Current Biology, Volume *18*

Supplemental Data

Microsporidia Evolved

from Ancestral Sexual Fungi

Soo Chan Lee, Nicolas Corradi, Edmond J. Byrnes III, Santiago Torres-Martinez, Fred S. Dietrich, Patrick J. Keeling, and Joseph Heitman

Table S1. Relaxed synteny conserved in *R. oryzae* and *E. cuniculi.*

***R. oryzae* proteins**	***E. cuniculi* proteins**	**blast score**	**e value**	***R. oryzae* proteins**	***E. cuniculi* proteins**	**blast score**	**e value**
RO3G_00003	gi 19068537	124	7.00E-30	RO3G_05865	gi 19168645	40	9.00E-04
RO3G_00004	gi 19069234	42	7.00E-05	RO3G_05866			
RO3G_00005				RO3G_05867			
RO3G_00006	gi 19068534	57	3.00E-09	RO3G_05868	gi 19168651	125	8.00E-30
RO3G_00009	gi 19170836	117	2.00E-28	RO3G_06214	gi 19168723	191	1.00E-49
RO3G_00010	gi 19170838	52	5.00E-09	RO3G_06215			
RO3G_01355	gi 19171027	102	2.00E-23	RO3G_06216			
RO3G_01356	gi 19171031	46	1.00E-06	RO3G_06217			
RO3G_01357				RO3G_06218			
RO3G_01358				RO3G_06219			
RO3G_01359				RO3G_06220			
RO3G_01360	gi 19171024	250	1.00E-67	RO3G_06221	gi 19168720	124	4.00E-30
RO3G_02423	gi 19168763	82	1.00E-17	RO3G_07828	gi 19068701	70	1.00E-13
RO3G_02424				RO3G_07829			
RO3G_02425	gi 19069007	59	3.00E-10	RO3G_07830			
RO3G_02426				RO3G_07831			
RO3G_02427				RO3G_07832			
RO3G_02428	gi 19168761	65	6.00E-12	RO3G_07833			
RO3G_03308	gi 19068592	170	4.00E-44	RO3G_07834			
RO3G_03309	gi 19068589	54	9.00E-09	RO3G_07835			
RO3G_03530	gi 19068611	236	2.00E-63	RO3G_07836			
RO3G_03531				RO3G_07837			
RO3G_03532	gi 19068606	79	3.00E-16	RO3G_07838			
RO3G_03533				RO3G_07839	gi 19069299	53	2.00E-09
RO3G_03534	gi 19170972	75	3.00E-15	RO3G_07840			
RO3G_03535	gi 19068608	58	1.00E-10	RO3G_07841			
RO3G_04457	gi 19168720	119	9.00E-29	RO3G_07842	gi 19068700	45	2.00E-06
RO3G_04458				RO3G_08158	gi 19170838	109	6.00E-26
RO3G_04459	gi 19168723	104	5.00E-24	RO3G_08159	gi 19170836	119	6.00E-29
RO3G_04463	gi 19068607	65	6.00E-13	RO3G_08282	gi 19170836	117	2.00E-28
RO3G_04464				RO3G_08283	gi 19170838	109	6.00E-26
RO3G_04465				RO3G_08889	gi 19168720	61	3.00E-11
RO3G_04466				RO3G_08890			
RO3G_04467				RO3G_08891			
RO3G_04468				RO3G_08892	gi 19168723	97	6.00E-22

RO3G_04469				RO3G_09031	gi 19069361	53	6.00E-08
RO3G_04470	gi 19171321	68	5.00E-13	RO3G_09032			
RO3G_04471				RO3G_09033			
RO3G_04472	gi 19068605	77	2.00E-15	RO3G_09034	gi 19068933	62	2.00E-11
RO3G_05485	gi 19068537	133	1.00E-32	RO3G_09035	gi 19069362	39	3.00E-04
RO3G_05486	gi 19171356	82	1.00E-17	RO3G_10613	gi 19170970	185	4.00E-48
RO3G_05487	gi 19068534	307	1.00E-84	RO3G_10614			
RO3G_05592	gi 19068534	52	2.00E-08	RO3G_10615			
RO3G_05593	gi 19068537	107	9.00E-24	RO3G_10616	gi 19170971	56	9.00E-10
RO3G_05835	gi 19069147	90	5.00E-20	RO3G_10956	gi 19069314	402	1.00E-113
RO3G_05836	gi 19069144	299	9.00E-83	RO3G_10957	gi 19069313	49	7.00E-07
RO3G_11255	gi 19168644	74	9.00E-15	RO3G_13173	gi 19068611	228	1.00E-60
RO3G_11256				RO3G_13174	gi 19068608	58	1.00E-10
RO3G_11257	gi 19068701	51	7.00E-08	RO3G_13322	gi 19069361	45	7.00E-06
RO3G_11258				RO3G_13323	gi 19170962	134	4.00E-33
RO3G_11259				RO3G_13324			
RO3G_11260				RO3G_13325			
RO3G_11261				RO3G_13326	gi 19069362	56	3.00E-09
RO3G_11262				RO3G_14144	gi 19068528	338	3.00E-94
RO3G_11263				RO3G_14145			
RO3G_11264				RO3G_14146			
RO3G_11265				RO3G_14147	gi 19170857	153	2.00E-38
RO3G_11266				RO3G_14148			
RO3G_11267				RO3G_14149			
RO3G_11268	gi 19168645	39	2.00E-04	RO3G_14150			
RO3G_12238	gi 19069100	105	3.00E-24	RO3G_14151	gi 19068525	39	2.00E-04
RO3G_12239				RO3G_14309	gi 19069097	52	2.00E-08
RO3G_12240				RO3G_14310	gi 19069100	110	2.00E-25
RO3G_12241				RO3G_14965	gi 19068691	37	5.00E-04
RO3G_12242				RO3G_14966			
RO3G_12243				RO3G_14967	gi 19068690	77	1.00E-15
RO3G_12244				RO3G_15055	gi 19170836	117	2.00E-28
RO3G_12245				RO3G_15056	gi 19170838	109	6.00E-26
RO3G_12246				RO3G_15651	gi 19069192	180	7.00E-47
RO3G_12247				RO3G_15652	gi 19069191	58	7.00E-10
RO3G_12248	gi 19068936	87	3.00E-18	RO3G_15944	gi 19069192	177	6.00E-46
RO3G_12249				RO3G_15945	gi 19069191	55	3.00E-09
RO3G_12250				RO3G_16080	gi 19168720	124	3.00E-30
RO3G_12251				RO3G_16081			
RO3G_12252	gi 19069102	67	2.00E-12	RO3G_16082			
RO3G_12682	gi 19168744	144	4.00E-36	RO3G_16083			
RO3G_12683				RO3G_16084			
RO3G_12684	gi 19069355	617	1.00E-178	RO3G_16085			
RO3G_12685				RO3G_16086			
RO3G_12686				RO3G_16087			
RO3G_12687				RO3G_16088	gi 19168723	159	3.00E-40
RO3G_12688							
RO3G_12689	gi 19168742	71	1.00E-13				
RO3G_12888	gi 19069579	420	1.00E-119				

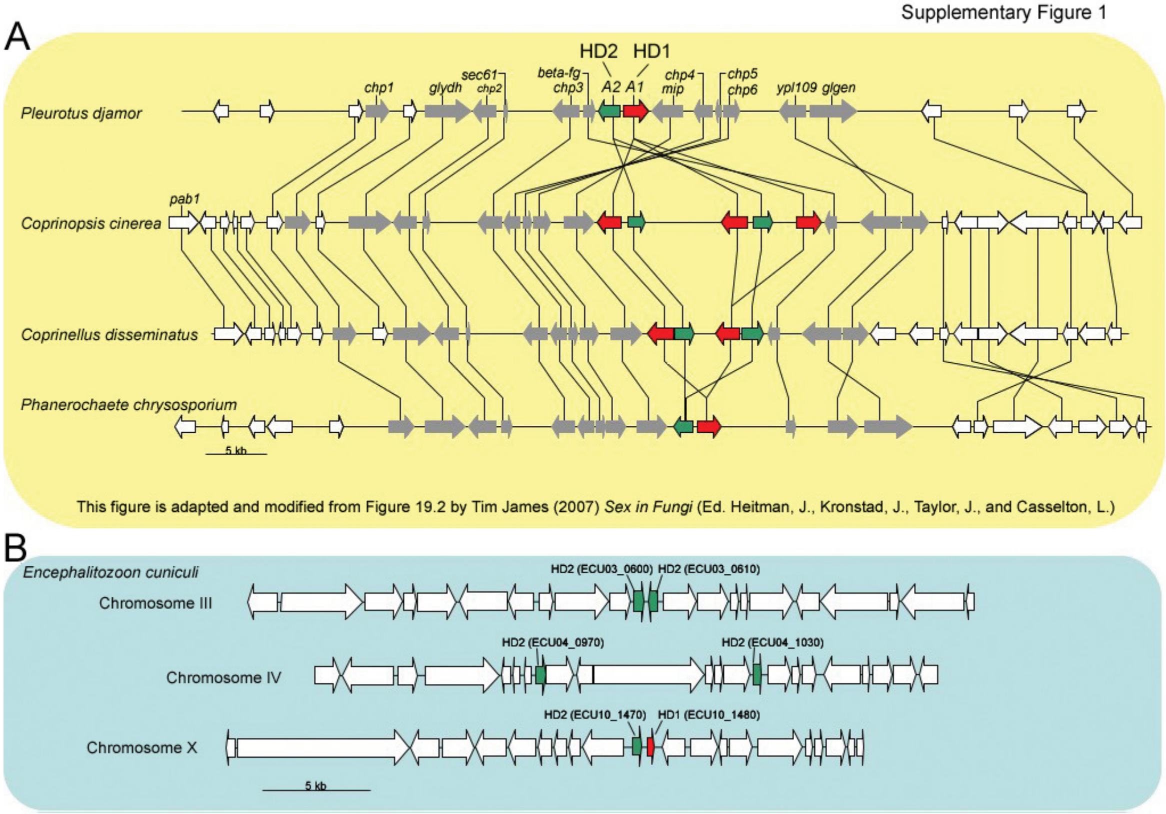
RO3G_12889							
RO3G_12890							
RO3G_12891	gi 19069662	103	5.00E-23				
RO3G_12892							
RO3G_12893							
RO3G_12894							
RO3G_12895	gi 19069580	60	2.00E-10				

Table S2. Microsporidian genes involved in meiosis.

Genes	*E. cuniculi*	accession number	*A. locustae* ID	contig	*E. bieneusi* ID	contig
Spo11	ECU04_1110	CAD25299	ORF1088	273	absent	absent
Mre11	ECU05_1280	CAD26648	newly annotated	604	newly annotated	1951
Rad50	ECU07_0610i	CAD25593	AAT12390	GenBank	ORF8569	30
Rad1	ECU08_0760	CAD26381	ORF1776	1391	absent	absent
Hop1	ECU02_0890	CAD25118	ORF156	366	absent	absent
Hop2	absent	absent	absent	absent	absent	absent
Mnd1	ECU10_1600	CAD25881	ORF2209	513	newly annotated	1117
Rad52/22	ECU09_0930	CAD27066	ORF258	232	N.A.	N.A.
Dmc1	absent	absent	absent	absent	absent	absent
Rad51	ECU11_0820	CAD25992	ORF1016	757	XM_00182796	Genbank
Msh2	ECU03_0540	CAD26200	newly annotated	0	newly annotated	678
Msh6	ECU10_0710	CAD25790	newly annotated	1037	ORF7694	168
Msh4	absent	absent	ORF713	509	newly annotated	678
Msh5	absent	absent	absent	absent	absent	absent
Mlh1	ECU05_0300	CAD26547	newly annotated	601	ORF7777	367
Mlh2	absent	absent	newly annotated	2654	absent	absent
Mlh3	absent	absent	absent	absent	absent	absent
Pms1	ECU11_1260	CAD26036	ORF1393	3	ORF8252	169
Mer3	absent	absent	absent	absent	absent	absent
Smc1	ECU04_0930	CAD25280	ORF303	1358	ORF8566	30
Smc2	ECU04_0930	CAD25280	ORF303	1358	ORF8566	30
Smc3	ECU09_1210	CAD27164	ORF506	289	ORF7826	1012
Smc4	ECU07_0680	CAD25600	ORF2406	169	ORF8576	30
Smc5	ECU11_2000	CAD26110	newly annotated	1454	ORF7827	1012
Rad18	ECU07_0390	CAD25571	newly annotated	3	NW_001849446	GenBank
Rad21	ECU04_1370	CAD25326	newly annotated	400	ORF7660	168
Rec8	ECU07_0390	CAD25571	newly annotated	3	NW_001849446	GenBank
Pds5	absent	absent	absent	absent	absent	absent
Scc3	absent	absent	absent	absent	absent	absent

Grey boxes indicate genes conserved in at least two microsporidia.

Supplementary Figure 1



This figure is adapted and modified from Figure 19.2 by Tim James (2007) *Sex in Fungi* (Ed. Heitman, J., Kronstad, J., Taylor, J., and Casselton, L.)

Figure S1. Comparison of genes surrounding homeodomain (HD) protein genes between four fungi (panel **A**) and *E. cuniculi* (panel **B**). The *E. cuniculi* genome contains three linked pairs of HD genes. 20 ORFs around the three HD pairs were compared to the A mating type locus alleles of four basidiomycetous fungi in which HD proteins [*A1* (HD1) and *A2* (HD2)] orchestrate the sexual cycle. No apparent gene homologs which flank the *MAT* locus of the basidiomycetes were identified around the *E. cuniculi* HD protein genes. Moreover, the HD genes in basidiomycete fungi are divergently transcribed, whereas the HD gene pairs of *E. cuniculi* are co-linear or convergent. Finally, the basidiomycete HD gene pairs always involve one HD1 class (green) and one HD2 class (red) protein, whereas in *E. cuniculi* only one of the three HD gene pairs matches this paradigm. Taken together, these features indicate that these pairs of HD genes are less likely to represent the *sex* locus of *E. cuniculi*.

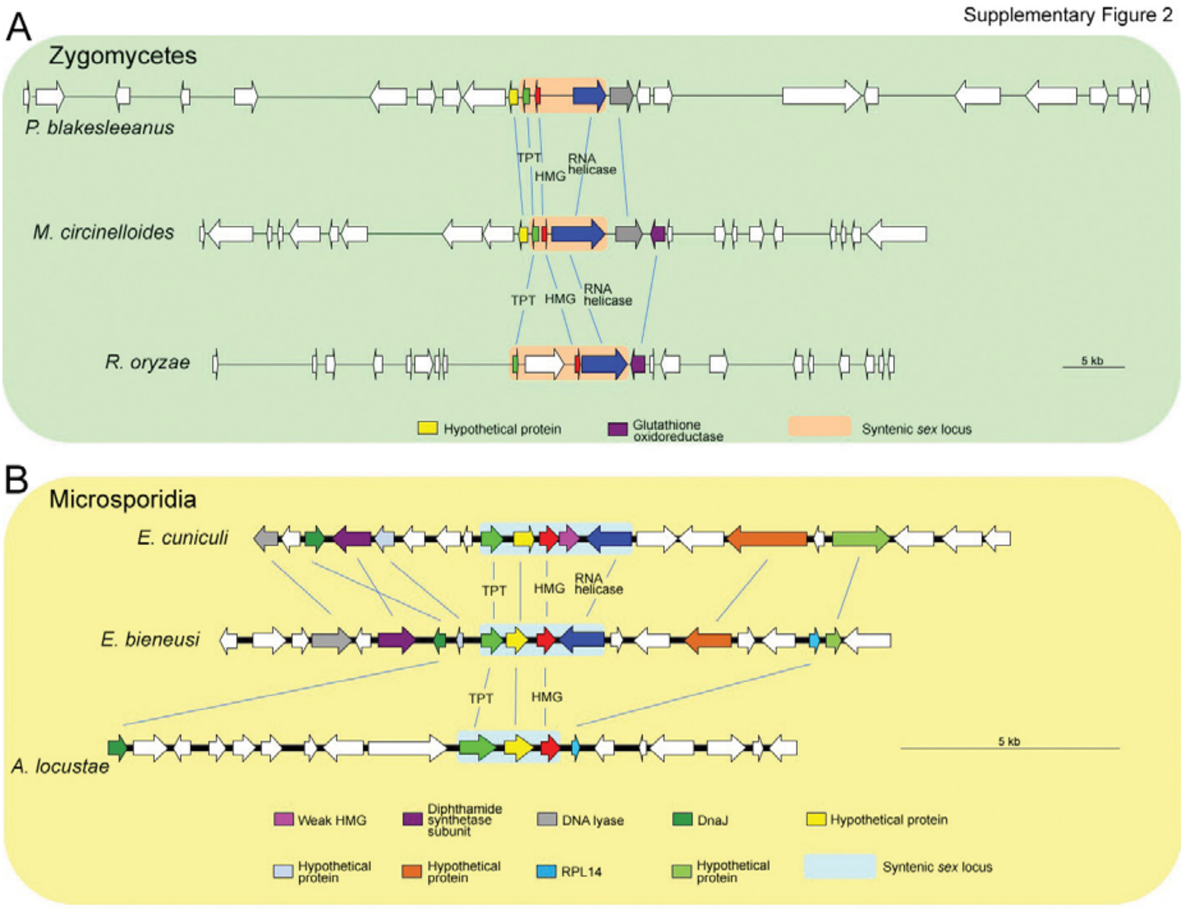


Figure S2. Analysis of ORFs up- and downstream of the *sex* locus of three zygomycetes and three microsporidia. **(A)**. *M. circinelloides* shares two genes encoding hypothetical proteins (5’ and 3’ to the *sex* locus) with *P. blakesleeanus* and shares a glutathione oxidoreductase gene with *R. oryzae*, whereas there are no common genes between *P. blakesleeanus* and *R. oryzae* outside the *sex* locus. This suggests that only the TPT and RNA helicase genes are syntenic at the *sex* locus among the zygomycetes. **(B)**. Additional genes are syntenic around the *sex* locus in the microsporidia. However, only the TPT, HMG, and RNA helicase genes are syntenic across the three zygomycetes and the microsporidia. The ORF information for *R. oryzae* and *E. cuniculi* was obtained from the Broad Institute/Fungal Genome Initiative (FGI). The *P. blakesleeanus* ORF information was obtained from the US Department of Energy Joint Genome Institute (DoE JGI) database. Genomic DNA sequences of *M. circinelloides* genome sequences were obtained from the DoE JGI and annotated by using FGENESH or Orf finder at NCBI. Note: the difference in scale bars for panels **A** and **B**.

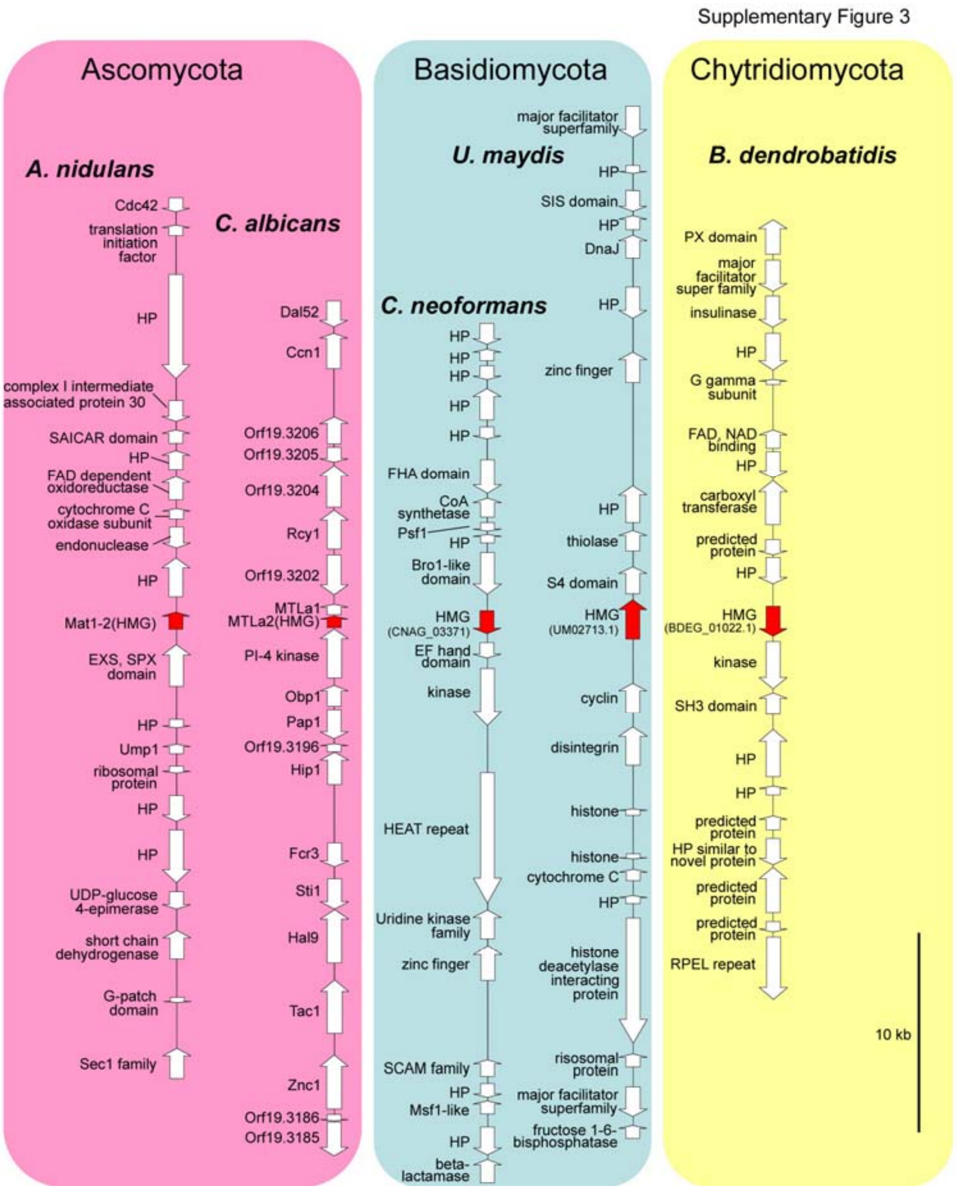


Figure S3. Genomic locations of *MAT*-related HMG domain protein genes of representative fungi in the Ascomycota, Basidiomycota, and Chytridiomycota phyla. HMG proteins were identified by BLASTp analysis with the *M. circinelloides* SexM protein against each fungal protein database. ORFs surrounding the most closely related HMG genes from the five fungi were analyzed. The fungi in the three phyla contain no TPT and RNA helicase genes that are syntenic with the HMG domain protein genes, indicating that syntenic TPT, HMG, and RNA helicase genes are specific to the Zygomycota. The ORF information for *Aspergillus nidulans*, *Cryptococcus neoformans* serotype A, *Ustilago maydis*, and *Batrachochytrium dendrobatidis* was acquired from the Broad Institute/FGI database. ORFs of *Candida albicans* were identified from the Candida Genome Database site. Red arrows indicate HMG proteins. HP: hypothetical protein.

Supplementary Figure 4

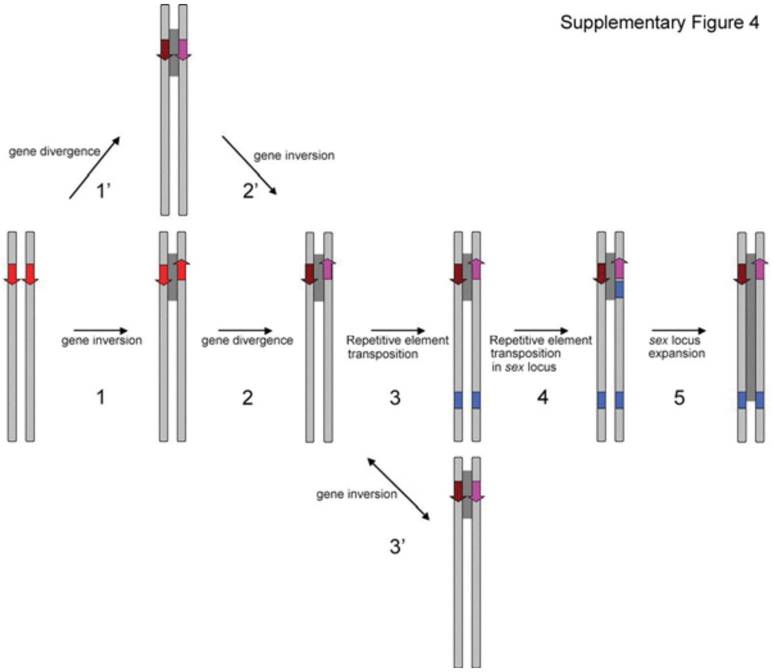


Figure S4. Possible models for the early steps in the evolution of sex determination and sex chromosomes. As posited by Ohno[1], in the first step, a sex determining gene emerges on an autosome. In model I (inversion-divergence model), one allele undergoes a gene inversion (1) resulting in an inhibition of recombination, enabling the inverted gene pair to diverge (2) because of a lack of recombination. A repetitive element (blue box) is integrated into the proto-sex chromosome (3) and transposition of the repetitive element occurs into one *sex* allele (4). Additional inversions due to the repetitive element enable expansion of the *sex* allele to capture the chromosome. The opposite orientation of the *sexP/M* genes of *P. blakesleeanus* supports steps in this model. In model II, the order of the divergence and inversion steps is altered (divergence-inversion model). Unlike *P. blakesleeanus*, the *M. circinelloides sex* loci encode the *sexP* and *sexM* genes in the same orientation. Thus, gene divergence (1') could precede gene inversion (2') suggesting another trajectory for the evolution of sex determining systems. Alternatively, the inverted *sexP/M* genes could have undergone a second gene inversion (3') resulting in the co-directional *sexP/M* gene pair in *M. circinelloides*.

Supplementary Figure 5

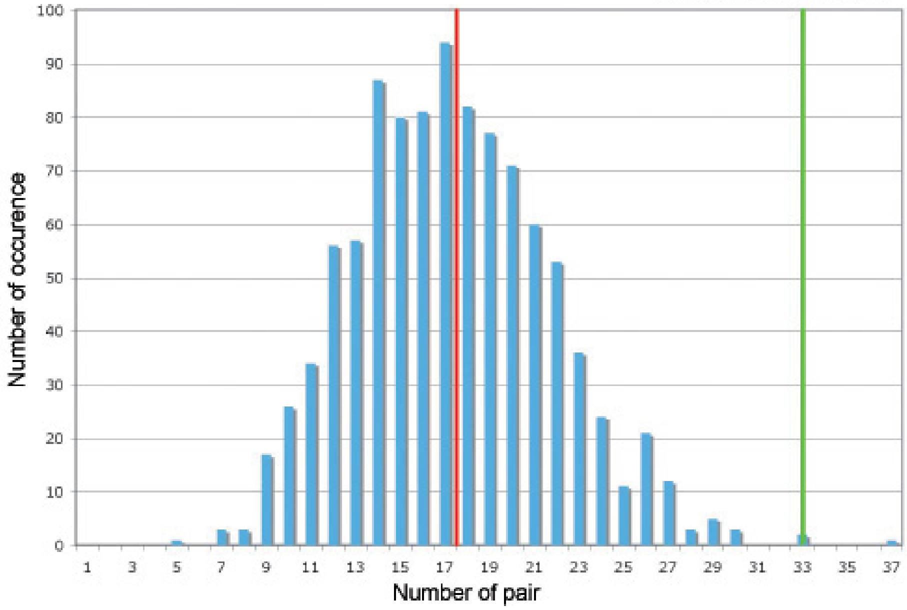
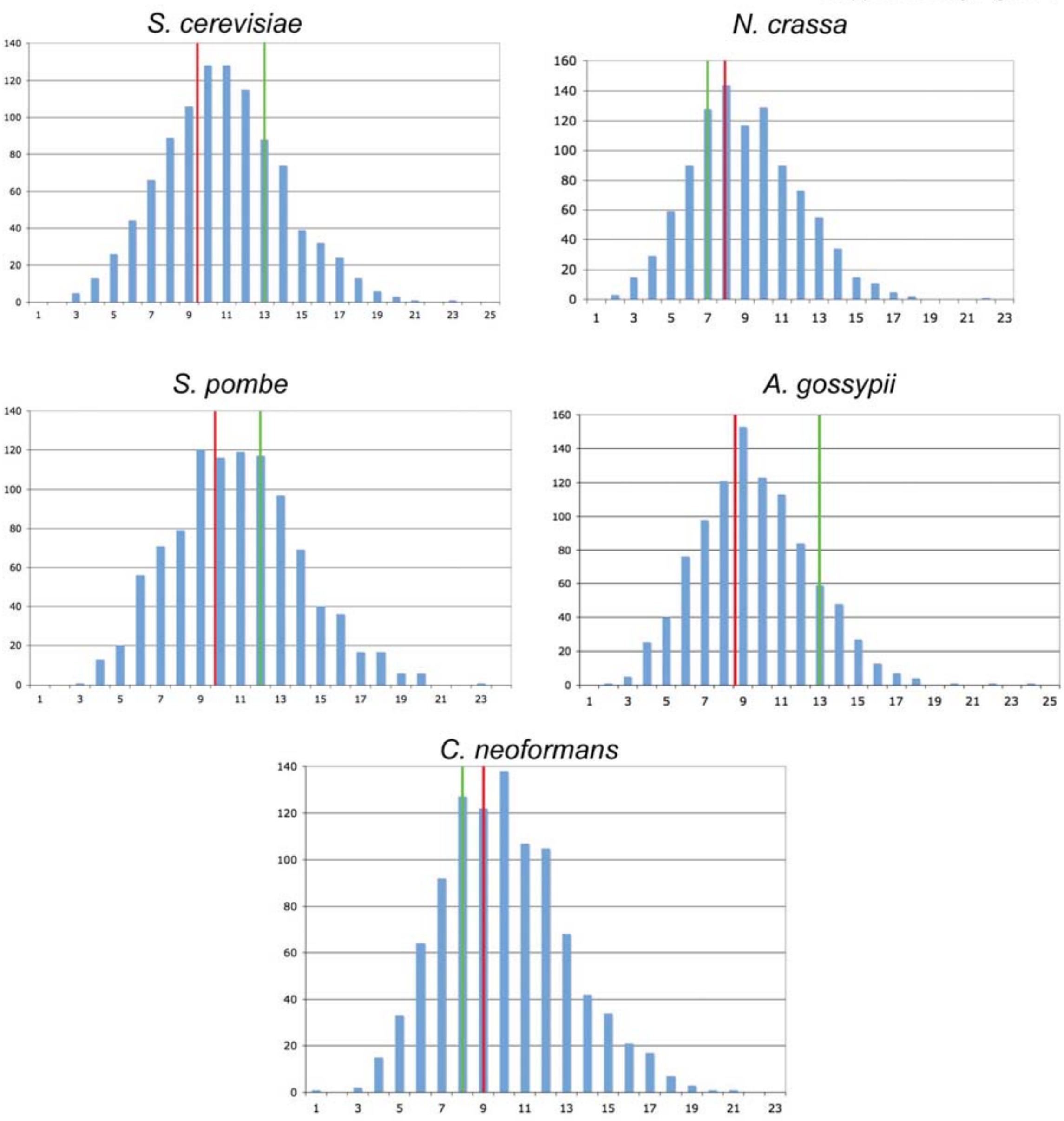


Figure S5. Synteny evaluation of *R. oryzae* and *E. cuniculi*. Examination of the 17,467 *R. oryzae* proteins (Broad Institute Database for *R. oryzae*) and the 1,996 *E. cuniculi* proteins[2] indicates that 3,457 *R. oryzae* proteins (20%) have homologs in *E. cuniculi*, and 978 *E. cuniculi* proteins (49%) have homologs in *R. oryzae*, where homologs are defined as having a BLASTp E-value of less than 1.0 x e^{-05} in an alignment of the two protein sets. The discrepancy in these gene sets is in part due to gene loss in one lineage or the other, and in part due to genes being too diverged to detect by blast. Thus for gene order analysis, we use a “relaxed synteny” definition where a pair of genes between these two species is defined to be syntenic if two genes in *E. cuniculi* that have at most three intervening genes have homologs in *R. oryzae* that are separated by at most 4 genes. If one of the intervening genes in *E. cuniculi* also has a homolog in *R. oryzae*, then the *R. oryzae* genes must be in a window of less than 15 genes. A total of 33 pairs (green line) of genes meet these criteria. When 1,000 randomizations of the *E. cuniculi* gene order were generated, an average of 17.3 pairs (red line) of “relaxed syntenic” genes were identified, with a standard deviation of 4.4. The number of 33 pairs of “relaxed syntenic” genes in the authentic genomes is 3.5 standard deviations higher than the average number of pairs of “relaxed syntenic” genes observed for the randomized gene order. All analyses were done using NCBI blast program and perl.

Supplementary Figure 6


S. cerevisiae

N. crassa

S. pombe

A. gossypii

C. neoformans

Figure S6. Synteny evaluation of *E. cuniculi* and other fungal lineages. To evaluate the significance of gene order conservation, the genomes of *Saccharomyces cerevisiae*, *Neurospora crassa*, *Schizosaccharomyces pombe*, *Ashbya gossypii*, and *Cryptococcus neoformans* were randomized 1000 times. Synteny was defined as described in the Figure 5S legend. Tandomly duplicated genes in each species were compressed into a single gene for purposes of this comparison. Blast cutoff threshold was 1 x e^{-5}. The numbers of gene pairs (green lines) observed between *E. cuniculi* and *S. cerevisiae* (13 pairs), *N, crassa* (7 pairs), *S. pombe* (12 pairs), *A. gossypii* (13 pairs), and *C. neoformans* (8 pairs) are not significantly different in comparison to those of the randomized gene orders (red lines, 9.79, 7.99, 9.82, 9.79, 8.59, and 9.02 pairs respectively). X axis: number of gene pairs, Y axis: number of occurrences.

Supplementary Figure 7

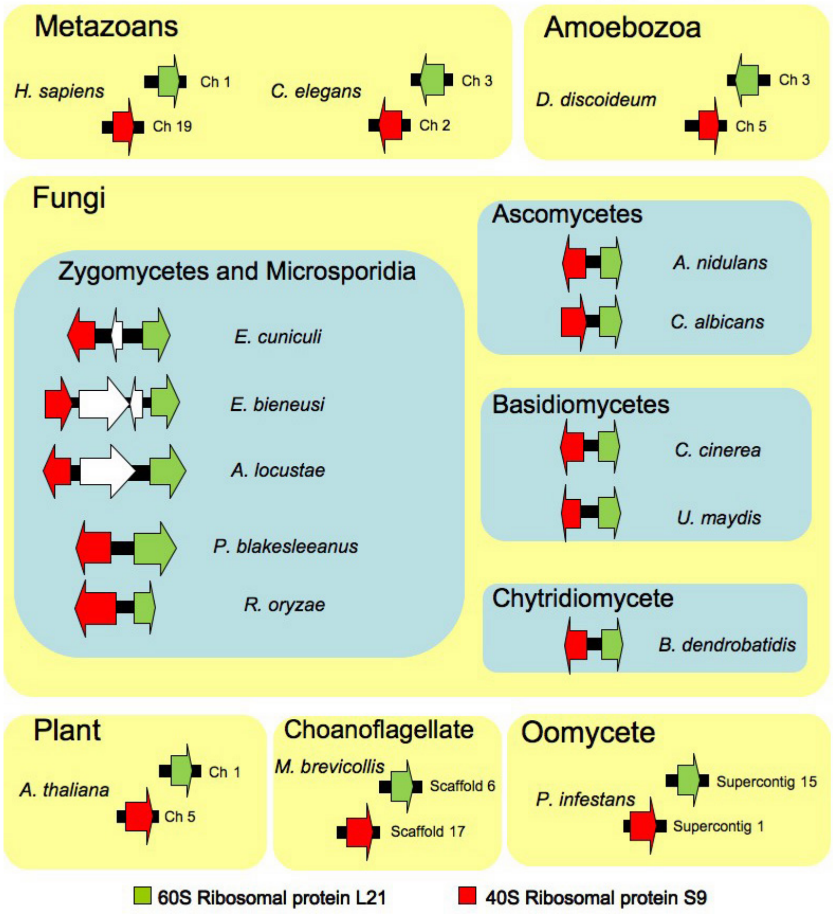


Figure S7. Synteny of the genes encoding 60S (L21) and 40S (S9) ribosomal proteins is fungal-specific. The synteny is conserved in ascomycetes (i.e., *Aspergillus nidulans* and *Candida albicans*), basidiomycetes (i.e., *Coprinopsis cinerea* and *Ustilago maydis)* and a chytridiomycete (i.e., *Batrachochytrium dendrobatidis*) as well as in *P. blakesleeanus*, *R. oryzae*, and the three microsporidia. However, the genes are not syntenic and are completely unlinked in metazoans (i.e., *Homo sapiens*, *Caenorhabditis elegans*), a amoebozoan (i.e., *Dictyostelium discoideum*), a choanoflagellate (i.e., *Monosiga brevicollis*), a plant (i.e., *Arabidopsis thaliana*), and an oomycete (i.e., *Phytophthora infestans*). Ch: chromosome. Gene sizes are not to scale.

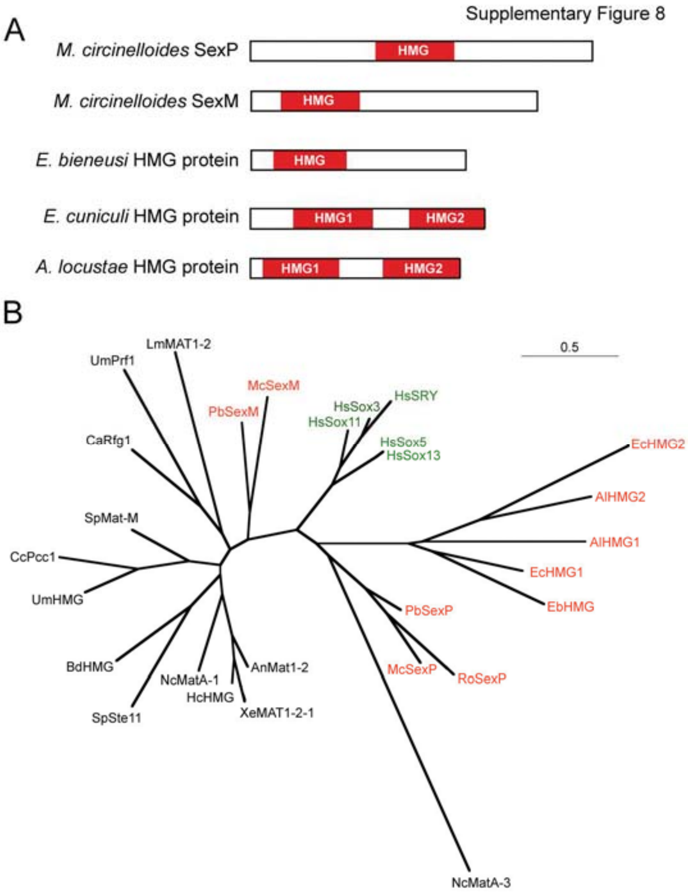


Figure S8. Analysis of HMG domain proteins. **(A)** *E. cuniculi* and *A. locustae* encode tandem HMG domains in a single HMG protein. However, *M. circinelloides* SexP and SexM contains a single HMG domain. *E. bieneusi*, a microsporidian, also contains only a single HMG domain in the HMG domain protein. **(B)** A maximum likelihood tree was drawn with the HMG domain sequences (approximately 70 amino acids) using ClustalW and the PHYLIP package. The tree suggests that the microsporidian HMG domain may have derived from an ancestral version of SexP. Human HMG domain proteins are shown in green. The HMG domains were identified with PROSITE tools. Scale bar indicates a branch length of 0.5 amino acid substitutions per position. An *E. cuniculi* protein with more limited HMG identity was highly diverged and is not included. Note that the alignable HMG domain is short, so none of the relationships are supported except the SexP and microsporidian clades. Ec: *Encephalitozoon cuniculi*, Al: *Antonospora locustae*, Eb: *Enterocytozoon bieneusi*, Pb: *Phycomyces blakesleeanus*, Ro: *Rhizopus oryzae*, Mc: *Mucor circinelloides*, Nc: *Neurospora crassa*, An: *Aspergillus nidulans*, Xe: *Xanthoria elegans*, Hc: *Histoplasma capsulatum*, Sp: *Schizosaccharomyces pombe*, Bd: *Batrachochytrium dendrobatidis*, Um: *Ustilago maydis*, Cc: *Coprinopsis cinerea*, Ca: *Candida albicans*, Lm: *Leptosphaeria maculans*, Hs: *Homo sapiens*.

Supplemental References

1. Ohno, S. (1967). Sex chromosomes and sex-linked genes, (New York, NY: Springer-Verlag).
2. Katinka, M.D., Duprat, S., Cornillot, E., Metenier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretaillade, E., Brottier, P., Wincker, P., et al. (2001). Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. Nature *414*, 450-453.