

Supplementary Figure Legends

Supplementary Figure S1. Schematic drawing of the *A. oryzae* genome sequence assembly. The 8 chromosomes of *A. oryzae* by the genome sequencing are represented from short arm to long arm together with the PFGE image of the *A. oryzae* chromosomes. The values at the left side of the PFGE indicate the positions of *Schizosaccharomyces pombe* chromosomes as size standards. Length of contigs are indicated in Mb except that a unit is indicated. No centromere is completely sequenced. Green, blue, yellow and white boxes represent a contig, a sequence gap, a physical gap linked to the neighboring contig by Southern hybridization and a physical gap linked to a chromosome but not linked to the neighboring contig, respectively. The longest sequence gap is 6.5 kb on chromosome 5. Telomeres and presumed centromeres are represented by red boxes and orange ovals. Additional 51,302 bp and 18,590 bp short contigs were mapped at the physical gaps on the chromosome 1 and 3, respectively, the latter was linked to the 2.3 Mb contig on the chromosome 3. The location of the rRNA repeat cluster is represented by a gray box. The direction of the repeat telomeric to centromeric is 28S to 18S. The entire nuclear genome size is calculated to be 38.4 Mb when the sizes of the unsequenced gaps estimated by the optical mapping are included.

Supplementary Figure S2. Distribution and expression of genes throughout the *A. oryzae* genome. Non metabolism, Q genes, Extra homologs, AO specific genes, Genes with EST and Gene density are aligned with the eight *A. oryzae* chromosomes together with synteny to the *A. fumigatus* and *A. nidulans* genomes. Symbols represent the same with those in Fig. 1.

Supplementary Figure S3. Gene redundancy of pathway members. The amino acid sequences of *S. cerevisiae* genes for metabolic pathways members were obtained by SGD database (Dolinski, K. et al., <http://www.yeastgenome.org/>) and were searched by BLASTP with E-value of $\leq 10^{-20}$. A set of three horizontal bars represents the numbers

of genes existing in the three *Aspergillus* species homologous to a particular gene in the pathways. Gene(s) involved in a particular pathway are indicated by a vertical line. The figures indicate a part of the pathways among 166 pathways analyzed; 1, S-adenosylhomocysteine catabolism; 2, phenylalanine degradation; 3, proline utilization; 4, L-serine degradation; 5, threonine catabolism; 6, tryptophan degradation; 7, tryptophan degradation via kynurenine; 8, valine degradation; 9, toluene degradation, via catechol; 10, aldoxime metabolism; 11, m-cresol degradation; 12, p-cymene degradation; 13, carbon monoxide dehydrogenase pathway; 14, RuMP cycle and formaldehyde assimilation (ribulose-5-phosphate); 15, RuMP cycle and formaldehyde assimilation (CO₂); 16, serine-isocitrate lyase pathway; Orange, purple and green bars represent the number of genes from *A. nidulans*, *A. fumigatus* and *A. oryzae*, respectively.

Supplementary Figure S4. Metabolic pathway view of the gene redundancy of pathway members. The gene redundancy of each member of the metabolic pathways (A, carbohydrate; B, amino acid; C, fatty acid) in *A. oryzae* genome as compared to *A. fumigatus* and *A. nidulans* was analyzed as same as in Supplementary Fig. S3, and was superimposed on the metabolic pathway map. The numbers of corresponding genes from *A. oryzae*, *A. nidulans*, *A. fumigatus* are indicated for each reaction. Red and green lines indicate that *A. oryzae* has ≥ 1.5 -fold of corresponding genes than *A. fumigatus* or *A. nidulans*, and that *A. fumigatus* or *A. nidulans* has ≥ 1.5 -fold of the corresponding genes than *A. oryzae*, respectively. The metabolic pathway view was constructed based on the pathways existing in *Saccharomyces cerevisiae* referring to KEGG (Kanehisa, M. et al., <http://www.genome.ad.jp/kegg/>) and SGD. Light gray lines indicate the genes that have not been well characterized in *S. cerevisiae* or the genes that were not identified in the *A. oryzae* genome by homology search. The three values delimited by slashes designate the number of corresponding genes found in *A. oryzae*, *A. fumigatus* and *A. nidulans* genomes, respectively.

Supplementary Figure S5. Phylogenetic analysis of proteases and glucosidases. Phylogenetic relationship of carboxypeptidases **(a)**, ATP-dependent proteinases **(b)**, maltase **(c)** and α -amylase genes **(d)** from the three Aspergilli were analyzed. The programs, algorithms and symbols are same with those in Fig. 3.

Supplementary Figure S6. Phylogenetic analysis of metabolic genes. Phylogenetic relationship of pyruvate decarboxylase **(a)**, saccharopine dehydrogenase, homoaconitase and saccharopine dehydrogenase (NADP⁺, L-glutamate forming) in lysine biosynthesis **(b)** from the three Aspergilli were analyzed. The programs, algorithms and symbols are same with those in Fig. 3.

Supplementary Figure S7. Analysis of a pair of syntenic genes highly homologous between *A. oryzae* and *A. tumefaciens*. **(a)** The upstream gene (AO070319000102 and AGR_L_1866) have no homologs to functionary known genes, and the downstream genes (AO070319000101 and AGR_L_1864) are annotated to be biotin carboxylase/acyl-CoA carboxylase. While the two genes of *A. tumefaciens* have a 4 bp overlap, the *A. oryzae* genes are separated by an approximately 300 bp insertion (dotted line). The upstream gene (AO070319000102) has a putative intron at 904-945 nt from the translation initiation codon. **(b)** Phylogenetic analysis using amino acid sequences of biotin carboxylases and acyl-CoA carboxylases from variety of organisms was carried out based on the same method as in Fig. 3. Orthologous clusters from the three Aspergilli are designated by Common to Aspergilli, 1 thru 5. The *A. oryzae* specific and the *Agrobacterium tumefaciens* enzymes were indicated by orange characters with an arrow. **(c)** Phylogenetic analysis using amino acid sequences of AO070319000102 and its orthologous genes.