 control and test sample). For the genomic library, both ends of each plasmid were sequenced using vector primers and BigDye3.1 (Applied Biosystems, Foster City, California) at 1/16th reaction strength. For the cDNA libraries, only the 5` ends of the plasmids were sequenced. The reactions were cleaned by ethanol precipitation and capillary electrophoresis was performed in a 3730 DNA analyzer (Applied Biosystems).

Pyrosequencing

Genomic DNA was nebulised, end repaired and adaptors ligated, which provide the priming sequences for amplification and sequencing of the sample library fragments. After DNA immobilisation on streptavidin beads (facilitated by a 5' biotin tag on Adaptor B) and DNA denaturation, the single-stranded DNA library was PCR-amplified and sequenced on a GS20 instrument (454 Life Sciences, Branford, Connecticut).

Computational analysis

Genome assembly was conducted using Arachne 2.0.1 (Broad Institute, Cambridge, Massachusetts). The pyrosequencing reads were first preassembled using Newbler 1.0.53 (454 Life Sciences). The resulting contigs were divided into 1 kb pieces with 500 bp overlaps (since Arachne 2.0.1 does not accept long contigs), and were then assembled together with the clone-end reads. FGENESH (Softberry, Inc., Mount Kisco, New York) was employed for gene prediction. The cDNA sequences underwent low quality and vector masking (cross-match), and were then assembled using phrap (Ewing and Green 1998, Gordon *et al*. 1998). The genomic sequences as well as the unigenes were annotated by BLAST (NCBI, Bethesda, Maryland) and Apple/Genentech (A/G) BLAST (Apple's Advanced Computation Group), and by InterPro (EBI, Cambridge, UK). To provide high performance computing both programs were executed on a cluster of more than 250 Apple Macintosh computers, using Apple's XGrid framework and customized wrappers developed at the University of Kentucky (Beech *et al*. 2006). Gene Ontology terms were assigned using blast2go (Conesa *et al*. 2005).

Results and Discussion

Genome sequencing

A total of 148,112 clone-end reads were sequenced, of which 141,604 high quality reads (867 ± 222 high quality bases) were included in the assembly. Considering an approximate genome size of 29 ± 3.5 Mb, this resulted in an estimated 4.2 x genomic coverage. After assembly with the 61,713 contigs from pyrosequencing (20 x raw coverage, 1.7 x preassembled), 3967 supercontigs covering 30.5 Mb were obtained. The first 500 supercontigs contained 86% of the genome, and 92% of the estimated genome size was covered by the 1004 supercontigs larger than 2 kb. The supercontig length ranged from 489 kb to 0.5 kb.

Pyrosequencing contigs did not cover areas of repeats and sequences rich in long stretches of single nucleotides, most of which were noncoding DNA. Clone-end reads tended to be scarce in areas coding for some secondary metablolism genes (e.g. no reads covered *lolC* from the loline biosynthesis gene cluster). The results indicate that both sequencing methods complement each other, and linkage information is essential for genome assembly conducted with pyrosequencing reads only.

The assembled genomic sequences can be queried by BLAST at http://www.genome.ou.edu/blast/ef_blastall.html.

Table 1 Genes identified on the mitochondrial genome of *E. festucae*.

Best meaningful Blastx hit against supercontig
1. I-PcI endonuclease [*Podospora curvicolla*]
2. ribosomal protein S3 [*Fusarium oxysporum*]
3. NADH dehydrogenase subunit 6 [*Metarhizium anisopliae*]
4. cytochrome oxidase subunit III [*Metarhizium anisopliae*]
5. putative maturase [*Cryphonectria parasitica*]
6. ATP synthase subunit 6 [*Niesslia exilis*].
7. ribosomal protein 3/homing endonuclease-like protein fusion [*Ophiostoma novo-ulmi*]
8. NADH dehydrogenase subunit 2 [*Hypocrea jecorina*].
9. probable maturase [*Hypocrea jecorina*].
10. NADH dehydrogenase subunit 3 [*Metarhizium anisopliae*]
11. cytochrome oxidase subunit II [*Metarhizium anisopliae*]
12. GIY COII i1 grp IB protein [*Podospora anserina*]
13. Dod ND5 i3 grp IB protein [*Podospora anserina*]
14. cytochrome oxidase subunit I [*Verticillium dahliae*].
15. GIY COI i14 grp IB protein [*Podospora anserina*]
16. COX1-I7 [*Penicillium marneffei*]
17. NADH dehydrogenase subunit 5 [*Metarhizium anisopliae*].
18. unknown [*Neurospora crassa*].
19. Probable intron-encoded endonuclease bI1 precursor. [*Ustilago maydis*]
20. cob gene intronic ORF [*Neurospora* mitochondria]

Table 2 Genes identified on supercontigs (Sc) with telomeric repeats.

Sc #	Sc length	Unigene #	Best meaningful blast match (Blastx against nr NCBI)
Sc1	36 kb	1	protein kinase [*Aspergillus fumigatus* Af293]
		2	hypothetical protein CIMG_02925 [*Coccidioides immitis* RS]
		3	putative SRPK1-like protein kinase [*Candida albicans* SC5314]
		4	sarcosine oxidase [*Cylindrocarpon didymum*]
		5	no hits found
		6	enoyl-CoA hydratase [*Aspergillus fumigatus* Af293]
		7	MepB [*Aspergillus fumigatus*]
		8	metallopeptidase MepB [*Aspergillus fumigatus* Af293]
Sc2	31 kb	9	transferase family protein
		10	hypothetical protein FG03735.1 [*Gibberella zeae* PH-1]
		11	dehydrogenase [*Aspergillus fumigatus* Af293]
		12	L-aminoadipate-semialdehyde dehydrogenase [*Coccidioides immitis*]
		13	MFS transporter [*Aspergillus fumigatus* Af293]
Sc3	30 kb	14	peptide synthetase [*Metarhizium anisopliae*]
Sc4	27 kb	15	no hits found
		16	no hits found
		17	hypothetical protein [*Aspergillus fumigatus*]
		18	kinase
		19	predicted protein [*Aspergillus terreus* NIH2624]
Sc5	20 kb	20	no hits found
		21	hypothetical protein FG03555.1 [*Gibberella zeae* PH-1]
		22	no hits found
Sc6	5 kb	23	Wdr1p [*Gibberella moniliformis*]
		24	no hits found
Sc7	8 kb	No hits	N/A
Sc8	1 kb	No hits	N/A

Genome annotation

Gene prediction was conducted using FGENESH trained with data from *Neurospora crassa* (Nc) and *Fusarium graminearum* (Fg). When run with default options, 9912 and 9741 predicted genes were obtained, respectively. These predictions are in accordance with the expected number of genes (ca. 10,000) as observed in related fungal genomes of similar size. Preliminary analyses indicate that intron-exon boundaries were predicted more accurately when the Fg parameter file was used. The unigene set (see below), as well as blast matches against the nr database at NCBI and Interpro-domains will be used to validate the predicted proteins.

Plans are to make the annotated genome available for visualisation and inspection via GBrowse (The Generic Model Organism Database Project; http://www.gmod.org/).

The mitochondrion

The *E. festucae* mitochondrial genome reads assembled into seven contigs covering 48 kb. So far, 20 mitochondrial genes have been identified using blastx against the nr database of NCBI (Table1).

Telomeres

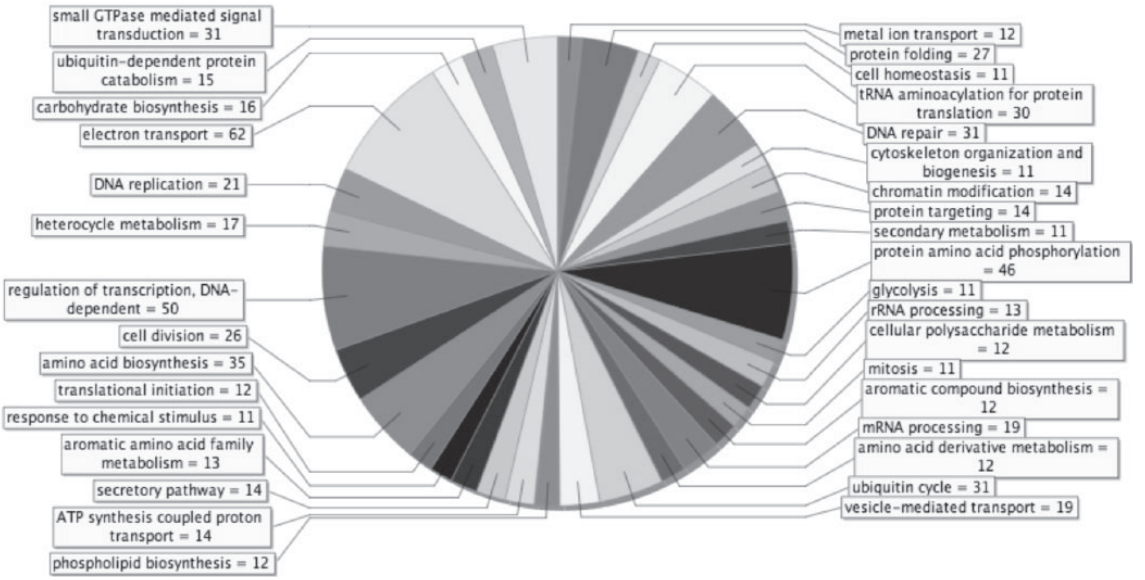
Southern blot analysis of genomic DNA digested with restriction enzymes and hybridised with telomere probes indicated that *E. festucae* probably has six chromosomes (K. Aandreeva, U. Hesse and M.L. Farman, unpublished). So far, eight supercontigs containing telomeric repeats have been identified in the assembled genome sequence. It is possible that sequence similarity of different telomeres resulted in co-assembly, and further sequence analysis is necessary to find all telomeres. The genes found on the supercontigs mainly encoded regulatory proteins, proteins involved in signal transduction and metabolism (Table 2).

Unigenes

Approximately 25,000 reads were sequenced from each of the two normalised cDNA libraries (from benignly infected inflorescences and stromata), respectively. The reads were mapped to the genome, resulting in 2055 hits from the inflorescence library and 10,832 hits from the stromata library. The difference in numbers is not surprising, considering that mycelium is clearly much more abundant in stromata. Assembly of the fungal reads resulted in identification of 5077 fungal unigenes. Sequence analysis indicated 85 pairs of overlapping unigenes encoded on opposite DNA strands. Furthermore, 308 differentially spliced unigenes were identified. A total of 4327 unigenes had significant blastx matches. Significant InterPro matches were assigned to 1860 unigenes, and 1804 sequences had validated GO terms.

Analysis of the sequence distribution (Fig.1) gives insight into the different physiological processes happening in the fungus. A high proportion of genes were found to be involved in electron transport, regulation of transcription, protein amino acid phosphorylation, and small GTPase mediated signal transduction. These are of particular interest, as they may be involved in the regulation of gene expression and fungal growth in the host plant. Secreted proteins and proteins involved in secondary metabolism are likely to affect the symbiotic interaction and its consequences

Figure 1 Distribution of fungal unigenes by GO terms for physiological processes (filtered by # of sequences: cutoff=10)



to host ecology. These unigenes provide a valuable gene pool for studying and understanding fungal/plant interactions.

ACKNOWLEDGEMENTS

The projects were supported by US National Science Foundation grant EF-0523661, US Department of Agriculture grants 2005-35319-16141 and 2005-34457-15712, and the Kentucky Biomedical Research Infrastructure Network, NIH-IDeA-INBRE Grant 2 P20 RR-16481 from the National Center for Research Resources.

REFERENCES

Ball, D.M.; Pedersen, J.F.; Lacefield, G.D. 1993. The tall-fescue endophyte. *American Scientist 81*: 370-379.

Bazely, D.R.; Vicari, M.; Emmerich, S.; Filip, L.; Lin, D.; Inman, A. 1997. Interactions between herbivores and endophyte-infected *Festuca rubra* from the Scottish islands of St. Kilda, Benbecula and Rum. *Journal of Applied Ecology 34*: 847-860.

Beech, W.E.; Bumgardner, V.C.; Hesse, H.; Van Horn, C.D.; Jaromczyk, J.W.; Schardl, C.L. 2006. XGrid Implementation of the InterProScan for the *Epichloë festucae* genome project. pp. *In*: 14th Annual International Conference on Intelligent Systems for Biomolecular Biology, ISMB 2006.

Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics 21*: 3674-3676.

Darbyshire, S.J.; Warwick, S.L. 1992. Phylogeny of North American *Festuca* (Poaceae) and related genera using chloroplast DNA restriction-site variation. *Canadian Journal of Botany 70*: 2415-2429.

Ewing, B.; Hillier, L.; Wendl, M.C.; Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research 8*: 175-185.

Funk, C.R.; White, J.R., Jr. 1997. Use of natural and transformed endophytes for turf improvement. pp. 229-239. *In*: *Neotyphodium*/Grass Interactions. Eds. Bacon, C. W.; Hill, N. S. Plenum Press, New York.

Gordon, D.; Abajian, C.; Green, P. 1998. Consed: a graphical tool for sequence finishing. *Genome Research 8*: 195-202.

Humphreys, M.W.; Thomas, H.M.; Morgan, W.G.; Meredith, M.R.; Harper, J.A.; Thomas, H.; Zwierzykowski, Z.; Ghesquiére, M. 1995. Discriminating the ancestral progenitors of hexaploid *Festuca arundinacea* using genomic in situ hybridization. *Heredity 75*: 171-174.

Kuldau, G.A.; Tsai, H.-F.; Schardl, C.L. 1999. Genome sizes of *Epichloë* species and anamorphic hybrids. *Mycologia 91*: 776-782.

Kutil, B.L.; Liu, G.; Vrebalov, J.; Wilkinson, H.H. 2004. Contig assembly and microsynteny analysis using a bacterial artificial chromosome library for *Epichloë festucae*, a mutualistic fungal endophyte of grasses. *Fungal Genetics and Biology 41*: 23-32.

Leuchtmann, A.; Schardl, C.L.; Siegel, M.R. 1994. Sexual compatibility and taxonomy of a new species of *Epichloë* symbiotic with fine fescue grasses. *Mycologia 86*: 802-812.

Leuchtmann, A.; Schmidt, D.; Bush, L.P. 2000. Different levels of protective alkaloids in grasses with stroma-forming and seed-transmitted *Epichloë/Neotyphodium* endophytes. *Journal of Chemical Ecology 26*: 1025-1036.

Moon, C.D.; Craven, K.D.; Leuchtmann, A.; Clement, S.L.; Schardl, C.L. 2004. Prevalence of interspecific hybrids amongst asexual fungal endophytes of grasses. *Molecular Ecology 13*: 1455-1467.

Schardl, C.L. 2001. *Epichloë festucae* and related mutualistic symbionts of grasses. *Fungal Genetics and Biology 33*: 69-82.

Schardl, C.L.; Leuchtmann, A. 2005. The epichloë endophytes of grasses and the symbiotic continuum. pp. 475-503. *In*: The Fungal Community: its Organization and Role in the Ecosystem. Eds. Dighton, J.; White, J. F.; Oudemans, P. CRC Press, Boca Raton, Florida.

Schardl, C.L.; Leuchtmann, A.; Tsai, H.F.; Collett, M.A.; Watt, D.M.; Scott, D.B. 1994. Origin of a fungal symbiont of

perennial ryegrass by interspecific hybridization of a mutualist with the ryegrass choke pathogen, *Epichloë typhina*. *Genetics 136*: 1307-1317.

Spiering, M.J.; Moon, C.D.; Wilkinson, H.H.; Schardl, C.L. 2005. Gene clusters for insecticidal loline alkaloids in the grass-endophytic fungus *Neotyphodium uncinatum*. *Genetics 169*: 1403-1414.

Tanaka, A.; Tapper, B.A.; Popay, A.; Parker, E.J.; Scott, B. 2005. A symbiosis expressed non-ribosomal peptide synthetase from a mutualistic fungal endophyte of perennial ryegrass confers protection to the symbiotum from insect herbivory. *Molecular Microbiology 57*: 1036-1050.

Wang, J.; Machado, C.; Panaccione, D.G.; Tsai, H.-F.; Schardl, C.L. 2004. The determinant step in ergot alkaloid biosynthesis by an endophyte of perennial ryegrass. *Fungal Genetics and Biology 41*: 189-198.

Wilkinson, H.H.; Siegel, M.R.; Blankenship, J.D.; Mallory, A.C.; Bush, L.P.; Schardl, C.L. 2000. Contribution of fungal loline alkaloids to protection from aphids in a grass-endophyte mutualism. *Molecular Plant-Microbe Interactions 13*: 1027-1033.

Xu, W.W.; Sleper, D.A. 1994. Phylogeny of tall fescue and related species using RFLPs. *Theoretical and Applied Genetics 88*: 685-690.

Young, C.; Bryant, M.; Christensen, M.; Tapper, B.; Bryan, G.; Scott, B. 2005. Molecular cloning and genetic analysis of a symbiosis-expressed gene cluster for lolitrem biosynthesis from a mutualistic endophyte of perennial ryegrass. *Molecular Genetics and Genomics 274*: 13-29.