Available online at www.sciencedirect.com
ScienceDirect
Fungal Genetics and Biology 43 (2006) 655–666
FUNGAL GENETICS AND BIOLOGY
www.elsevier.com/locate/yfgbi

Mating factor linkage and genome evolution in basidiomycetous pathogens of cereals

Guus Bakkeren [a,*], Guoqiao Jiang [b], René L. Warren [c], Yaron Butterfield [c], Heesun Shin [c], Readman Chiu [c], Rob Linning [a], Jacqueline Schein [c], Nancy Lee [b], Guanggan Hu [b], Doris M. Kupfer [d], Yuhong Tang [d], Bruce A. Roe [d], Steven Jones [c], Marco Marra [c], James W. Kronstad [b]

[a] *Pacific Agri-Food Research Centre, Agriculture and Agri-Food Canada, Summerland, BC, Canada V0H 1Z0*
[b] *The Michael Smith Laboratories, Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada V6T 2Z4*
[c] *Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada V5Z 4S6*
[d] *Advanced Center for Genome Technology, University of Oklahoma, Stephenson Research and Technology Center, Norman, OK 73019-0370, USA*

Received 1 March 2006; accepted 7 April 2006
Available online 21 June 2006

Abstract

Sex in basidiomycete fungi is controlled by tetrapolar mating systems in which two unlinked gene complexes determine up to thousands of mating specificities, or by bipolar systems in which a single locus (*MAT*) specifies different sexes. The genus *Ustilago* contains bipolar (*Ustilago hordei*) and tetrapolar (*Ustilago maydis*) species and sexual development is associated with infection of cereal hosts. The *U. hordei MAT-1* locus is unusually large (~500 kb) and recombination is suppressed in this region. We mapped the genome of *U. hordei* and sequenced the *MAT-1* region to allow a comparison with mating-type regions in *U. maydis*. Additionally the rDNA cluster in the *U. hordei* genome was identified and characterized. At *MAT-1*, we found 47 genes along with a striking accumulation of retrotransposons and repetitive DNA; the latter features were notably absent from the corresponding *U. maydis* regions. The tetrapolar mating system may be ancestral and differences in pathogenic life style and potential for inbreeding may have contributed to genome evolution.

Keywords: BAC mapping; *MAT* locus; Retrotransposons; rRNA unit; Sex chromosome; *Ustilago hordei*

1. Introduction

Ustilago hordei is a fungal pathogen of small grain cereals that is generally found in nature as black masses of diploid teliospores on infected floral tissue of the host (Fisher and Holton, 1957). Teliospores germinate and undergo meiosis to produce haploid progeny that segregate for the mating-type locus (*MAT*). Haploid cells of opposite mating type fuse to form the infectious dikaryotic cell type that grows with the developing seedling, proliferates extensively within floral tissue and eventually forms masses of teliospores in place of the seeds (Hu et al., 2002). Thus, mating is required for formation of the infectious cell type and infection of an appropriate host is necessary for completion of the sexual phase of the life cycle. The mating-type locus is therefore considered to be a pathogenicity locus in *U. hordei* (Kronstad and Staben, 1997).

U. hordei, like most *Ustilago* species, has a bipolar mating system with two opposite specificities called *MAT-1* and *MAT-2* at the *MAT* locus. In contrast, the well-characterized species *Ustilago maydis* has a tetrapolar mating system controlled by two unlinked loci designated *a* and *b* (Fisher and Holton, 1957). The *a* locus encodes pheromone and pheromone receptor functions and the *b* locus encodes homeodomain transcription factors (Feldbrugge et al., 2004). We previously found that the *MAT* locus of *U. hordei* also contains gene complexes equivalent to the *a* and *b*

* Corresponding author. Fax: +1 250 494 0755.
E-mail address: BakkerenG@agr.gc.ca (G. Bakkeren).

1087-1845/$ - see front matter Crown copyright © 2006 Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.fgb.2006.04.002

loci of *U. maydis* and that the difference between bipolar and tetrapolar mating systems for these fungi is due to linkage of *a* and *b* within the *MAT* locus in *U. hordei* and the separation of the loci on different chromosomes in *U. maydis* (Bakkeren et al., 1992; Bakkeren and Kronstad, 1993, 1994, 1996; Lee et al., 1999). We also found that recombination between the *a* and *b* gene complexes is suppressed in *U. hordei* and that the *MAT* locus is unexpectedly large with distances between *a* and *b* estimated at 500 kb for *MAT-1* and 430 kb for *MAT-2* (Bakkeren and Kronstad, 1994; Lee et al., 1999). The mechanisms underlying the suppression of recombination are not known, but sequence rearrangements and indels are evident between *MAT-1* and *MAT-2* (Lee et al., 1999). The *MAT* region of *U. hordei* is unusually large compared to the mating-type loci in other fungi. These other loci generally encode transcription factors that control mating and sexual development, although the basidiomycete fungi additionally possess mating-type loci that encode pheromones and pheromone receptors (Kronstad and Staben, 1997; Fraser and Heitman, 2005).

In this report, we describe the construction of a physical map for the *U. hordei* genome, the sequence of the *MAT-1* region and the comparison of this sequence with the mating-type regions of *U. maydis*. Previous characterization of the genome by electrophoretic karyotyping identified 15–19 chromosome-sized bands and a haploid genome size of approximately 20 Mb (McCluskey and Mills, 1990; Abdennadher and Mills, 2000). Chromosome IV was the most variable between strains and the variation (size range 1.6–1.9 Mb) was proposed to result from rearrangements within the rDNA cluster (Gaudet and Kiesling, 1991; McCluskey et al., 1994; Gaudet et al., 1998). The physical map described here represents an important resource for characterization of genomic features such as chromosome length polymorphism and for eventual completion of the genomic sequence. The physical map was also a key resource to identify the clones spanning the ~500-kb *MAT-1* locus, defined here as the region between the known *a* and *b* mating-type gene complexes. Detailed sequence analysis of the *MAT-1* region revealed a remarkable abundance of retrotransposons and repeats, a finding reminiscent of retroelement accumulation on sex chromosomes in higher eukaryotes (Erlandsson et al., 2000). However, this accumulation was not evident in corresponding regions harboring the *U. maydis a* and *b* mating type gene complexes, as determined from the genome sequence of this species, suggesting that a tetrapolar system may be ancestral in this genus.

2. Materials and methods

2.1. BAC clone fingerprinting and physical mapping

A BAC library from *U. hordei MAT-1* strain 4857-4 (Linning et al., 2004) containing 2304 clones with an average insert size of 113,138 bp, corresponding to approximately 11.5 genome equivalents, was used for map construction. High throughput, agarose gel-based BAC fingerprinting, fingerprint map assembly and manual editing were performed as previously described (Marra et al., 1997; Marra et al., 1999; McPherson et al., 2001; Schein et al., 2002) except that restriction fragment identification, fragment mobility and size determination were performed using automated analysis software (Fuhrmann et al., 2003). Additional details are provided in the Supplemental Materials.

2.2. Sequence assembly, annotation, and comparisons

The sequences of the five BAC clones spanning the *MAT-1* locus were obtained as described in the Supplemental Materials. The 526,707 bp *MAT-1* locus was assembled using SeqMan II software from DNASTAR (assembly date: 9 August 2004). The final assembled sequence of 526,707 bp included the overlapping portion between clone H002B07 and a 15,566 bp sequence containing the *a1* gene complex (Accession No. U07939; G. Jiang, unpublished). We also identified the overlapping region between clone H005D09 and a 5674 bp region of the *b1* gene complex (accession Z18532; G. Jiang, unpublished). The repetitive nature of the region made assembly challenging and the BAC end sequences of mapped clones across the region were used as support. However, given the high density of related elements, particularly LTRs, it is possible that some regions within BAC clones may have sequences that are inverted relative to their actual orientation. In addition, four sequence gaps that are estimated to be less than 100 bp each remain in the assembled sequence. Repeat sequences were identified and classified using BLASTn (Altschul et al., 1997) and ClustalW (Thompson et al., 1994). ARTEMIS (Rutherford et al., 2000; release 5; eukaryotic mode) was used to establish the positions and classes of repeats and to annotate the genes present in the sequenced region. ORFs greater than 100 aa were examined using BLASTx 2.2.8 (Altschul et al., 1997; performed on 13 October 2004). The *MAT-1* open reading frames and LTRs/repeats were compared with the corresponding gene models at the MIPS *U. maydis* database (MUMDB; http://mips.gsf.de/genre/proj/ustilago). The assembled *U. maydis* genome and contig information can be found at http://www.broad.mit.edu/annotation/fungi/ustilago_maydis/. The programs GLIMMERM (Majoros et al., 2003) and FGENESH (Salamov and Solovyev, 2000) were also used for gene prediction.

The analysis of conservation of synteny between the *U. hordei MAT-1* region and the *U. maydis* contigs was performed with PatternHunter (Ma et al., 2002) and the diagram in Fig. 2 was generated using the UCSB/BSI Genomic Data Interactive Visualization Utility, version 1.0.4 (L. J. Miller; http://www.cs.ucsb.edu/~ljmiller/bioinf/bioprojects.html) using BLASTn output coordinates. To generate Fig. 4, regions were aligned using CrossMatch (Green, 1994; http://www.phrap.org), a general-purpose utility that uses an implementation of the Smith–Waterman algorithm for comparing DNA sequences. For every comparison, a minimum score of 30 and a minimum seed length of 17

nucleotides were used to nucleate Smith–Waterman alignments. Colinear regions were plotted using XMatchView (R. Warren, unpublished), a versatile python application to (1) identify colinear blocks, (2) view relationships between colinear blocks, (3) assess sequence identity between repeated segments, and (4) view the repeat frequency. Additional methods are presented in the Supplemental Materials.

2.3. Database submissions

The sequence data of the five BAC clones spanning the *MAT-1* locus have been submitted to GenBank Accession Nos.: H002B07 (uhobac-2b7), AC114900; H001C16 (uhobac-1c16), AC114898; H004K21 (uhobac-4k21), AC116560; H001N01 (uhobac-1n1), AC114899; and H005D09 (uhobac-5d9), AC119572. The 526,707 bp *MAT-1* locus was deposited in the EMBL database, Accession No. AM118080.

3. Results

3.1. Construction of a physical map of the U. hordei genome and identification of tiling path clones for the MAT-1 locus

We previously found that the *a* and *b* gene complexes are present on different chromosomes in *U. maydis* compared with linkage at the *MAT* locus on a single chromosome in *U. hordei* (Bakkeren and Kronstad, 1994; Fig. 1). Therefore, differences in the genomic organization of sex-determining loci represent a key aspect of the mating systems in these fungi. In this context, we constructed a physical map of the *U. hordei* genome by BAC fingerprinting (Table 1) to initiate an analysis of the genome, to characterize the *MAT-1* allele at the *MAT* locus and to facilitate a detailed comparison of mating-type sequences between *U. hordei* and *U. maydis*. The genome analysis also revealed the rDNA cluster and we characterized this region in detail in parellel work (see Supplemental materials). The map was used to identify the contig containing the *MAT-1* locus by hybridization with DNA sequences from the *a1* and *b1* gene complexes that we previously characterized from *U. hordei* (Bakkeren and Kronstad, 1994, 1996). Specifically, hybrid-

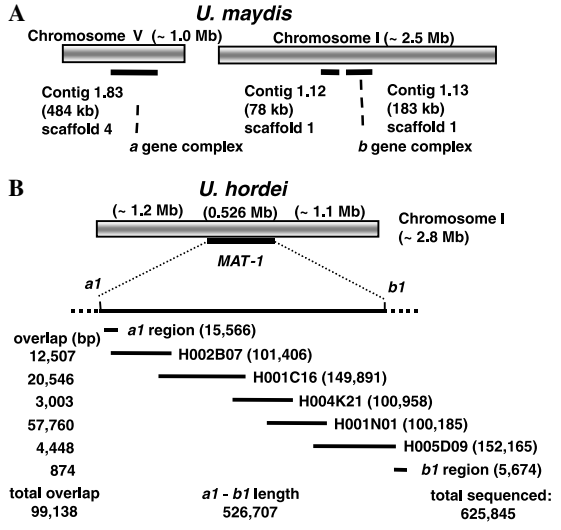

Fig. 1. Genome organization of mating-type loci controlling bipolar and tetrapolar mating in *U. hordei* and *U. maydis*. (A) Location of the *a* and *b* gene complexes on chromosome V and I, respectively, in *U. maydis* and (B) on *MAT-1* in *U. hordei*. The minimal tiling path of five BAC clones spanning the region between the *a1* and *b1* gene complexes of *U. hordei* plus the overlapping *a1* and *b1* regions sequenced previously, are highlighted. The numbers are shown in base pairs and those on the left indicate the extent of overlap between each of the clones used for the assembly.

Table 1
BAC fingerprint contigs and hybridization with selected LTR sequences

Contig data			Hybridization to LTR probes		
Number	# Clones	size (bp)	LTR2	LTR3	LTR9
Contig 10	228	1785060	5 (5)	27 (11)	5 (11)
Contig 101	**131**	**1324529**	**44 (44)**	**32 (13)**	**14 (30)**
Contig 2	147	1313164	4 (4)	23 (9)	14 (30)
Contig 3	130	1280348	11 (11)	20 (8)	0
Contig 4	110	1084447	4 (4)	9 (4)	4 (9)
Contig 20	124	928076	9 (9)	7 (3)	0
Contig 1	109	769735	0	11 (4)	1 (2)
Contig 44	74	731469	2 (2)	5 (2)	0
Contig 65	91	729651	1 (1)	19 (8)	0
Contig 5	73	728275	0	0	0
Contig 266	49	610877	0	10 (4)	0
Contig 96	77	584528	1 (1)	0	0
Contig 37	52	552771	0	2 (0.8)	0
Contig 47	45	503061	0	0	0
Contig 9	57	494948	0	0	0
Contig 95	30	490510	0	12 (5)	0
Contig 100	71	478717	0	1 (0.4)	0
Contig 57	40	464425	4 (4)	1 (0.4)	0
Contig 209	41	434856	8 (8)	15 (6)	0
Contig 76	41	396850	0	13 (5)	2 (5)
Contig 12	41	324445	0	1 (0.4)	0
Contig 55	25	300701	0	0	0
Contig 23	29	282345	0	3 (1)	0
Contig 86	16	281637	0	0	0
Contig 62	18	241350	0	13 (5)	1 (2)
Contig 43	20	236345	0	15 (6)	0
Contig 115	14	216012	0	0	0
Contig 6	26	203613	6 (6)	1 (0.4)	0
Contig 118	9	183022	0	0	0
Contig 195	6	178250	0	4 (1.6)	0
Contig 162	6	173338	0	0	0
Contig 8	12	161655	0	1 (0.4)	0
Contig 160	9	156350	0	0	3 (7)
Contig 148	6	156280	0	0	0
Contig 111	7	148079	0	0	0
Contig 35	22	147371	0	0	0
Contig 186	8	128279	0	1 (0.4)	0
Contig 265	2	66168	0	0	0
Singletons	34		0	1 (0.4)	1
Orphans	6		2 (2)	2 (0.8)	2 (4)
Total		19205369	101	249	46

All contigs are listed in descending order by size in base pairs and those hybridizing to the respective LTR probes are indicated. The number of positive BAC clones for each probe is given with the percentage of the total hybridizing clones in parentheses. Contig 101 containing the *MAT-1* locus is indicated in bold.

ization with the probe aw-1 revealed that BAC clones H006N17 and H006I02 carried the *a1* gene complex, and hybridization with the probe b1 indicated that the *b1* gene complex was located on clones H005D09, H002C14, H004I03 and H006N10. Each of these BAC clones mapped to the same contig (101), as expected from the physical linkage of the *a1* and *b1* gene complexes (Lee et al., 1999). The five BAC tiling clones that span the *MAT-1* locus on contig 101 are shown in Fig. 1B and the BAC clones that make up the complete contig are shown in Supplementary Fig. 5. In addition to the five BAC clones for the *MAT-1* locus, the minimum tiling set for map contig 101 included 14 clones to the left and three clones to the right of the locus. Previous physical mapping (Lee et al., 1999) indicates that the *MAT* locus is approximately in the center of chromosome I (Fig. 1B).

3.2. Assembly and annotation of the MAT-1 sequence

The physical map provided the framework to identify and subsequently sequence the inserts of five BAC clones, H002B07, H001C16, H004K21, H001N01, and H005D09, that spanned the *MAT-1* locus (Fig. 1B). The assembled sequence included flanking sequences at the *a1* and *b1* gene complexes to establish a genomic region of 526,707 bp for subsequent annotation; this length is close to the estimate of 500 kb from CHEF-PFGE[1] analysis (Lee et al., 1999). The assembled sequence for the *MAT-1* region was annotated using ARTEMIS (Methods) to identify candidate genes and to characterize repetitive elements at the locus. The identification of open reading frames (ORFs) greater than 100 aa in length combined with gene predictions revealed 47 candidate protein-coding genes (excluding the retrotransposon coding regions described below; Table 2 and Supplementary Table 4). Twenty of the forty-seven ORFs were found to encode functionally uncharacterized proteins (designated as hypothetical proteins). The remaining candidate genes encode proteins with similarity to known proteins functioning in signalling (e.g., GTPases, glycogen synthase kinase), gene expression (e.g., ribosomal and TATA-binding proteins), and metabolism (e.g., trehalose phosphatase, α-mannosidase, and ferric reductase). The previously identified genes at the *a1* locus included the pheromone receptor gene, *Uhpra1* (Bakkeren and Kronstad, 1994), the pheromone gene, *Uhmfa1* (Anderson et al., 1999) and a pantothenic acid biosynthesis gene, *pan1*. Adjacent genes were identified for a DAHP synthase, an oligopeptide transporter and a ribosomal protein. The *a*1 locus also contained an ortholog of the *rba2* gene found at the *a* locus in *U. maydis* (Bolker et al., 1992; Urban et al., 1996). The homeodomain genes *UhbE1* and *UhbW1* were found at the *b1* locus as expected, along with an adjacent gene for a predicted N-terminal acetyltransferase.

[1] *Abbreviations used:* aa: amino acid; CHEF-PFGE: contour-clamped homogeneous electric field pulsed field gel electrophoresis.

3.3. Repeated sequences and retrotransposons in the MAT-1 region

The analysis of the *MAT-1* sequence revealed a remarkable accumulation of partial and intact retrotransposons and related repeats (LTRs), as well as putative transposons (Table 3 and Supplementary Table 4). The high density of these elements at the *MAT-1* locus (~50% of the total sequence) resulted in islands of one to four genes separated by large regions of repetitive DNA, an organization remarkably similar to that of genes and repeats in the cereal hosts (e.g., barley and wheat) of *Ustilago* pathogens (Fig. 2; Ramakrishna et al., 2002; Anderson et al., 2003). The largest region of contiguous genes (i.e., lacking the identified repeat sequences) contains seven ORFs at the *a1* gene complex on the left end of *MAT-1*. The numerous copies of retrotransposons (~100) in the *MAT-1* region were related to copia or gypsy-type elements as determined by the similarities of the coding sequences to parts of the polyproteins encoded by these elements. The majority of the candidate retrotransposons were partial copies and only 12 appeared to be full length, although each of the predicted coding regions for the polyproteins contained several stop codons suggesting that the elements may not be functional. The frequency of each element within the *MAT-1* region was tabulated (Table 3) and we noted particularly large clusters of elements in a 37-kb region between 295 and 330 kb and in a 64-kb region between 420 and 480 kb (Fig. 2B, Supplementary Table 4). These clusters resemble putative centromeric sequences identified in genome sequencing projects for several fungi including *Cryptococcus neoformans* (Loftus et al., 2005), *U. maydis* (Kahmann and Kamper, 2004) and *Neurospora crassa* (Borkovich et al., 2004).

In addition to repeated copies of the retroelements, the *MAT-1* sequence contained a substantial number of repetitive elements up to 1 kb in length and many of these had characteristics of LTRs; in fact, some were associated with retroelements (Table 3). For example, it was noteworthy that there was a preponderance of the element designated LTR1 in the *MAT-1* sequence (see below) and this element was associated with the retrotransposons Tuh1 and Tuh2. BAC end-sequences generated for the tiling set of 14 clones on the left side of the *MAT-1* region also revealed that the retroelements (Tuh3, 4, and 5), transposons (Tho3) and repeats (LTR1 and 9) were present elsewhere on the chromosome (J. Kronstad, unpublished data). BAC end sequences for the three tiling set clones on the other side of *MAT-1* also detected Tuh3 and several LTRs. These results suggested that the repetitive sequences were not located solely within the *MAT-1* locus and prompted further use of the mapped BAC clones to examine repeat distribution across the genome.

Previously, we found that some repetitive elements, such as a RAPD marker that contained part of an LTR1-like sequence, hybridized widely across the genome when tested on chromosomes separated by CHEF-PFGE (Linning et al., 2004). This result, coupled with the BAC end

Table 2
List of genes located on the *U. hordei MAT-1* locus compared with the *U. maydis* genome

U. hordei gene	Position	Or.	*Um* Ortholog	*e*-Value	Contig	Position	Or.	Scaffold	LG
Pra1	52–1357	–	um02383	e−126	83	(117504–116194)	–	4	5
Rba2	2542–2956	–	um02384	3e−46	83	(118859–118217)	–	4	5
Pan1	3290–4414	–	um10139	e−155	83	(120405–119104)	–	4	5
DAHP synthetase	7203–8384	–	um02385	e−198	83	(124008–122827)	–	4	5
S19 rib. protein	9430–9786	+	um02386	2e−40	83	(124925–125379)	+	4	5
Oligopeptide transporter	10221–13173	–	um02387	0	83	(128502–125545)	–	4	5
Put. tyrosine recombinase	15423–15938	–	None						
*Isocitrate dehydrogense	28533–28994	+	um06111	1e−31	230	(40899–42373)	+	21	21
*Vacuolar protein sorting	37313–40852	–	um10136	0	83	(71389–67849)	–	4	5
Hypothetical (ThiF?)	60541–61362	–	um05834	2e−17	214	(93650–95500)	+	20	20
26S Protease regulatory subunit	80979–82139	–	um10459	0	13	(1137–2384)	+	1	1a
Hypothetical	82806–83372	+	um10458	5e−50	13	(120–515)	–	1	1a
Hypothetical	83713–84807	–	um00545	e−131	12	(77024–78175)	+	1	1a
Zinc finger	113783–117204	–	um02389	0	83	(134025–130609)	–	4	5
GDP/GTP exchange factor?	146212–148205	+	um10463	0	13	(35785–34193)	–	1	1a
Hypothetical	168408–169715	+	um02392	1e−175	83	(148497–147226)	–	4	5
Hypothetical	170190–171257	–	um10141	1e−110	83	(145641–146698)	+	4	5
GTPase act. *Sec2*	197391–200441	+	um00553	0	13	(28440–25309)	–	1	1a
Hypothetical *Sec7*	211798–216855	–	um10461	0	13	(20150–25094)	+	1	1a
Hypoth. α-trehalose phosphatase	224509–228447	–	um02390	0	83	(144142–140099)	–	4	5
Hypothetical	256732–257121	+	None (C. neo)						
Hypothetical	257295–258125	–	um00536	e−122	12	(46270–45443)	–	1	1a
Hypothetical	270231–272144	+	um00537	0	12	(50087–51974)	+	1	1a
Hypothetical	272811–273917	+	um00538	1e−78	12	(54079–55143)	+	1	1a
Hypothetical	274115–279556	–	um00539	0	12	(61982–56514)	+	1	1a
Hypothetical UTR1 ferric reductase	291983–295180		um00549	0	13	(8062–11163)	+	1	1a
Hypothetical	331788–332334	+	um10142	1e−74	83	(149236–150188)	+	4	5
Hypothetical *TFIID*	333456–334354	–	um10143	e−114	83	(151814–150966)	–	4	5
Hypothetical (ferric reductase)	335724–337560	–	um02395	0	83	(154980–153193)	–	4	5
Hypothetical Zn finger (RING)	360513–362729	+	um00542	0	12	(67324–69576)	+	1	1a
Hypothetical	362927–367105	–	um00543	0	12	(73791–69730)	–	1	1a
Put. tyrosine recombinase	380379–380972	+	None						
Hypoth. nucleoside diphosphatase	381827–383770	+	um10460	0	13	(6641–4929)	–	1	1a
Hypoth. Δ1-pyrr-5-carb. reductase	384000–384926	–	um00547	e−144	13	(3827–4750)	+	1	1a
Hypothetical	393820–395171	+	um10468	e−122	13	(108549–109346)	+	1	1a
Hypothetical	401311–402795	+	um00587	e−112	13	(129487–128036)	–	1	1a
Hypothetical	403104–405251	–	um00586	e−144	13	(125486–127657)	+	1	1a
Hypothetical	426939–428216	–	um00562	e−124	13	(50845–49604)	–	1	1a
Hypothetical rib. protein L30	442552–443076	+	um00563	6e−73	13	(53556–54068)	+	1	1a
Hypothetical, related to *Bub2*	469378–470715	+	um00561	e−144	13	(49135–47769)	–	1	1a
Hyp. Ser/Thr prot. kinase (*STE11*)	471494–473004	–	um00560	0	13	(45490–47044)	+	1	1a
Hypothetical	476110–478578	+	None (*M. grisea*)						
AMS1 α-mannosidase	478878–482315	–	um00557	0	13	(37905–41357)	+	1	1a
N-term. acetyltransferase	519537–522227	+	um00579	0	13	(107589–104908)	–	1	1a
bW1 Mating gene	522588–524813	–	um00578	9e−96	13	(102445–104596)	+	1	1a
bE1 Mating gene	525018–526508	+	um00577	9e−97	13	(102190–100958)	–	1	1a

Only ORFs with predicted protein sequences larger than 100 aa were analyzed. Note that *mfa1*, the mating-type pheromone gene, (GenBank Accession No. AF043940) is not listed but is located just upstream of *pra1*. Indicated are their position and orientation (or) of transcription (refer to EMBL accession AM118080). Similarity to *U. maydis* homologs in a BLASTp search with expect values and their location on contig, scaffold (supercontig) and linkage groups (LG, chromosome) according to http://www.broad.mit.edu/annotation/fungi/ustilago_maydis/supercontig_table.html#c1, is also given; only matches at <e−10 are given. Coordinates and annotations of the *U. maydis* homologs can be found at MUMDB (http://mips.gsf.de/genre/proj/ustilago/).
*Genes with insertions of LTR repeats. Refer to Fig. 2; for a complete list including genes from transposable and retro elements, see Supplementary Table 4 which has color-coding matching Fig. 2.

sequencing, raised the possibility that the repeat abundance at *MAT-1* reflected the organization of the entire genome for *U. hordei*. To address this possibility, we examined the distribution of seven representative repeat sequences from *MAT-1* by hybridization of amplified probe sequences to the genome arrayed on filters containing the mapped BAC clones. Consistent with our previous results, the RAPD probe hybridized to a large proportion of the BAC clones on the array. A distinct 644 bp segment from the LTR1 sequence hybridized to only seven clones that were all contained in *MAT-1* and this hybridization pattern accounted for the 159 repeat units identified by sequence analysis (Figs. 3A and B; Table 3). Subsequent analysis revealed considerable sequence variability between different examples of LTR1 indicating that it may be possible to identify subclasses for the element. Similar to the result with the

Table 3
Overview of repeats in the *U. hordei MAT-1* region

LTR	Copy number	Retrotransposon (associated with LTR)	Related sequences in *U. maydis*[*]
LTR1	159	Tuh1/Tuh2 (Ty3/Gypsy)	~5
LTR2	5	Tuh4 (Ty1/copia)	—
LTR3	9		—
LTR4	30	Tuh1	—
LTR5	9	Tuh3 (Ty1/copia)	—
LTR6	7		—
LTR7	33	Tuh2 (Ty3/Gypsy)	—
LTR8	12		—
LTR9	16	Tuh2	—
LTR10	3		~3
LTR12	2		—
LTR13	4	Tuh5 (Ty1/copia)	—

Retrotransposon	Type	Copy number intact	Partial	Related sequences in *U. maydis*
Tuh1	(Ty3/Gypsy)	3	10	~5
Tuh2	(Ty3/Gypsy)	0	6	~3
Tuh3	(Ty1/copia)	5	0	~26
Tuh4	(Ty1/copia)	3	2	~7
Tuh5	(Ty1/copia)	1	7	~10
Transposon				
Tho1		2	10	~2
Tho2		1	5	~3
Tho3		1	2	—
Tho4		1	7	—

Note that the element annotated as LTR1 shows variability in terms of the component repetitive sub-sequences. [*]Based on BLASTn (relaxed parameters).

RAPD probe, a pattern of wide-spread occurrence was observed with a probe covering the Gypsy-type gag/pol polyprotein from the Tuh1 element, and with the LTR13 probe (not shown). The LTR3 probe revealed an intermediate occurrence in that hybridization was detected to 249 BAC clones on 24 out of the 38 contigs, including 32 BACs from the map contig carrying the *MAT-1* locus (Tables 1 and 3). By contrast, the Tuh3, LTR2, and LTR9 probes were present at a lower frequency in the genome (Figs. 3C and D) but showed a more frequent occurrence on the *MAT-1* contig. For example, LTR2 hybridized to 101 BAC clones and 44 of these were from the *MAT-1* contig (Table 1, Fig. 3D). We should note that the number of repeats per BAC clone outside of the *MAT-1* locus is unknown and it is possible that certain parts of the genome are not represented in the map, although clones that did not assemble into fingerprint contigs (singletons), or which failed to generate successful fingerprints (orphans), were also represented on the filters (Table 1). Overall, these results suggest that the *U. hordei* genome is rich in repetitive elements and that specific elements have accumulated preferentially at *MAT-1*.

3.4. Comparison between the U. hordei and U. maydis mating-type regions

The genome sequence for *U. maydis* has been determined and annotation is in progress (Broad Institute, Munich Information Centre for Protein Sequences: MIPS) thus allowing comparisons with *U. hordei* in the mating-type regions. The gene and repeat organization between the two regions is particularly striking when the conservation of synteny for genes in *MAT-1* is compared with the regions containing the *a* and *b* loci in *U. maydis* (Fig. 2). Annotation of the genes in the *MAT-1* region revealed that in the majority of cases the closest ortholog was a *U. maydis* gene (Table 2, Supplementary Fig. 6). In addition, these orthologs mapped to only three contigs in the *U. maydis* genome; two of these (1.12 and 1.13) appear to be linked on chromosome I where contig 1.13 harbors the *b* locus, and the other (contig 1.83) carries the *a* locus on chromosome V (Fig. 2). Direct comparisons of the organization of genes in *MAT-1* and the *a* and *b* loci reinforced the view that the *MAT-1* locus has been subjected to a striking accumulation of repeats, as well as inversions, deletions and translocations. As described above, the *U. hordei* genes occur in clusters of two or three genes and these generally showed conservation of synteny with the corresponding *U. maydis* orthologs. This pattern was evident from the progression of numbers assigned to the *U. maydis* genes during annotation (Table 2 and Supplementary Table 4) and is illustrated by the color coded ORFs in Fig. 2. Of the 47 *U. hordei* genes identified in the *MAT-1* sequence, only four were not found in the corresponding sequence contigs in *U. maydis*; two of these encoded hypothetical proteins with highest similarity to proteins in other fungi (although one did have similarity to a *U. maydis* sequence) and two encoded putative tyrosine recombinases (Supplementary Fig. 6). We expanded our search for potential conservation of synteny among

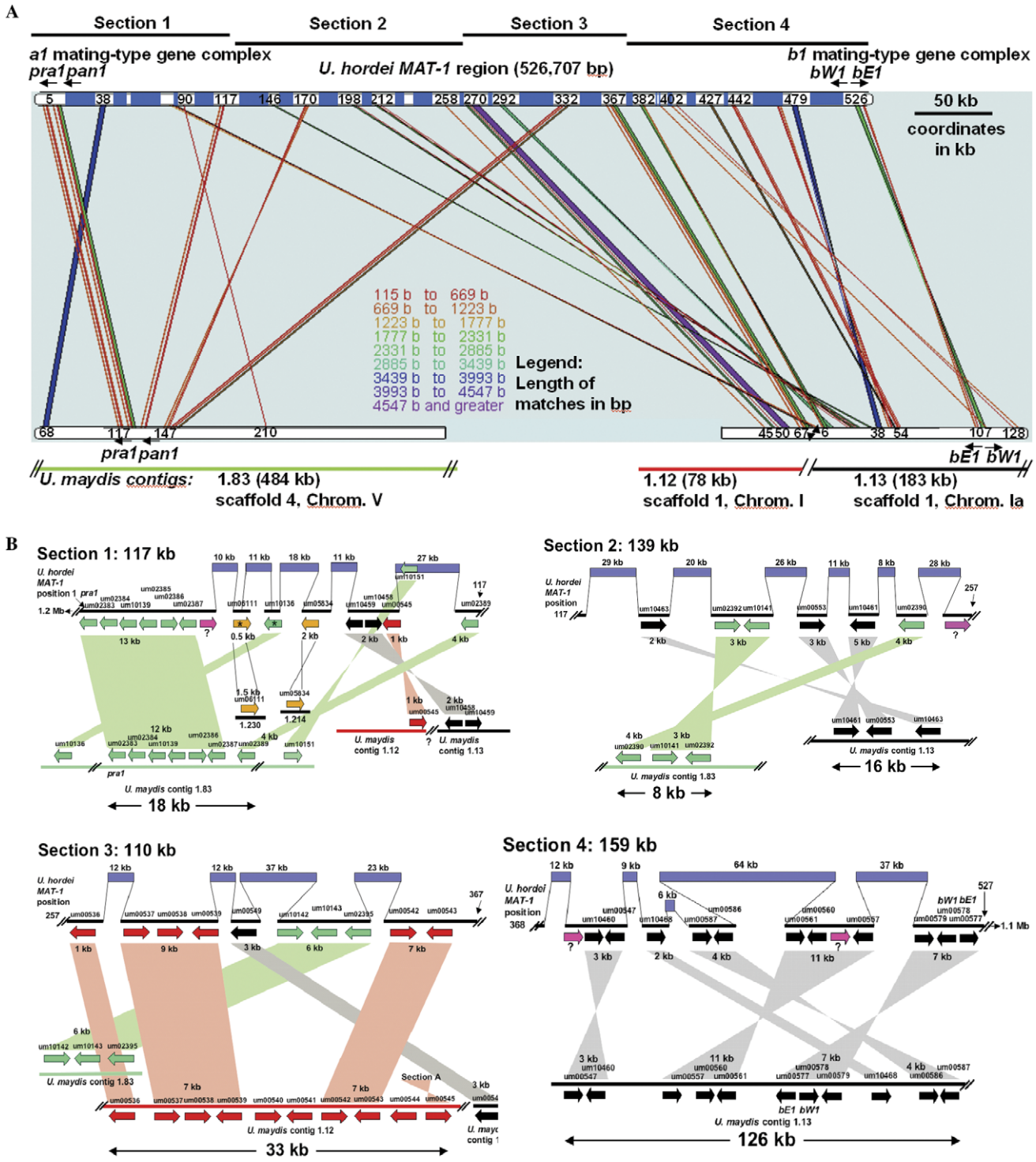

Fig. 2. Comparison of the regions containing the mating-type gene complexes from *U. hordei* and *U. maydis*. (A) PatternHunter output (Ma et al., 2002) to provide an overview of the regions being compared between the two species and to illustrate the gross rearrangements present in these areas. *Pra1*: *a1* pheromone receptor gene; *pan1*: pantoate *b*-alanine ligase gene; *bW1*: bWest1 gene; *bE1*: bEast1 gene (Bakkeren and Kronstad, 1993, 1996; Feldbrugge et al., 2004). Numbers refer to coordinates in kb. (B) Detailed view of four arbitrarily chosen sections labelled 1–4 in (A) of the 527-kb *MAT-1* region aligned with the three *U. maydis* contigs to illustrate the conservation of synteny in local regions, as well as the rearrangements and the accumulation of repetitive sequences in *U. hordei MAT-1*. A detailed list of the genes and their sizes and coordinates is provided in Table 2 and Supplementary Table 4, and EMBL accession AM118080. Note that the genes are not drawn to scale; indicated are approximate sizes of the blocks of genes to display their arrangements and orientations relative to the repeats found in the *MAT-1* locus.

three other more distant basidiomycete species whose complete genomes have been released recently. Targeted similarity searches with the *U. hordei MAT-1* proteins revealed homologs with varying degrees of confidence (Supplementary Fig. 6) but no obvious conservation of synteny was apparent.

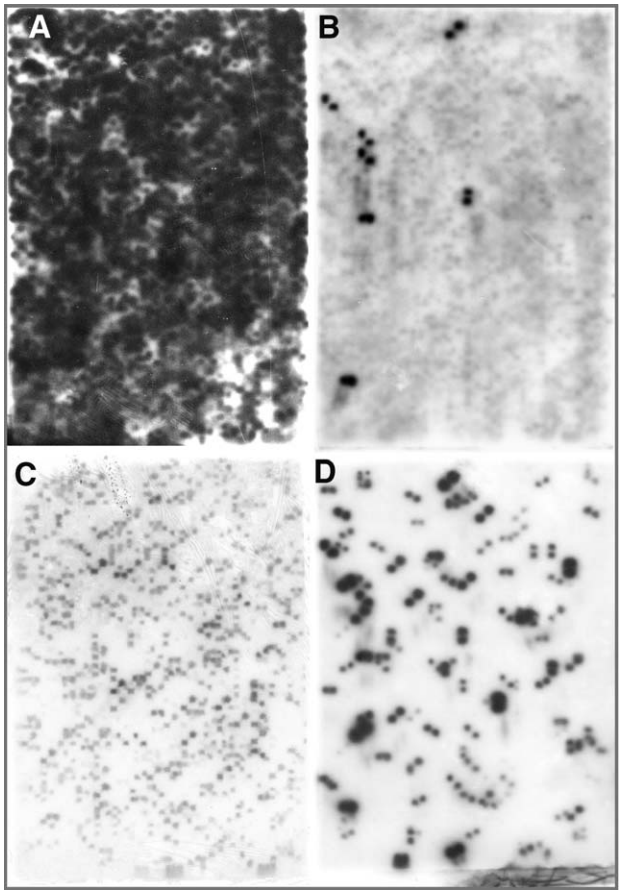

Fig. 3. Hybridization to BAC clone arrays to estimate the frequency of occurrence of LTRs in the genome of *U. hordei*. BAC clone filter arrays carrying the 2304 clones in the library were hybridized with probes as indicated: (A) 1.5-kb RAPD marker 359 (LTR1-like sequence; Linning et al., 2004); (B) 644-bp fragment of LTR1; (C) 4 kb sequence for Tuh3; (D) 261-bp fragment of LTR2. Each BAC clone is spotted twice on the filter. A single positive clone resulting from the probe hybridization is thus indicated by twin spots.

Searches were performed to characterize the distribution of sequences related to the *U. hordei* repeats in the assembled *U. maydis* genome as well as in all of the available sequence reads from the sequencing projects with this species. Targeted searches were also performed against the *U. maydis* contigs carrying the *a* and *b* loci. Short sequences with weak similarity to LTRs 1 and 10, and to the putative transposon Tho1, were found in the *U. maydis* genome (Table 3 and Supplementary Table 4). Additionally, three sequences related to Tho2 were found, but only one of these showed extensive similarity to the putative transposon (65% identity over 1385 bp). We found sequences related to all five of the *U. hordei* retrotransposons in the *U. maydis* genome with sequences related to the gag-pol region of Tuh3 being the most common (Supplementary Table 4). The best match for all of these sequences was an 1189 bp segment of Tuh3 that shared 65% identity with a sequence on contig 1.86 in the *U. maydis* genome (on scaffold 5). Importantly, none of these sequences were found on the contigs 1.12 (78 kb), 1.13 (183 kb), and 1.83 (484 kb) that carry the mating type sequences, although some are present on the same scaffold. We conclude that repetitive sequences that are closely related to the elements at *MAT-1* are not recognizable in the *U. maydis* genome. This suggests these elements likely diverged in sequence between the two species and that *U. hordei* may have acquired some of these elements after divergence from a common ancestor.

We also performed a Smith–Waterman alignment using CrossMatch (Green, 1994) to examine the density of repeats associated with the *U. hordei MAT-1* locus versus the *U. maydis* mating-type regions. As shown in Fig. 4, comparison of the sequence of the *MAT-1* locus with itself identified the same distribution of the highly repetitive sequences that was characterized in detail by annotation. For example, the most abundant repeat class found by CrossMatch (>8× and color coded yellow) corresponded to the LTR1 element. The gene islands (Fig. 2) are also evident from this analysis. In stark contrast, very few repetitive sequences of limited copy number (~2×) were found on the *U. maydis* contigs 1.12 and 1.13 that contain the region with the *b1* mating-type locus. Contig 1.83 that carries the *a1* mating gene complex in *U. maydis* possessed a higher frequency of repetitive sequences, some of which were present in multiple copies (>8×), but this region was not nearly as repetitive as the *MAT-1* locus. In addition, closer inspection of the repeated sequences found on the *U. maydis* contigs revealed that they were only 25–100 bp in length. Overall, this analysis highlights the dramatic accumulation of repeated elements at the *MAT-1* locus and reveals that the comparable regions from *U. maydis* do not have nearly the same density of repetitive sequences.

4. Discussion

4.1. The U. hordei MAT-1 locus is unusual compared with other fungal mating-type loci

At 526 kb, the *MAT-1* locus of *U. hordei* is substantially larger than other mating-type loci thus far characterized in fungi. In general, mating-type loci in ascomycetes are relatively short regions (<10 kb) containing genes for transcription factors (e.g., HMG or homeodomain proteins) that regulate the expression of unlinked mating functions. In basidiomycetes, mating-type loci have generally been defined as regions encoding either homeodomain transcription factors or pheromones and pheromone receptors that are essential during mating. In our work, we have defined the *U. hordei MAT* locus as the region bordered by the two known *a* and *b* mating-type gene complexes (Bakkeren and Kronstad, 1993, 1994). The largest mating-type regions characterized to date in basidiomycetes are the *MAT*a and *MAT*α loci of *Cryptococcus neoformans* and *Cryptococcus gattii* (Lengeler et al., 2002; Fraser and Heitman, 2005). These 105- to 130-kb regions contain approximately 20 genes including those encoding pheromones, pheromone receptors and homeodomain proteins, as well as components of the pheromone-response signalling pathway (e.g., Ste11). Given the large size of the *Cryptococcus* locus and

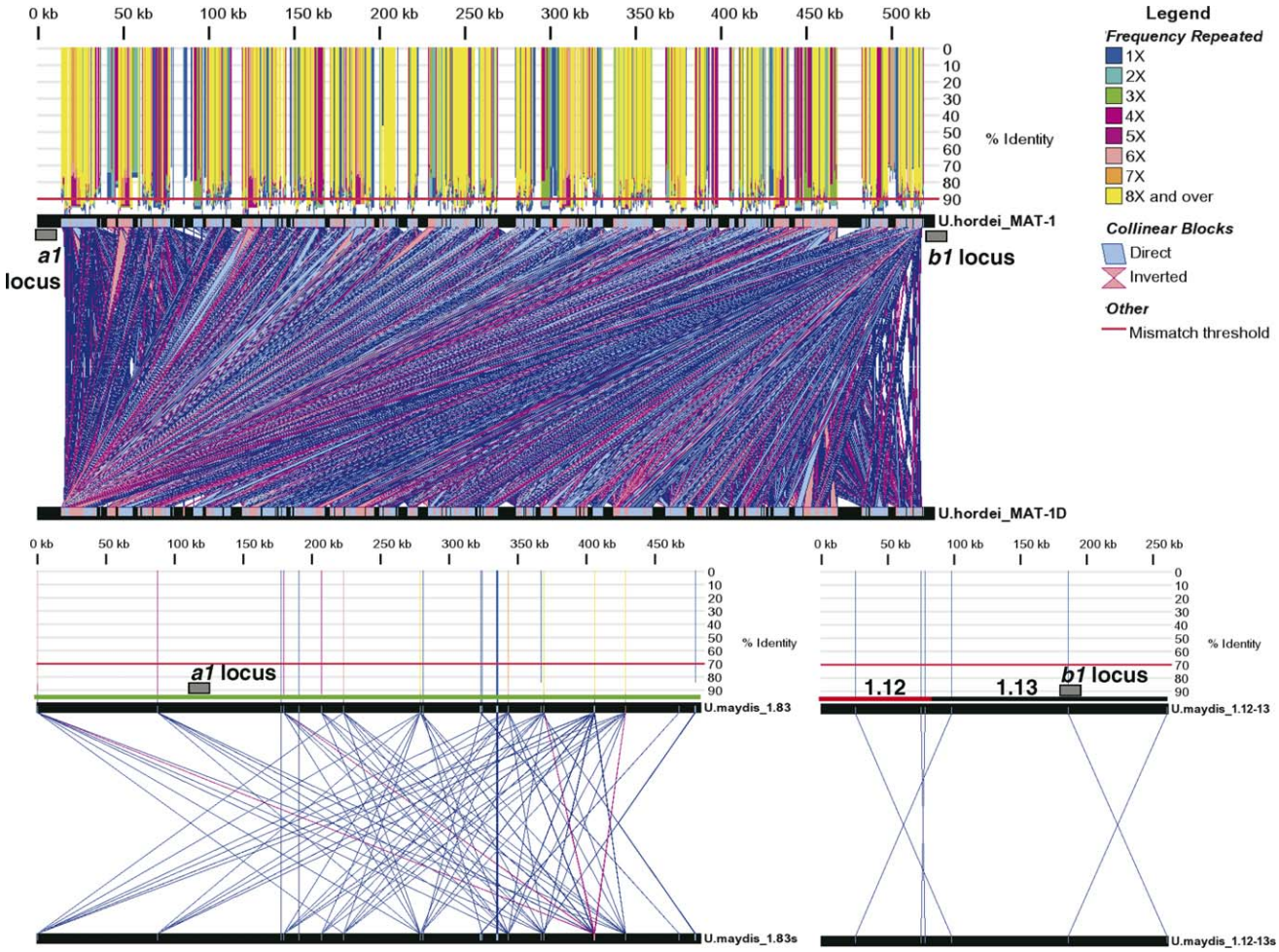

Fig. 4. CrossMatch comparison of the repeat density at the *U. hordei MAT-1* locus versus the densities for the three sequence contigs carrying the *a* and *b* mating-type gene complexes of *U. maydis*. The chromosomal regions harboring the mating-type loci are represented by the black rectangles; *U. maydis* contigs 1.12 and 1.13 are linked as shown in Fig. 2 and the locations of the mating-type gene complexes are indicated. Direct and inverted repeats are depicted by blue- and salmon-colored boxes on the DNA sequence, respectively. Only co-linear blocks having less than 10% base mismatch are displayed, along with their relationship. A sliding window of 50 nucleotides was used to determine the frequency of repeated bases between or within regions. These frequencies are indicated in the legend and represented by the colored lines above the top DNA sequence. The length of each line represents the percent sequence identity shared by the two sequences inside that window.

the bipolar mating system in this fungus, we initially postulated that the *MAT-1* locus in *U. hordei* also might contain additional mating-related genes (Lee et al., 1999). We reasoned that the lack of recombination at this locus might have led to the accumulation of other genes involved in mating and pathogenesis because the *a* and *b* gene complexes control the formation of the infectious cell type for *U. hordei*. However, our analysis revealed only one putative protein kinase gene with most of the other predicted genes encoding functionally uncharacterized proteins or metabolic functions not obviously related to mating. It is possible that other genes with functions involved in mating lie proximal to these regions and that recombination is suppressed beyond the sequenced region, but the analysis of these regions will await future study.

We did find the *pan1* gene (encoding a pantothenic acid biosynthetic enzyme) at the *MAT-1* locus in *U. hordei* and at the mating-type loci in *U. maydis* (gene um10139) and *C. neoformans*. The functional significance of this association, if any, is not apparent. Similarly, the *opt1* gene, which encodes an oligopeptide transporter, was found near the *a* gene complexes in both *U. hordei* and *U. maydis* (gene um02387). It is not known whether this gene is involved in mating in the *Ustilago* species, but a gene encoding a putative oligopeptide transporter (*mtd1*) is regulated by the *B* mating-type locus in *Schizophyllum commune* (Lengeler and Kothe, 1999). Genes that are not clearly related to mating or sexual development have been found at the *MAT* loci of *Candida albicans* and *Cryptococcus* species (Hull and Johnson, 1999; Fraser et al., 2004). In addition to *pan1* and *opt1*, sixteen additional *U. hordei MAT-1* proteins were conserved with a reasonable degree of similarity (<e−20) in all five basidiomycete genomes that we examined (Supplementary Fig. 6).

4.2. Repetitive sequences in U. hordei

The *U. hordei MAT-1* locus contains a remarkable number of repetitive elements and some of these sequences have

a wide genomic distribution (Figs. 2 and 3, Table 1). Repetitive sequences and retroelements also account for up to 15% of the *MAT* locus of *C. neoformans* (Fraser et al., 2004) and are present at high frequency on the mating-type chromosomes of the bipolar anther smut fungus *Microbotryum violaceum* (Hood, 2005). The *MAT-1* locus of *U. hordei* appears to be an extreme case because ~50% of the region is composed of repetitive DNA with alternating repetitive regions and small islands of genes, an organization reminiscent of that found in cereal genomes. Analysis of the sequences of BAC clones from various grasses reveal gene islands in between clusters of repeats and retroelements (Ramakrishna et al., 2002; Anderson et al., 2003). These similarities could simply reflect the repeat-rich nature of the cereal and *U. hordei* genomes, although the possibility of parallel genome evolution for a pathogen and its host is intriguing. We found that the *U. maydis* genome does not harbor the abundance of retroelements found in *U. hordei* and that the mating-type regions are much less repetitive. However, some sequences similar to the *U. hordei* Tuh retrotransposons were found in *U. maydis*. For example, an element related to Tuh3 is the most frequently represented with 26 similar sequences in *U. maydis*; this element resembles the copia-like element HobS that has been suggested to be associated with centromeres in *U. maydis* (Kahmann and Kamper, 2004). It is possible that a centromere is present at the *MAT-1* locus given the lack of recombination and the high density of repeated sequences in the region. Overall, it will be interesting to explore whether the lack of repeat accumulation in the genome is a general feature of species with tetrapolar mating systems.

4.3. Evolution of bipolar and tetrapolar mating systems in the genus Ustilago

In general, size polymorphisms, suppression of recombination and repeat accumulation are well-documented features of sex chromosomes, particularly in mammalian systems (e.g., the Y chromosome in humans; Lahn and Page, 1999; Fraser and Heitman, 2005). Previously we showed that the *MAT* locus of *U. hordei* has features in common with sex chromosomes, e.g., physical linkage of the sex-determining gene complexes, loss of recombination, and a size difference between *MAT-1* and *MAT-2* (Bakkeren and Kronstad, 1994; Lee et al., 1999). We show here that the *MAT-1* locus also shares the feature of repeat accumulation with sex chromosomes. The features for *MAT* in *U. hordei* are quite similar to the sex chromosomes in the bipolar smut *M. violaceum*. Hood (2002; 2005) has shown that the chromosomes carrying the mating-type locus in this fungus are dimorphic and rich in repetitive sequences. The shared features raise the possibility that a bipolar mating system might contribute to the accumulation of repetitive elements at the mating-type locus and throughout the genome in these fungi. Similarities with the bipolar *MAT* locus in *Cryptococcus* species support this idea. However, our comparison with *U. maydis* revealed that the tetrapolar mating system in this species departs from the paradigm of repeat accumulation suggesting that sex chromosomes may evolve differently in the context of this type of fungal mating system. Comparisons of sex determining sequences in other genera that have both tetrapolar and bipolar mating systems are needed to examine whether tetrapolar systems generally accumulate fewer repetitive sequences.

One factor that may contribute to the accumulation of repetitive elements in a bipolar mating system, compared with a tetrapolar system, is the potential for inbreeding versus outbreeding in the two systems. The tetrapolar mating system is generally thought to promote outbreeding and, in the case of the *Ustilago* species, this system may be less favorable for the accumulation of repetitive elements. Inbreeding in the case of bipolar *Ustilago* species may be particularly common given that teliospore germination occurs with infection of the host (concomitantly with seed germination) and there may be limited opportunities for non-sibling interactions among meiotic progeny. Arkhipova and Meselson (2005) hypothesized that, relative to asexual organisms, sexual activity may limit the proliferation of transposable elements within a genome even though new elements may be introduced through sex. In this context, inbreeding may have a similar influence as asexuality with regard to the accumulation of transposable elements for *U. hordei*. The bipolar and tetrapolar mating systems in the *Ustilago* species also differ in the number of different mating specificities encoded by the alleles of the *b* genes. Multiple allelic specificities are known for the *b* locus (at least 25) in *U. maydis* and these presumably evolved since the time of divergence from an ancestor shared with *U. hordei* (two *b* specificities). For *U. hordei*, the association of the *a* and *b* gene complexes in a locus with suppressed recombination may also have presented limitations to the development of multiple specificities. This idea is consistent with recent findings for *Sporisorium reilianum*, a tetrapolar relative of the *Ustilago* species that has three specificities at the *a* locus and multiple specificities at the *b* locus (Schirawski et al., 2005). In the future, it will be interesting to examine the mating systems in the smuts in parallel with the emerging view of the phylogeny of these fungi (Stoll et al., 2005).

Fraser et al. (2004) presented phylogenetic arguments based on the comparative sequence analysis of the *Cryptococcus MAT* locus that support the derivation of a bipolar mating system from a tetrapolar system. These authors proposed the fusion of ancient loci encoding a homeodomain transcription factor and a pheromone/receptor function with the accompanying trapping of additional genes. Our data for two *Ustilago* species supports the scenario of evolution in the tetrapolar to bipolar direction because of the dramatic accumulation of repetitive elements in the *MAT-1* region compared with the regions surrounding *a* and *b* in *U. maydis*. That is, blocks of genes show conservation of synteny when the regions are compared (Fig. 2), but gene order has apparently been interrupted or rearranged by repeat sequences in the *U. hordei* regions since the time of divergence. Such overall synteny with local rearrangements

has been observed by Galagan et al. (2005) when comparing genomes of three *Aspergillus* species around the *MAT* loci (although for these ascomycete fungi, evolutionary trends resulted in a discrimination between homo- and heterothalism). Concerning evolutionary events in the smut fungi, one view is that contigs 1.12 and 1.13 are contiguous in *U. maydis* and the syntenous chromosome section in an ancestor of *U. hordei* was initially translocated to a chromosome arm syntenous to *U. maydis* contig 1.83, possibly facilitated by repetitive elements or a centromere.

Genome rearrangements, likely mediated by repeats and including inversions (demonstrated around the *a* and *b* loci; Lee et al., 1999), probably contribute to the suppression of recombination in the region. It has been postulated that the lack of "purifying recombination" in sex-determining regions leads to the accumulation of transposable elements and repeats (Charlesworth and Langley, 1989) offering an explanation for the increased abundance of such elements compared to the rest of the genome. Finally, a small number of genes with similarity to *U. maydis* genes not associated with mating type seem to have been trapped at the *MAT* locus in *U. hordei*. This suggests that rearrangements have occurred at *MAT-1* that involved other regions of the *U. hordei* genome.

Acknowledgments

We thank the mapping, sequencing and systems personnel of the Michael Smith Genome Sciences Centre (J. Asano, I. Bosdet, S. Chan, S. Chittaranjan, C. Fjell, N. Girn, C. Gray, R. Guin, M. Krzywinski, R. Kutsche, S. Leach, D. Lee, S. Lee, B. Li, C. Mathewson, C. McLeavy, S. Ness, T. Olson, P. Pandoh, A. Prabhu, P. Saeedi, D. Smailus, L. Spence, J. Stott, S. Taylor, M. Tsai, N. Wye, and G. Yang) for their contributions to this work. The authors thank Joe Heitman and James Fraser for helpful discussions, Jessica Sawkins for help with annotation and Limei Yan, Sunkyoung So, and Sulan Qi for help with sequencing. We also acknowledge the Broad Institute and the Munich Information Center for Protein Sequences for the *U. maydis* genome sequence. This work was supported by the NHGRI Mouse Genome Sequencing Network, by Discovery and Genomics grants from NSERC (Canada), and by a scholar award from the Burroughs Wellcome Fund to J. K., M. M., and S.J. are Michael Smith Foundation for Health Research Biomedical Scholars and M.M., is a Terry Fox/NCIC Young Investigator.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fgb. 2006.04.002.

References

Abdennadher, M., Mills, D., 2000. Telomere-associated RFLPs and electrophoretic karyotyping reveal lineage relationships among race-specific strains of *Ustilago hordei*. Curr. Genet. 38, 141–147.

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

Anderson, C.M., Willits, D.A., Kosted, P.J., Ford, E.J., Martinez-Espinoza, A.D., Sherwood, J.E., 1999. Molecular analysis of the pheromone and pheromone receptor genes of *Ustilago hordei*. Gene 240, 89–97.

Anderson, O.D., Rausch, C., Moullet, O., Lagudah, E.S., 2003. The wheat D-genome HMW-glutenin locus, BAC sequencing, gene distribution, and retrotransposon clusters. Funct. Integr. Genomics 3, 56–68.

Arkhipova, I., Meselson, M., 2005. Deleterious transposable elements and the extinction of asexuals. Bioessays 27, 76–85.

Bakkeren, G., Gibbard, B., Yee, A., Froeliger, E., Leong, S., Kronstad, J., 1992. The *a* and *b* loci of *Ustilago maydis* hybridize with DNA sequences from other smut fungi. Mol. Plant Microbe Interact. 5, 347–355.

Bakkeren, G., Kronstad, J.W., 1993. Conservation of the *b* mating-type gene complex among bipolar and tetrapolar smut fungi. Plant Cell 5, 123–136.

Bakkeren, G., Kronstad, J.W., 1994. Linkage of mating-type loci distinguishes bipolar from tetrapolar mating in basidiomycetous smut fungi. Proc. Natl. Acad. Sci. USA 91, 7085–7089.

Bakkeren, G., Kronstad, J.W., 1996. The pheromone cell signaling components of the *Ustilago a* mating-type loci determine intercompatibility between species. Genetics 143, 1601–1613.

Bolker, M., Urban, M., Kahmann, R., 1992. The a mating type locus of *Ustilago* maydis specifies cell signaling components. Cell 68, 441–450.

Borkovich, K.A., Alex, L.A., Yarden, O., Freitag, M., Turner, G.E., Read, N.D., Seiler, S., Bell-Pedersen, D., Paietta, J., Plesofsky, N., et al., 2004. Lessons from the genome sequence of *Neurospora crassa*, tracing the path from genomic blueprint to multicellular organism. Microbiol. Mol. Biol. Rev. 68, 1–108.

Charlesworth, B., Langley, C.H., 1989. The population genetics of *Drosophila* transposable elements. Annu. Rev. Genet. 23, 251–287.

Erlandsson, R., Wilson, J.F., Paabo, S., 2000. Sex chromosomal transposable element accumulation and male-driven substitutional evolution in humans. Mol. Biol. Evol. 17, 804–812.

Feldbrugge, M., Kamper, J., Steinberg, G., Kahmann, R., 2004. Regulation of mating and pathogenic development in *Ustilago maydis*. Curr. Opin. Microbiol. 7, 666–672.

Fisher, G.W., Holton, C.S., 1957. Biology and Control of the Smut Fungi. Ronald Press, New York.

Fraser, J.A., Diezmann, S., Subaran, R.L., Allen, A., Lengeler, K.B., Dietrich, F.S., Heitman, J., 2004. Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. PLoS Biol. 2, e384.

Fraser, J.A., Heitman, J., 2005. Chromosomal sex-determining regions in animals, plants and fungi. Curr. Opin. Genet. Dev. 15, 645–651.

Fuhrmann, D.R., Krzywinski, M.I., Chiu, R., Saeedi, P., Schein, J.E., Bosdet, I.E., Chinwalla, A., Hillier, L.W., Waterston, R.H., McPherson, J.D., et al., 2003. Software for automated analysis of DNA fingerprinting gels. Genome Res. 13, 940–953.

Galagan, J.E., Calvo, S.E., Cuomo, C., Ma, L.J., Wortman, J.R., Batzoglou, S., Lee, S.I., Basturkmen, M., Spevak, C.C., Clutterbuck, J., et al., 2005. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. Nature 438, 1105–1115.

Gaudet, D.A., Kiesling, R.L., 1991. Variation in aggressiveness among and within races of *Ustilago hordei* on barley. Phytopathology 81, 1385–1390.

Gaudet, D.A., Gusse, J., Laroche, A., 1998. Origins and inheritance of chromosome-length polymorphisms in the barley covered smut fungus, *Ustilago hordei*. Curr. Genet. 33, 216–224.

Green, P., 1994. Ancient conserved regions in gene sequences. Curr. Opin. Struct. Biol. 4, 404–412.

Hood, M.E., 2002. Dimorphic mating-type chromosomes in the fungus *Microbotryum violaceum*. Genetics 160, 457–461.

Hood, M.E., 2005. Repetitive DNA in the automictic fungus *Microbotryum violaceum*. Genetica 124, 1–10.

Hu, G.G., Linning, R., Bakkeren, G., 2002. Sporidial mating and infection process of the smut fungus, *Ustilago hordei*, in susceptible barley. Can. J. Bot. 80, 1103–1114.

Hull, C.M., Johnson, A.D., 1999. Identification of a mating type-like locus in the asexual pathogenic yeast *Candida albicans*. Science 285, 1271–1275.

Kahmann, R., Kamper, J., 2004. *Ustilago maydis*, how its biology relates to pathogenic development. New Phytol. 164, 31–42.

Kronstad, J.W., Staben, C., 1997. Mating type in filamentous fungi. Annu. Rev. Genet. 31, 245–276.

Lahn, B.T., Page, D.C., 1999. Four evolutionary strata on the human X chromosome. Science 286, 964–967.

Lee, N., Bakkeren, G., Wong, K., Sherwood, J.E., Kronstad, J.W., 1999. The mating-type and pathogenicity locus of the fungus *Ustilago hordei* spans a 500-kb region. Proc. Natl. Acad. Sci. USA 96, 15026–15031.

Lengeler, K.B., Kothe, E., 1999. Mated, a putative peptide transporter of *Schizophyllum commune* expressed in dikaryons. Curr. Genet. 36, 159–164.

Lengeler, K.B., Fox, D.S., Fraser, J.A., Allen, A., Forrester, K., Dietrich, F.S., Heitman, J., 2002. Mating-type locus of *Cryptococcus neoformans*, A step in the evolution of sex chromosomes. Eukaryotic Cell 1, 704–718.

Linning, R., Lin, D., Lee, N., Abdennadher, M., Gaudet, D., Thomas, P., Mills, D., Kronstad, J.W., Bakkeren, G., 2004. Marker-based cloning of the region containing the *UhAvr1* avirulence gene from the basidiomycete barley pathogen *Ustilago hordei*. Genetics 166, 99–111.

Loftus, B.J., Fung, E., Roncaglia, P., Rowley, D., Amedeo, P., Bruno, D., Vamathevan, J., Miranda, M., Anderson, I.J., Fraser, J.A., et al., 2005. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. Science 307, 1321–1324.

Ma, B., Tromp, J., Li, M., 2002. PatternHunter, faster and more sensitive homology search. Bioinformatics 18, 440–445.

Majoros, W.H., Pertea, M., Antonescu, C., Salzberg, S.L., 2003. GlimmerM, Exonomy and Unveil, three ab initio eukaryotic genefinders. Nucleic Acids Res. 31, 3601–3604.

Marra, M.A., Kucaba, T.A., Dietrich, N.L., Green, E.D., Brownstein, B., Wilson, R.K., McDonald, K.M., Hillier, L.W., McPherson, J.D., Waterston, R.H., 1997. High throughput fingerprint analysis of large-insert clones. Genome Res. 7, 1072–1084.

Marra, M., Kucaba, T., Sekhon, M., Hillier, L., Martienssen, R., Chinwalla, A., Crockett, J., Fedele, J., Grover, H., Gund, C., et al., 1999. A map for sequence analysis of the *Arabidopsis thaliana* genome. Nat. Genet. 22, 265–270.

McCluskey, K., Mills, D., 1990. Identification and characterization of chromosome length polymorphisms among strains representing fourteen races of *Ustilago hordei*. Mol. Plant Microbe Interact. 3, 366–373.

McCluskey, K., Agnan, J., Mills, D., 1994. Characterization of genome plasticity in *Ustilago hordei*. Curr. Genet. 26, 486–493.

McPherson, J.D., Marra, M., Hillier, L., Waterston, R.H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E.R., Wilson, R.K., et al., 2001. A physical map of the human genome. Nature 409, 934–941.

Ramakrishna, W., Dubcovsky, J., Park, Y.J., Busso, C., Emberton, J., San-Miguel, P., Bennetzen, J.L., 2002. Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. Genetics 162, 1389–1400.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., Barrell, B., 2000. Artemis, sequence visualization and annotation. Bioinformatics 16, 944–945.

Salamov, A.A., Solovyev, V.V., 2000. Ab initio gene finding in *Drosophila* genomic DNA. Genome Res. 10, 516–522.

Schein, J.E., Tangen, K.L., Chiu, R., Shin, H., Lengeler, K.B., MacDonald, W.K., Bosdet, I., Heitman, J., Jones, S.J., Marra, M.A., et al., 2002. Physical maps for genome analysis of serotype A and D strains of the fungal pathogen *Cryptococcus neoformans*. Genome Res. 12, 1445–1453.

Schirawski, J., Heinze, B., Wagenknecht, M., Kahmann, R., 2005. Mating type loci of *Sporisorium reilianum*, novel pattern with three *a* and multiple *b* specificities. Euk. Cell 4, 1317–1327.

Stoll, M., Begerow, D., Oberwinkler, F., 2005. Molecular phylogeny of *Ustilago*, *Sporisorium*, and related taxa based on combined analyses of rDNA sequences. Mycol. Res. 109, 342–356.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. ClustalW, 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.

Urban, M., Kahmann, R., Bolker, M., 1996. The biallelic *a* mating type locus of *Ustilago maydis*, remnants of an additional pheromone gene indicate evolution from a multiallelic ancestor. Mol. Gen. Genet. 250, 414–420.