Current Biology 22, 1403–1409, August 7, 2012 ©2012 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2012.05.033

Report

The Evolutionary Imprint of Domestication on Genome Variation and Function of the Filamentous Fungus *Aspergillus oryzae*

John G. Gibbons,[1] Leonidas Salichos,[1] Jason C. Slot,[1] David C. Rinker,[1,2] Kriston L. McGary,[1] Jonas G. King,[1] Maren A. Klich,[3] David L. Tabb,[4] W. Hayes McDonald,[5] and Antonis Rokas[1,2,*]

[1]Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA
[2]Center for Human Genetics Research, Vanderbilt University, Nashville, TN 37232, USA
[3]Southern Regional Research Center, Agricultural Research Service, United States Department of Agriculture, New Orleans, LA 70124, USA
[4]Department of Biomedical Informatics
[5]Department of Biochemistry
Vanderbilt University Medical Center, Nashville, TN 37232, USA

Summary

The domestication of animals, plants, and microbes fundamentally transformed the lifestyle and demography of the human species [1]. Although the genetic and functional underpinnings of animal and plant domestication are well understood, little is known about microbe domestication [2–6]. Here, we systematically examined genome-wide sequence and functional variation between the domesticated fungus *Aspergillus oryzae*, whose saccharification abilities humans have harnessed for thousands of years to produce sake, soy sauce, and miso from starch-rich grains, and its wild relative *A. flavus*, a potentially toxigenic plant and animal pathogen [7]. We discovered dramatic changes in the sequence variation and abundance profiles of genes and wholesale primary and secondary metabolic pathways between domesticated and wild relative isolates during growth on rice. Our data suggest that, through selection by humans, an atoxigenic lineage of *A. flavus* gradually evolved into a "cell factory" for enzymes and metabolites involved in the saccharification process. These results suggest that whereas animal and plant domestication was largely driven by Neolithic "genetic tinkering" of developmental pathways, microbe domestication was driven by extensive remodeling of metabolism.

Results and Discussion

Examination of several plants and animals suggests that domestication was driven by genetic changes in diverse developmental pathways that ultimately led to large fruits, naked grains, small brains, and big bodies [1, 8, 9]. Although the molecular genetics and phenotypic outcomes of crop and livestock domestication have been extensively studied [8–10], the evolutionary paths traversed by domesticated microbes remain poorly understood [2–6]. In China, evidence for a fermented beverage based on rice mixed with honey and fruit dates back to 7,000 B.C.E. [11]. Over the millennia that followed, the gradual development of the saccharification process, in which filamentous fungi break down the starch-rich rice to sugars that yeast ferments, morphed the beverage into the high-alcohol rice wine known as sake [11–15]. The filamentous fungus used in saccharification for making sake, as well as other traditional Japanese food products such as soy sauce and miso, is *Aspergillus oryzae* (class Eurotiomycetes, phylum Ascomycota). For sake making, *A. oryzae* spores (koji-kin) are first spread onto steamed rice. After an ~2-day growth period, the resulting *A. oryzae*-rice mixture (koji) is mixed with additional steamed rice and water and fermented by *Saccharomyces cerevisiae*, such that the breakdown of the rice starch by *A. oryzae* occurs in parallel with the conversion of sugars to alcohol by *S. cerevisiae* [16]. However, the saccharific and more generally proteolytic and metabolic activities of *A. oryzae* not only fuel the yeast but also contribute metabolites that influence the flavor and aroma of sake [16].

A. oryzae is closely related to the wild species *A. flavus* [17, 18], the two species sharing 99.5% genome-wide nucleotide similarity [19]. However, *A. oryzae* is an atoxigenic domesticate recognized by the United States Department of Agriculture as a Generally Regarded As Safe (GRAS) organism [7], whereas *A. flavus* is a destructive agricultural pest of several seed crops and producer of the potent natural carcinogen aflatoxin [20]. This striking contrast between genomic and phenotypic variation makes the *A. oryzae*-*A. flavus* lineage an excellent microbe domestication model for the study of the functional changes associated with microbe domestication and the impact of the process on genome variation [6, 7, 21, 22].

Domesticated organisms have typically been selected for beneficial traits conferred by certain genetic loci and have undergone several rounds of population bottlenecks. Although we previously did not find evidence that the *A. oryzae* genome exhibited a relaxation of selective constraints, a common characteristic accompanying plant and animal domestication [6], whether the *A. oryzae* genome has experienced positive selection during the domestication process remained an open question. To address this question, we Illumina sequenced 14 geographically and industrially diverse isolates from *A. oryzae* and *A. flavus* and jointly analyzed them with the two species' reference genomes [7, 23] (*A. oryzae* RIB 40 and *A. flavus* NRRL 3357; see Tables S1 and S2 available online). Analysis of the genome-wide nucleotide diversity across the 16 isolates showed that the genetic diversity of the *A. oryzae* isolates was ~25% of that found in the *A. flavus* isolates (chromosome average nucleotide variation $\Theta_{A.\ oryzae}$ = 0.0006 versus $\Theta_{A.\ flavus}$ = 0.0024; t test, p = 4.1×10^{-7}), consistent with previous gene-level estimates [17, 24, 25] (Figure 1A). Evolutionary analysis of 100,084 high-quality SNPs (see Supplemental Experimental Procedures) suggested that the *A. oryzae* isolates were monophyletic, in agreement with the previous hypotheses that *A. oryzae* originated via a single domestication event [17, 25], and did not group by geography or ecology (Figure 1B). Interestingly, two *A. flavus* isolates (SRRC 1357 and SRRC 2112) showed closer affinity to *A. oryzae* than to other *A. flavus* isolates (Figure 1B), suggesting that *A. oryzae* originated from within *A. flavus*.

*Correspondence: antonis.rokas@vanderbilt.edu

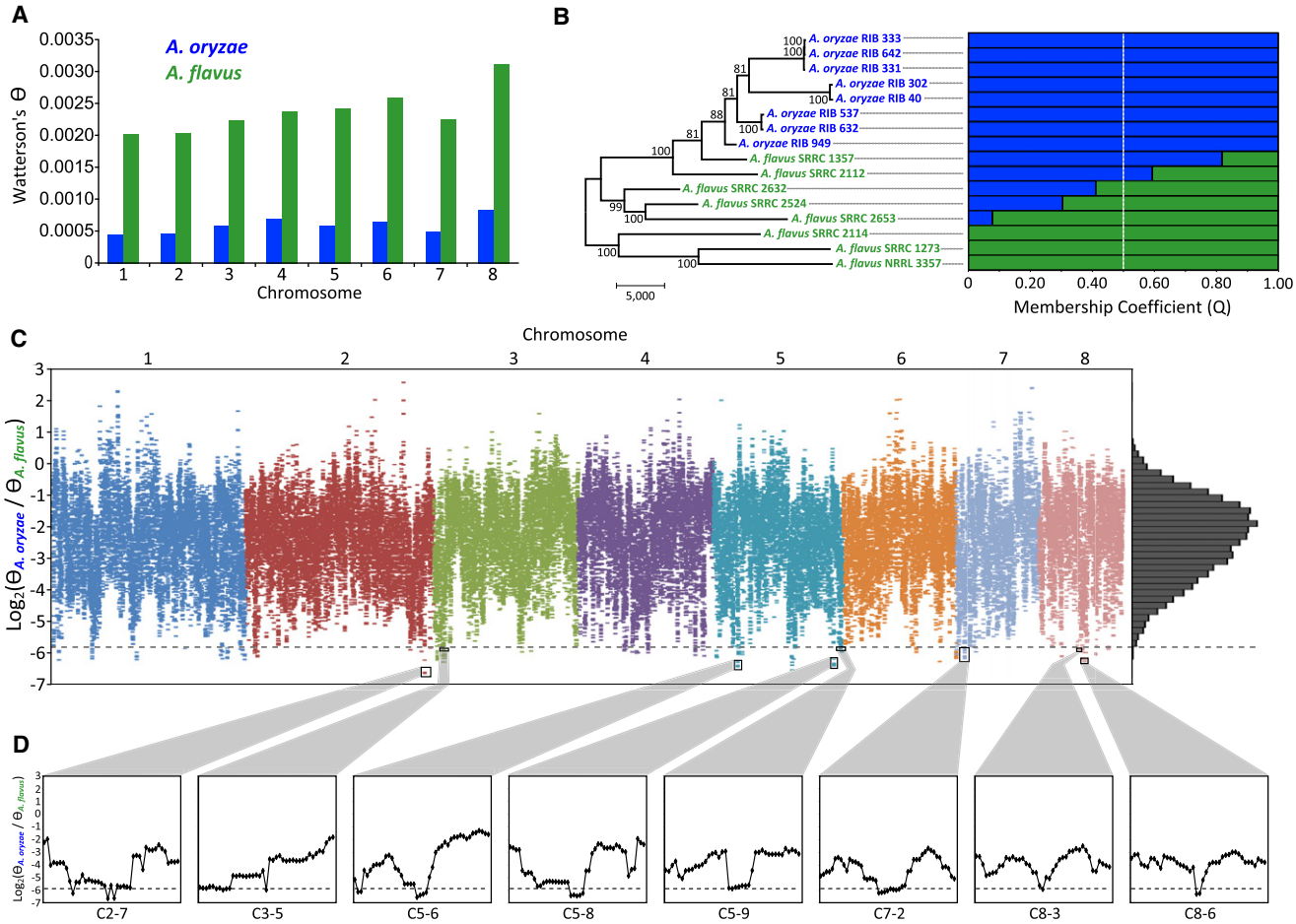

Figure 1. Phylogenetic Relationship and Genomic Patterns of Variation in *Aspergillus oryzae* and *A. flavus*

(A) Chromosomal levels of nucleotide variation (Θ) in *A. oryzae* (blue) and *A. flavus* (green).

(B) Left: parsimony-inferred phylogeny of the 16 *A. oryzae* (blue) and *A. flavus* (green) isolates from the 100,084 high-quality genome-wide variant sites. Values near internodes indicate bootstrap support, generated by 1,000 replicates. The scale bar represents the number of changes. Right: STRUCTURE-based membership coefficient for each isolate (population number K = 2). The *A. oryzae* and *A. flavus* genetic backgrounds are shown in blue and green, respectively.

(C) Relative nucleotide diversity scores (Θ_{OF}) for 5 kb windows (65,894 windows) with a 500 bp step size scanning the eight chromosomes. Points below the dashed line represent genomic regions below the empirical 0.25% quantile (164 windows) and comprise the candidate putative selective sweep regions (PSSRs). The distribution of Θ_{OF} scores is shown at right.

(D) Close-ups of representative PSSRs and flanking regions.

One of the footprints of recent selection on the genome is the reduction in variation of regions that are close to the variants under selection [26]. When a beneficial allele is rapidly driven toward fixation, nearby neutral variants are likely to also become fixed as a result of the low rate of recombination between closely linked sites [27]. By estimating the relative genome-wide nucleotide diversity $\Theta_{OF} = \log_2 (\Theta_{A.\ oryzae}/\Theta_{A.\ flavus})$, we identified 61 putative selective sweep regions (PSSRs) (Figures 1C and 1D; Table S3; see Supplemental Experimental Procedures). Examination of PSSR gene content indicated that the main targets of selection were genes and pathways involved in primary metabolism (PM) and secondary metabolism (SM). For example, the 148 PSSR genes were significantly overrepresented for SM (Fisher's exact test [FET], p = 0.0004), whereas five PSSRs contained SM gene clusters, including one for the biosynthesis of the tremorgenic mycotoxin aflatrem (PSSR C5-9; Figures 1C and 1D) [28]. These results were particularly noteworthy because SM gene families are thought to have expanded and to be located in unique genomic regions of the *A. oryzae-A. flavus* lineage compared to the far more distantly related species *A. fumigatus* and *A. nidulans* [7]. Furthermore, several PSSR genes are involved in protein and peptide degradation (genes in PSSRs C2-7 and C5-8) and carbohydrate metabolism (C3-5 and C5-6) (Figures 1C and 1D). One of the strongest supported PSSRs (C8-6) contained a glutaminase (Figures 1C and 1D), which catalyzes the hydrolysis of carbon-nitrogen bonds of L-glutamine to glutamic acid, a widely used food flavor enhancer found at considerable levels in sake [29]. Strikingly, whereas there were six polymorphic sites within the *A. oryzae* isolates (two promoter and four intron), *A. flavus* isolates were polymorphic at 86 sites (14 synonymous, 2 nonsynonymous, 18 promoter region, and 52 intron) (Figure S1).

We also examined the isolate genome data to identify differences in genome architecture between the two species (see Supplemental Experimental Procedures). Although our search identified only five genes shared uniquely by all *A. oryzae* isolates and none by *A. flavus* isolates (Table S4), it did also

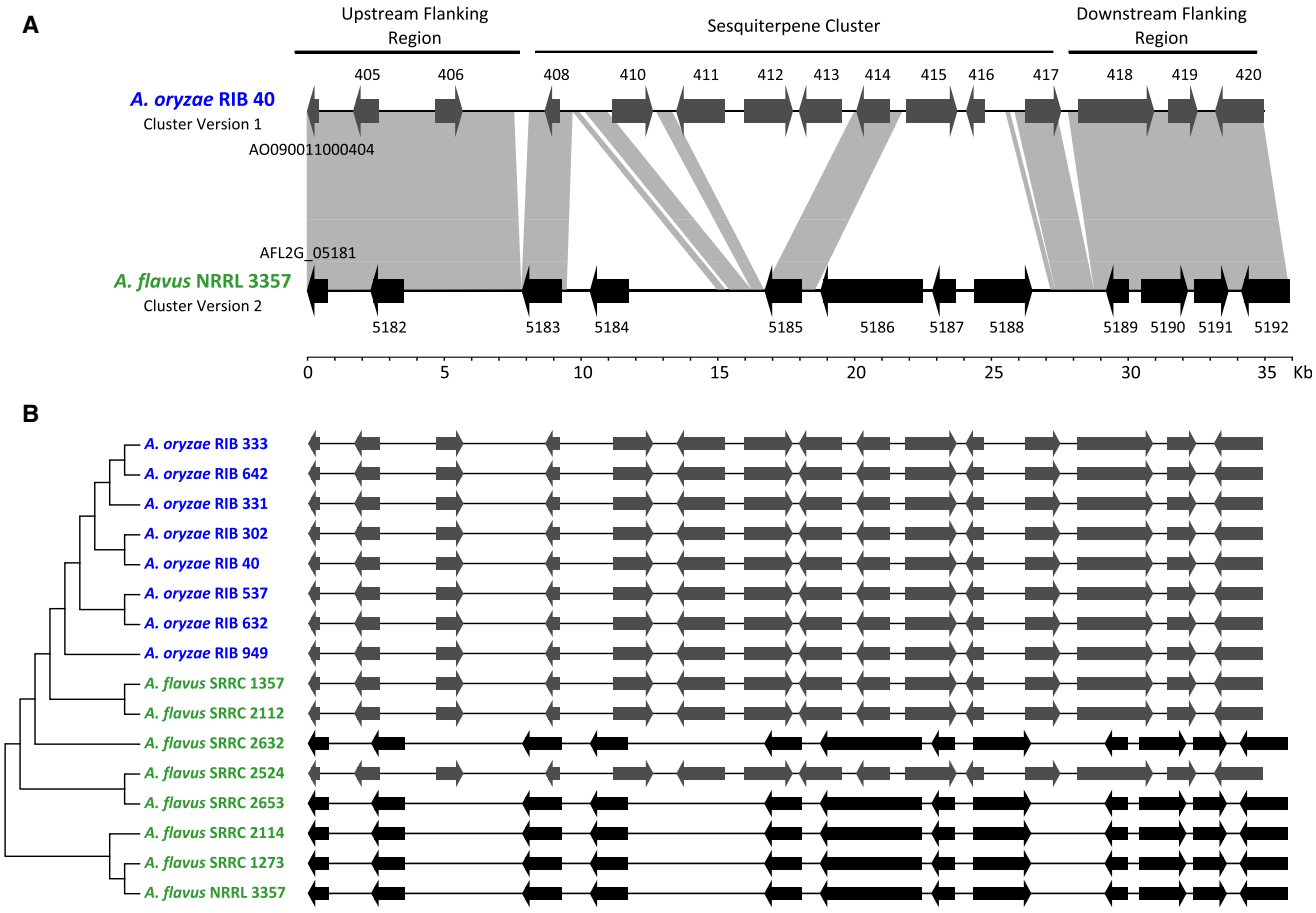

Figure 2. The Variable Genome Architecture of the Sesquiterpene Cluster Locus

(A) Microsynteny of the locus harboring the sesquiterpene-encoding gene cluster and its flanking regions in *A. oryzae* RIB 40 and *A. flavus* NRRL 3357 isolates. Gray blocks represent genomic regions exhibiting significant sequence similarity between species. Genes and the direction of transcription are symbolized by arrows and labeled. The *A. oryzae* RIB 40 genome contains a nine-gene cluster "allele," whereas the *A. flavus* NRRL 3357 genome contains a six-gene cluster "allele." Only the terpene cyclase (AO090011000408) and GAPDH (AO090011000414), as well as a few noncoding regions, are homologous between the two "alleles."

(B) Graph showing the allele present in each of the 16 isolates. Note that all eight *A. oryzae* isolates contain the nine-gene cluster "allele," whereas *A. flavus* is polymorphic.

identify a locus that contains a nine-gene cluster in the *A. oryzae* genome but contains a six-gene cluster in the *A. flavus* NRRL 3357 genome (Figure 2A). Interestingly, the nine-gene cluster is very similar to the sesquiterpene gene cluster in *Trichoderma virens* [30, 31], whose product belongs to a class of food-flavoring aromatic compounds [32], whereas the six-gene cluster comprises a terpene cyclase and GAPDH from the nine-gene cluster together with four other unrelated genes (Figure S2). Remarkably, although *A. oryzae* is fixed for the nine-gene cluster, *A. flavus* is polymorphic: three isolates contained the nine-gene cluster, whereas the other five contained the alternative six-gene cluster (Figure 2B). Furthermore, the genes contained in the two alternative cluster "alleles" at this locus have different evolutionary histories (Figure S2). Most unique genes of the nine-gene cluster group with sequences from *A. clavatus* and divergent fungi related to *T. virens*, consistent with horizontal transfer, whereas most *A. flavus* unique genes of the alternative cluster group with sequences from *A. aculeatus*, suggesting a very different history.

A. oryzae has been grown continually on starch-rich substrates, such as rice and soy, for thousands of years [7, 14]. To identify functional differences and putative adaptations to this starch-rich diet, we examined the transcriptome profiles of three phylogenetically distinct isolates of sake-derived *A. oryzae*, as well as the proteome profiles of the reference isolate of each species, during growth on rice. Similar to the analyses of the PSSR gene content, comparison of the transcriptome and proteome profiles between *A. oryzae* and *A. flavus* identified several differentially abundant transcripts, proteins, and pathways involved in PM and SM.

All *A. oryzae* isolates possess two or three copies of α-amylase [7, 21], the enzyme that hydrolyzes the α-D-glycosidic bonds of starch to produce dextrin, compared to a single copy in *A. flavus*. We found that the transcript and protein abundance of α-amylase was the highest of any *A. oryzae* gene or protein and was significantly upregulated compared to *A. flavus* (gene expression: FET, $p < 1 \times 10^{-300}$; protein abundance: >30-fold, FET, $p = 2.15 \times 10^{-51}$) (Figure 3; Tables S5–S8). Several other *A. oryzae* upregulated genes are involved in carbohydrate PM, including the genome neighbors amylolytic transcriptional activator *amyR* [33] (FET, $p = 1.68 \times 10^{-97}$) and saccharide-metabolizing enzyme maltase glucoamylase (FET, $p = 1.79 \times 10^{-17}$), as well as the

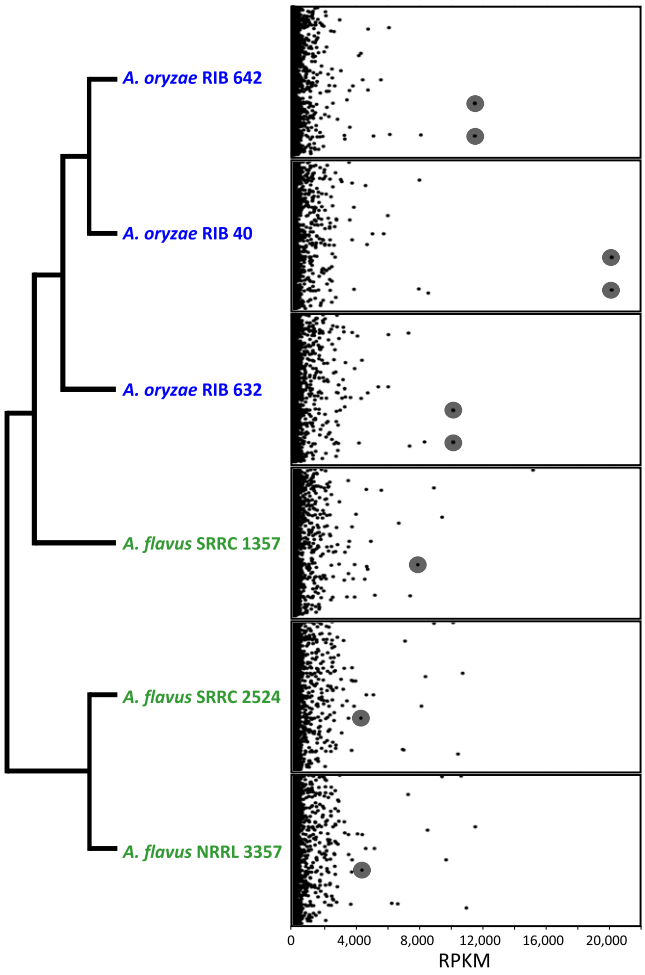

Figure 3. α-Amylase Is the Most Highly Expressed Transcript in *A. oryzae*

Expression levels (in reads per kilobase per million mapped reads; RPKM) (x axis) of all genes (y axis) for each of the six isolates organized by their phylogenetic relatedness. The two α-amylase paralogs are highlighted in gray. Expression levels for the two paralogs are depicted as equal because they have identical coding sequences and differentiation of their expression levels is not possible.

glucose-metabolizing enzyme sorbitol dehydrogenase (FET, $p = 8.22 \times 10^{-252}$) (Figure S3; Tables S6 and S8). Importantly, comparison of the transcriptional profile of the two species showed that both the upregulated and downregulated gene sets in *A. oryzae* were overrepresented for carbohydrate PM (FET; $p = 6.24 \times 10^{-5}$ and $p = 4.22 \times 10^{-12}$, respectively), suggesting that differential regulation of PM is a key functional difference between the two species.

A. oryzae is also equipped with an arsenal of secreted enzymes that break down the proteins and complex polysaccharides of grain outer layers, providing access to the starch-rich interior layers [7, 13, 16, 34]. Several protease-encoding genes are located in PSSRs (e.g., the methionine aminopeptidase located in the PSSR C5-8), are upregulated (e.g., extracellular cellulase *celA*), or both (e.g., the upregulated proteinase located in PSSR C2-7) (Figures 1C and 1D; Tables S3 and S5–S8). In contrast, 16 of the 27 plant polysaccharide-degrading genes are downregulated (a few of them are also located in PSSRs, e.g., endoglucanase and feruloyl esterase in PSS C3-5 and endo-1,4-β-xylanase in PSS C5-6) (Tables S5, S7, and S8). The broad downregulation of this subset of genes likely reflects differences between *A. oryzae* and *A. flavus*.

Comparison of the gene expression profiles of 610 genes in all 55 predicted SM gene clusters [35] against background genes in the two species showed that another general characteristic of the *A. oryzae* transcriptome during growth on rice is SM downregulation (FET, $p = 7.3 \times 10^{-10}$). This is consistent with the wholesale downregulation of five SM gene clusters in *A. oryzae* (Figure 4). Importantly, both the cyclopiazonic acid and the aflatoxin SM pathways in *A. oryzae* were downregulated (Figures 4A and 4B), explaining a key phenotypic difference between *A. oryzae* and *A. flavus*, the inability of the first to produce either of the two toxins [7, 22, 36]. We further investigated sequence variation in the isolates using expression data with respect to five previously characterized types of mutations observed at the aflatoxin gene cluster locus: (1) transcription binding-site mutations in the *aflR* promoter [37], (2) an ~250 bp 3′ deletion in the *aflT* coding region [37], (3) a frameshift mutation in the *norA* coding region [37], (4) multiple nonsynonymous mutations in the *verA* coding region [37], and (5) an ~40 kb deletion spanning the genomic region between the *norB* and the *norA* genes, inclusive [38]. This analysis revealed mutation 1 in *A. oryzae* RIB 632 and mutations 1–4 in *A. oryzae* RIB 632 and RIB 40 [37] when compared to *A. flavus* NRRL 3357. Furthermore, the *A. oryzae*-like isolate *A. flavus* SRRC 1357 contained 5 and 13 nonsynonymous mutations in the *aflT* (2) and *verA* (4) genes respectively, whereas *A. flavus* SRRC 2524 was nearly identical to *A. flavus* NRRL 3357 (three and one synonymous mutations in the *norA* [3] and *verA* [4] genes). Interestingly, aflatoxin is genotoxic to *S. cerevisiae* [39], suggesting that the atoxicity of *A. oryzae* might have been driven by its impact on yeast survival and, as a consequence, fermentation for making sake.

A. flavus natural isolates show substantial variation in SM production, and several are known to be atoxigenic [20, 22, 36, 38, 40–42]. Interestingly, the SM expression profile of the atoxigenic *A. flavus* SRRC 1357, the isolate most closely related to *A. oryzae* (Figure 1A), was more similar to *A. oryzae* than to those of the other *A. flavus* isolates [Figures 4A–4E (C7-2); Table S6], consistent with the hypothesis that *A. oryzae* was domesticated from an atoxigenic clade of *A. flavus*.

During malt rice (koji) production, *A. oryzae* also produces a variety of aromatic, flavor-producing volatile compounds and associated enzymes [16, 43, 44]. In addition to the sequence and genome architecture differences observed in the glutaminase and sesquiterpene loci, we also detected functional differences in other industrially associated genes. Two particularly interesting examples of upregulated genes include a glycosyl transferase (FET, $p = 1.75 \times 10^{-237}$), a member of a broad sugar modifier family involved in the making of many sweeteners [45] (Tables S6 and S8), and an asparaginase (gene expression: FET, $p = 1.29 \times 10^{-15}$; protein abundance: FET, $p = 0.006$), an enzyme used commercially to reduce acrylamide levels in starch-rich foods, such as rice [46] (Tables S6–S8). Surprisingly, however, of the more than 500 genes annotated as MFS or ABC transporters, only six were upregulated in all *A. oryzae* isolates when compared to all *A. flavus* isolates, and an additional six were upregulated in the *A. oryzae* isolates and the closely related *A. flavus* isolate when compared against all other *A. flavus* isolates (Figure S3).

In summary, our systematic comparison of sequence, gene expression, and protein abundance variation in the *A. oryzae*-*A. flavus* lineage indicates that *A. oryzae* domestication was

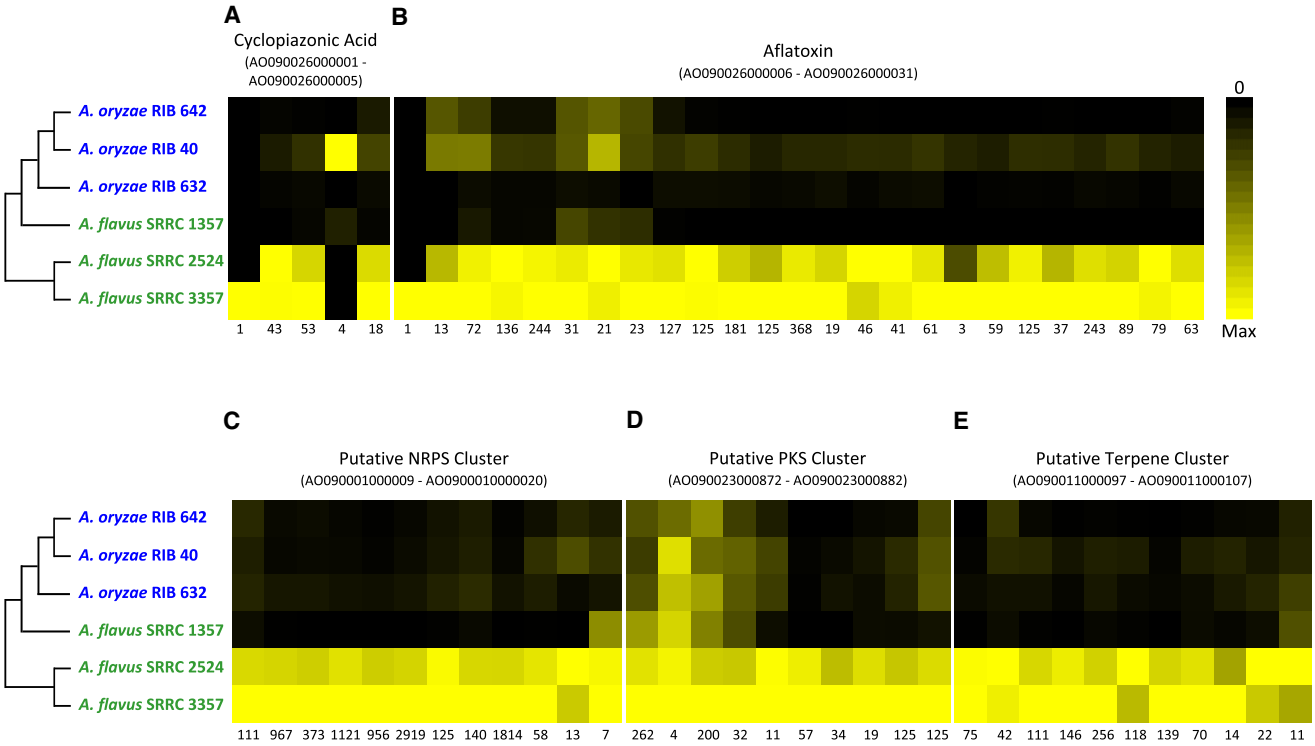

Figure 4. The *A. oryzae* Secondary Metabolism Transcriptome Is Widely Downregulated during Growth on Rice

Expression levels for the three *A. oryzae* and three *A. flavus* isolates for five secondary metabolism biosynthesis gene clusters: cyclopiazonic acid (A), aflatoxin (B), putative nonribosomal peptide metabolite (C), putative polyketide synthase metabolite (D), and putative terpene (E). The range of genes included in each gene cluster is given under the cluster's name. For each gene, the color of the heat map cell corresponds to its expression level (in RPKM units), where black is zero expression and yellow is the maximum RPKM for that gene (listed below each gene).

accompanied by dramatic changes in primary and secondary metabolism. In a span of a few millennia, unintentional human breeding of predominantly segregating variation present in *A. flavus* resulted, through the gradual accumulation of small-scale (e.g., Figures 1C, 1D, and S1) and large-scale (e.g., Figures 2 and 3) genetic and functional changes (e.g., Figures 3 and 4), in the evolution of the saccharific and proteolytic *A. oryzae* "cell factory." Although alterations in metabolic pathways were also likely targets of selection during both plant and animal domestication [47], the majority of these changes were driven primarily by modifications in developmental pathways that affect growth and form. In stark contrast, the present study and previous findings [4, 48–53] argue that the molecular foundations of microbe domestication rested largely in the restructuring of metabolism.

Experimental Procedures

See Supplemental Experimental Procedures for full description.

Sequencing and Proteomics

gDNA and mRNA libraries were prepared as described previously [54, 55] and sequenced on an Illumina Genome Analyzer II. We used multidimensional protein identification technology (MudPIT) to examine the proteomic profile of *A. oryzae* RIB40 and *A. flavus* NRRL 3357. For mRNA and proteomics samples, isolates were grown on rice at 30°C for 24 hr to mimic sake-making conditions. Raw Illumina sequence reads were submitted to the NCBI Sequence Read Archive (SRA) (see below for Accession Numbers). Raw proteomics data were submitted to Tranche and can be downloaded from the Vanderbilt MSRC Bioinformatics Data site (http://www.mc.vanderbilt.edu/root/vumc.php?site=msrc/bioinformatics&doc=21164).

SNP Detection and Evolutionary Analysis

We used the Maq package [56] to identify SNPs for each isolate by mapping genomic reads against the *A. oryzae* RIB40 reference genome. We required that SNP sites have $\geq 5\times$ coverage, have an average quality score ≥ 20, and not be ambiguously called in any isolate. We then extracted the nucleotide from each variant site in all isolates. The alignment of variant sites was used to infer the phylogenetic relationships and population structure of our isolates.

Selective Sweep Detection

Using VariScan [57], we measured relative nucleotide diversity $\Theta_{OF} = \log_2(\Theta_{A.\ oryzae}/\Theta_{A.\ flavus})$ in 5 kb windows, with a 500 bp step size to detect regions of the *A. oryzae* genome with relatively reduced levels of variation. We considered the lower 0.25% quantile of Θ_{OF} values as putative selective sweep regions (PSSRs).

Gene Expression and Protein Abundance Analysis

Using the rSeq package [58], mRNA reads were mapped against the *A. oryzae* RIB40 reference transcriptome, and gene expression was quantified in terms of reads per kilobase per million mapped reads (RPKM) [59]. We identified differentially expressed genes and differentially abundant proteins between the reference strains of each species by comparing the proportion of mapped reads using Fisher's exact tests. Species-level and clade-level gene expression upregulation was further identified where all isolates of a group were expressed $\geq$10 RPKM and upregulated by at least 1.5-fold versus all isolates of the other group.

Accession Numbers

Raw Illumina sequence reads have been submitted to the NCBI Sequence Read Archive (SRA) with the accession numbers SRA0502658 (*A. oryzae* gDNA), SRA052664 (*A. flavus* gDNA), SRA0502666 (*A. oryzae* RNaseq), and SRA052667 (*A. flavus* RNaseq).

Supplemental Information

Supplemental Information includes three figures, eight tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2012.05.033.

Acknowledgments

We thank members of the Rokas lab, Chris Hittinger, Shannon Beltz, David Geiser, Kathy Friedman, David Friedman, Jim Patton, Julian Hillyer, Kamya Rajaram, Abby Olena, Scott Egan, Jonathan Flowers, David McCauley, Travis Clark, Chelsea Baker, Osamu Yamada, and the National Research Institute of Brewing of Japan. J.G.G. is funded by the Graduate Program in Biological Sciences at Vanderbilt University and the National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH, NIAID: F31AI091343-01). Research in A.R.'s lab is supported by the Searle Scholars Program and the National Science Foundation (DEB-0844968).

Received: April 4, 2012
Revised: May 13, 2012
Accepted: May 15, 2012
Published online: July 12, 2012

References

1. Diamond, J. (2002). Evolution, consequences and future of plant and animal domestication. Nature *418*, 700–707.
2. Hyma, K.E., Saerens, S.M., Verstrepen, K.J., and Fay, J.C. (2011). Divergence in wine characteristics produced by wild and domesticated strains of *Saccharomyces cerevisiae*. FEMS Yeast Res. *11*, 540–551.
3. Legras, J.L., Merdinoglu, D., Cornuet, J.M., and Karst, F. (2007). Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. Mol. Ecol. *16*, 2091–2102.
4. Libkind, D., Hittinger, C.T., Valério, E., Gonçalves, C., Dover, J., Johnston, M., Gonçalves, P., and Sampaio, J.P. (2011). Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. Proc. Natl. Acad. Sci. USA *108*, 14539–14544.
5. Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V., et al. (2009). Population genomics of domestic and wild yeasts. Nature *458*, 337–341.
6. Rokas, A. (2009). The effect of domestication on the fungal proteome. Trends Genet. *25*, 60–63.
7. Machida, M., Asai, K., Sano, M., Tanaka, T., Kumagai, T., Terai, G., Kusumoto, K., Arima, T., Akita, O., Kashiwagi, Y., et al. (2005). Genome sequencing and analysis of *Aspergillus oryzae*. Nature *438*, 1157–1161.
8. Doebley, J.F., Gaut, B.S., and Smith, B.D. (2006). The molecular genetics of crop domestication. Cell *127*, 1309–1321.
9. Purugganan, M.D., and Fuller, D.Q. (2009). The nature of selection during plant domestication. Nature *457*, 843–848.
10. Andersson, L., and Georges, M. (2004). Domestic-animal genomics: deciphering the genetics of complex traits. Nat. Rev. Genet. *5*, 202–212.
11. McGovern, P.E., Zhang, J.H., Tang, J.G., Zhang, Z.Q., Hall, G.R., Moreau, R.A., Nuñez, A., Butrym, E.D., Richards, M.P., Wang, C.S., et al. (2004). Fermented beverages of pre- and proto-historic China. Proc. Natl. Acad. Sci. USA *101*, 17593–17598.
12. Abe, K., Gomi, K., Hasegawa, F., and Machida, M. (2006). Impact of *Aspergillus oryzae* genomics on industrial production of metabolites. Mycopathologia *162*, 143–153.
13. Kobayashi, T., Abe, K., Asai, K., Gomi, K., Juvvadi, P.R., Kato, M., Kitamoto, K., Takeuchi, M., and Machida, M. (2007). Genomics of *Aspergillus oryzae*. Biosci. Biotechnol. Biochem. *71*, 646–670.
14. Machida, M., Yamada, O., and Gomi, K. (2008). Genomics of *Aspergillus oryzae*: learning from the history of Koji mold and exploration of its future. DNA Res. *15*, 173–183.
15. Teramoto, Y., Hano, T., and Ueda, S. (2000). Production and characteristics of an ancient form of sake made with shitogi. J. Inst. Brewing *106*, 95–99.
16. Yoshizawa, K. (1999). Sake: Production and flavor. Food Rev. Int. *15*, 83–107.
17. Geiser, D.M., Pitt, J.I., and Taylor, J.W. (1998). Cryptic speciation and recombination in the aflatoxin-producing fungus *Aspergillus flavus*. Proc. Natl. Acad. Sci. USA *95*, 388–393.
18. Kurtzman, C.P., Smiley, M.J., Robnett, C.J., and Wicklow, D.T. (1986). DNA relatedness among wild and domesticated species in the *Aspergillus flavus* group. Mycologia *78*, 955–959.
19. Rokas, A., Payne, G., Fedorova, N.D., Baker, S.E., Machida, M., Yu, J., Georgianna, D.R., Dean, R.A., Bhatnagar, D., Cleveland, T.E., et al. (2007). What can comparative genomics tell us about species concepts in the genus *Aspergillus?* Stud. Mycol. *59*, 11–17.
20. Murakami, H., Takase, S., and Ishii, T. (1967). Non-productivity of aflatoxin by Japanese industrial strains of *Aspergillus*. J. Gen. Appl. Microbiol. *13*, 323–334.
21. Hunter, A.J., Jin, B., and Kelly, J.M. (2011). Independent duplications of α-amylase in different strains of *Aspergillus oryzae*. Fungal Genet. Biol. *48*, 438–444.
22. Kato, N., Tokuoka, M., Shinohara, Y., Kawatani, M., Uramoto, M., Seshime, Y., Fujii, I., Kitamoto, K., Takahashi, T., Takahashi, S., et al. (2011). Genetic safeguard against mycotoxin cyclopiazonic acid production in *Aspergillus oryzae*. ChemBioChem *12*, 1376–1382.
23. Payne, G.A., Nierman, W.C., Wortman, J.R., Pritchard, B.L., Brown, D., Dean, R.A., Bhatnagar, D., Cleveland, T.E., Machida, M., and Yu, J. (2006). Whole genome comparison of *Aspergillus flavus* and *A. oryzae*. Med. Mycol. *44*, S9–S11.
24. Chang, P.K., and Ehrlich, K.C. (2010). What does genetic diversity of *Aspergillus flavus* tell us about *Aspergillus oryzae?* Int. J. Food Microbiol. *138*, 189–199.
25. Geiser, D.M., Dorner, J.W., Horn, B.W., and Taylor, J.W. (2000). The phylogenetics of mycotoxin and sclerotium production in *Aspergillus flavus* and *Aspergillus oryzae*. Fungal Genet. Biol. *31*, 169–179.
26. Sabeti, P.C., Schaffner, S.F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T.S., Altshuler, D., and Lander, E.S. (2006). Positive natural selection in the human lineage. Science *312*, 1614–1620.
27. Smith, J.M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. Genet. Res. *23*, 23–35.
28. Nicholson, M.J., Koulman, A., Monahan, B.J., Pritchard, B.L., Payne, G.A., and Scott, B. (2009). Identification of two aflatrem biosynthesis gene loci in *Aspergillus flavus* and metabolic engineering of *Penicillium paxilli* to elucidate their function. Appl. Environ. Microbiol. *75*, 7469–7481.
29. Fujisawa, K., and Yoshino, M. (1998). Formation of inosinic acid as the taste compound in the fermentation of Japanese sake. Dev. Food Sci. *40*, 227–231.
30. Mukherjee, M., Horwitz, B.A., Sherkhane, P.D., Hadar, R., and Mukherjee, P.K. (2006). A secondary metabolite biosynthesis cluster in *Trichoderma virens*: evidence from analysis of genes underexpressed in a mutant defective in morphogenesis and antibiotic production. Curr. Genet. *50*, 193–202.
31. Mukherjee, P.K., Horwitz, B.A., and Kenerley, C.M. (2012). Secondary metabolism in *Trichoderma*—a genomic perspective. Microbiology *158*, 35–45.
32. Janssens, L., De Pooter, H.L., Schamp, N.M., and Vandamme, E.J. (1992). Production of flavors by microorganisms. Process Biochem. *27*, 195–215.
33. Gomi, K., Akeno, T., Minetoki, T., Ozeki, K., Kumagai, C., Okazaki, N., and Iimura, Y. (2000). Molecular cloning and characterization of a transcriptional activator gene, *amyR*, involved in the amylolytic gene expression in *Aspergillus oryzae*. Biosci. Biotechnol. Biochem. *64*, 816–827.
34. Kitamoto, K. (2002). Molecular biology of the Koji molds. Adv. Appl. Microbiol. *51*, 129–153.
35. Khaldi, N., Seifuddin, F.T., Turner, G., Haft, D., Nierman, W.C., Wolfe, K.H., and Fedorova, N.D. (2010). SMURF: Genomic mapping of fungal secondary metabolite clusters. Fungal Genet. Biol. *47*, 736–741.
36. Rank, C., Klejnstrup, M., Petersen, L., Kildgaard, S., Frisvad, J., Gotfredsen, C., and Larsen, T. (2012). Comparative chemistry of *Aspergillus oryzae* (RIB40) and *A. flavus* (NRRL 3357). Metabolites *2*, 36–56.
37. Tominaga, M., Lee, Y.H., Hayashi, R., Suzuki, Y., Yamada, O., Sakamoto, K., Gotoh, K., and Akita, O. (2006). Molecular analysis of an inactive aflatoxin biosynthesis gene cluster in *Aspergillus oryzae* RIB strains. Appl. Environ. Microbiol. *72*, 484–490.
38. Chang, P.K., Horn, B.W., and Dorner, J.W. (2005). Sequence breakpoints in the aflatoxin biosynthesis gene cluster and flanking regions in nonaflatoxigenic *Aspergillus flavus* isolates. Fungal Genet. Biol. *42*, 914–923.

39. Keller-Seitz, M.U., Certa, U., Sengstag, C., Würgler, F.E., Sun, M.Z., and Fasullo, M. (2004). Transcriptional response of yeast to aflatoxin B1: recombinational repair involving RAD51 and RAD1. Mol. Biol. Cell *15*, 4321–4336.
40. Amaike, S., and Keller, N.P. (2011). Aspergillus flavus. Annu. Rev. Phytopathol. *49*, 107–133.
41. Georgianna, D.R., Fedorova, N.D., Burroughs, J.L., Dolezal, A.L., Bok, J.W., Horowitz-Brown, S., Woloshuk, C.P., Yu, J.J., Keller, N.P., and Payne, G.A. (2010). Beyond aflatoxin: four distinct expression patterns and functional roles associated with *Aspergillus flavus* secondary metabolism gene clusters. Mol. Plant Pathol. *11*, 213–226.
42. Kusumoto, K., Nogata, Y., and Ohta, H. (2000). Directed deletions in the aflatoxin biosynthesis gene homolog cluster of *Aspergillus oryzae*. Curr. Genet. *37*, 104–111.
43. Ito, K., Yoshida, K., Ishikawa, T., and Kobayashi, S. (1990). Volatile compounds produced by the fungus *Aspergillus oryzae* in rice koji and their changes during cultivation. J. Ferment. Bioeng. *70*, 169–172.
44. Yoshizaki, Y., Yamato, H., Takamine, K., Tamaki, H., Ito, K., and Sameshima, Y. (2010). Analysis of volatile compounds in Shochu Koji, sake Koji, and steamed rice by gas chromatography-mass spectrometry. J. Inst. Brewing *116*, 49–55.
45. Scott, D. (1989). Specialty enzymes and products for the food industry. In Biocatalysis in Agricultural Biotechnology (ACS Symposium Series 389), J.R. Whitaker and P.E. Sonnet, eds. (Washington, DC: American Chemical Society), pp. 176–192.
46. Friedman, M. (2003). Chemistry, biochemistry, and safety of acrylamide. A review. J. Agric. Food Chem. *51*, 4504–4526.
47. Aharoni, A., Giri, A.P., Verstappen, F.W.A., Bertea, C.M., Sevenier, R., Sun, Z.K., Jongsma, M.A., Schwab, W., and Bouwmeester, H.J. (2004). Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. Plant Cell *16*, 3110–3131.
48. Bachmann, H., Starrenburg, M.J.C., Molenaar, D., Kleerebezem, M., and van Hylckama Vlieg, J.E. (2012). Microbial domestication signatures of *Lactococcus lactis* can be reproduced by experimental evolution. Genome Res. *22*, 115–124.
49. Kelly, W.J., Ward, L.J.H., and Leahy, S.C. (2010). Chromosomal diversity in *Lactococcus lactis* and the origin of dairy starter cultures. Genome Biol. Evol. *2*, 729–744.
50. Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., Pavlov, A., Pavlova, N., Karamychev, V., Polouchine, N., et al. (2006). Comparative genomics of the lactic acid bacteria. Proc. Natl. Acad. Sci. USA *103*, 15611–15616.
51. Makarova, K.S., and Koonin, E.V. (2007). Evolutionary genomics of lactic acid bacteria. J. Bacteriol. *189*, 1199–1208.
52. Borneman, A.R., Desany, B.A., Riches, D., Affourtit, J.P., Forgan, A.H., Pretorius, I.S., Egholm, M., and Chambers, P.J. (2011). Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. PLoS Genet. *7*, e1001287.
53. Douglas, G.L., and Klaenhammer, T.R. (2010). Genomic evolution of domesticated microorganisms. Annu. Rev. Food Sci. Technol. *1*, 397–414.
54. Hittinger, C.T., Gonçalves, P., Sampaio, J.P., Dover, J., Johnston, M., and Rokas, A. (2010). Remarkably ancient balanced polymorphisms in a multi-locus gene network. Nature *464*, 54–58.
55. Gibbons, J.G., Janson, E.M., Hittinger, C.T., Johnston, M., Abbot, P., and Rokas, A. (2009). Benchmarking next-generation transcriptome sequencing for functional and evolutionary genomics. Mol. Biol. Evol. *26*, 2731–2744.
56. Li, H., Ruan, J., and Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. *18*, 1851–1858.
57. Vilella, A.J., Blanco-Garcia, A., Hutter, S., and Rozas, J. (2005). VariScan: Analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. Bioinformatics *21*, 2791–2793.
58. Jiang, H., and Wong, W.H. (2009). Statistical inferences for isoform expression in RNA-Seq. Bioinformatics *25*, 1026–1032.
59. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.