

Genome Comparisons between *A. flavus* and *A. oryzae* Reveal Unique Genes



Beth L. Pritchard¹, Jiujiang Yu², Jennifer Wortman³, Natalie Fedorova³, Masayuki Machida⁴, Katsuya Gomi⁵, Keietsu Abe⁵, Ralph A. Dean¹, Deepak Bhatnagar², Thomas E. Cleveland², William Nierman³, Doug E. Brown¹, and Gary A. Payne¹

¹ Center for Integrated Fungal Research, North Carolina State University, ² USDA/ARS/SRRC, ³ The Institute for Genomic Research, ⁴ Applied Gene Technology Research Group, National Institute of Advanced Industrial Science and Technology (AIST), ⁵ The Graduate School of Agricultural Science, Tohoku University



Abstract

Aspergillus flavus and *A. oryzae* are closely related fungi that inhabit very different ecological niches. *A. flavus* is a plant and animal pathogen that also produces the toxic and carcinogenic secondary metabolite, aflatoxin. In contrast, *A. oryzae* is the major fungus used in food fermentation and is generally regarded as safe (GRAS). Whole genome sequences are now available for both of these fungi, which allows for a careful comparison of the two species. Overall, these two fungi are very similar in genome size, gene organization and nucleotide identity. However analysis of genes involved in secondary metabolism revealed that there are subtle differences between the two organisms. *A. flavus* contains 35 PKS genes, 24, NRPS and 122 P450s in comparison to 31 PKS, 24 NRPS and 151 P450s in *A. oryzae*. The *A. oryzae* genome has been assembled into chromosomes using optical mapping, the high correspondence between the two genomes allowed alignment of the *A. flavus* genomic scaffolds to the *A. oryzae* chromosomes. The 16 largest genomic scaffolds from *A. flavus* essentially correspond to the 16 arms of the 8 predicted chromosomes for *A. oryzae*. However our initial studies show small differences in genome organization due to small-scale insertion-deletion events and transversions as well as evidence of a translocation event in *A. flavus* between chromosomes II and VI. The translocation break sites and many of the indels are associated with families of uncharacterized repeat elements. Analysis of these repeat elements is ongoing, but amongst those studied to date three putative types of transposable elements have been predicted. These transposable elements seem to be present in larger copy numbers in *A. oryzae* than in *A. flavus*. Interestingly, each species has approximately 350 genes unique to that species. Most of the genes are of unknown function, but within this group are genes for secondary metabolism, including polyketide synthases and non-ribosomal peptide synthases. Examination of several of these unique PKS and NRPS genes shows they are commonly associated with regions of the genome where indels have occurred. This may indicate that there is evolutionary pressure on genes involved in secondary metabolism due to the different ecologies of these fungi. Further analysis of the unique gene sets and comparison of their expression profile will help to identify genes responsible for aflatoxin production and pathogenicity.



Figure 2. Translocation event.

Aligning the *A. flavus* scaffolds to the *A. oryzae* chromosomes revealed the occurrence of a putative translocation event involving ~1MB of sequence from chromosomes II and VI. This translocation event has been confirmed experimentally using PCR and is not due to an error in the assembly of the *A. flavus* genome. The region where the translocation break point appears is shown here. This figure is a view taken from the *A. flavus* genome browser available at <http://www.aspergillusflavus.org>. It shows the genomic scaffold from *A. flavus* that maps to two different chromosomes in *A. oryzae*. The break point is highlighted with a red line. The data track in the browser called "Correspondence to *A. oryzae*" displays matches between the *A. flavus* genome and *A. oryzae* genomic scaffolds. To the left of the red line the best match between the *A. flavus* scaffold and the *A. oryzae* genome is to scaffold SC038 but to the right of the line the best match is to scaffold SC001. Optical mapping of the *A. oryzae* genome has assigned these two scaffolds to different chromosomes indicating that this is a major rearrangement between the two closely related fungi. Interestingly there is evidence of a region of repetitive DNA associated with the translocation break point (circled in red).

Unique Genes

Comparative genomic analysis between the two species reveals that despite the high level of similarity between the two genomes, both contain genes that are not shared by the other species. In this case we have classed a gene as being unique to one species if we can not construct a local alignment of that gene to any region in the other genome. This approach identifies 319 genes that are unique to *A. flavus* and 490 genes that are unique to *A. oryzae*. The majority of these genes are of unknown function, but within this group are polyketide synthases and non-ribosomal peptide synthases. These unique genes are distributed throughout the genome (Figure 4), but are often associated with regions where indels have occurred. This may indicate that there is evolutionary pressure on genes involved in secondary metabolism due to the different ecologies of these fungi. Analysis of these unique gene sets and their expression profiles should help to reveal the differences between the two fungi.

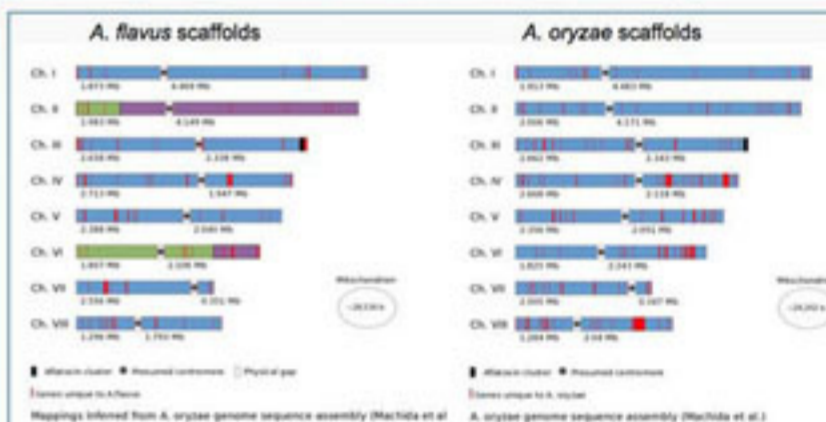


Figure 4. Distribution of unique genes

The locations of the genes unique to each species marked in red on the appropriate genome.

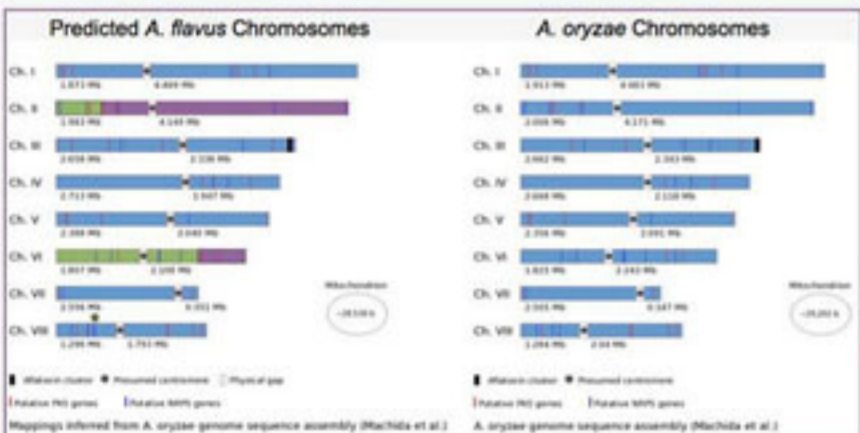


Figure 1. *A. flavus* and *A. oryzae* Chromosome Structure

Optical mapping was used to assemble the *A. oryzae* genome into chromosomes (Machida et al). The high degree of DNA similarity between the two genomes allowed for putative *A. flavus* chromosomes to be predicted based on alignment of the *A. flavus* genomic scaffolds to the *A. oryzae* chromosomes. Essentially the 16 largest scaffolds from *A. flavus* correspond to the 16 chromosome arms. There is evidence for one large translocation event in *A. flavus* between chromosomes II and VI, shown here in green and purple. There is also small scale local reorganization of the two genomes, mainly through insertion-deletion events and transversions. Analysis of the locations of genes involved in secondary metabolism shows that gene locations are largely conserved, however there are differences, for example there is a NRPS gene on Ch. VII in *A. flavus* (marked with a star) that is absent at the same locus in *A. oryzae*.



Figure 3. Small scale rearrangements - Indels.

Throughout both genome sequences we see evidence for local small scale rearrangements. In most case these are small insertion or deletion events where we see expansion in one genome when compared to the same locus in the other genome. This example shows the NRPS gene identified in Figure 1 that is present in *A. flavus* but not *A. oryzae*. This NRPS gene is unique to *A. flavus*. The gap in the correspondence track (red arrow) indicates that there is no good blast match between this region of the *A. flavus* genome and any part of the *A. oryzae* genome. Again we see a group of repetitive sequences associated with the break in correspondence between the two genomes.

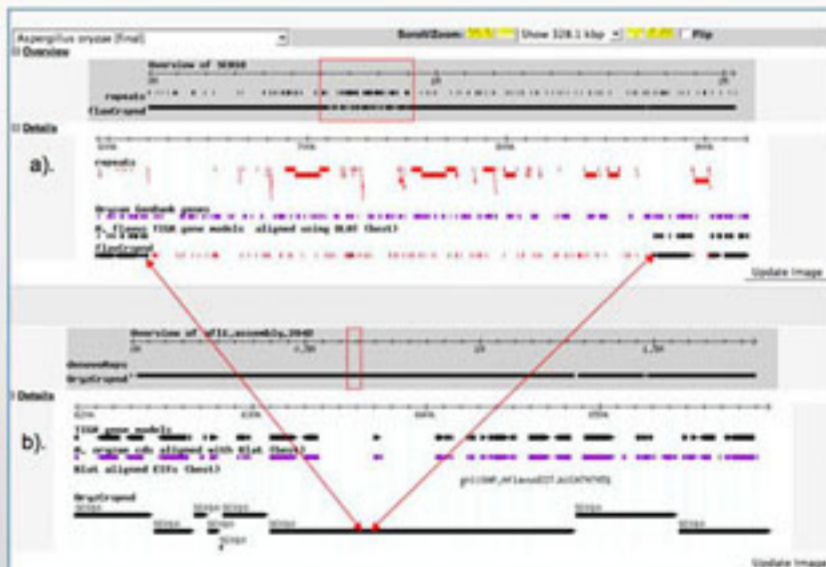


Figure 5. Analysis of a genomic locus with a high concentration of unique genes

Visualisation of the locations of unique genes in the two genomes revealed a large area on chromosome VIII of *A. oryzae* that contained genes not present in *A. flavus*. Examination of this region in the genome browser for *A. oryzae* (a) revealed an expansion of ~250 kb relevant to *A. flavus* (b). This expanded region contains a high density of repetitive DNA families, shown in the track labeled repeats. Blast analysis of these repeat sequences indicates that they belong to putative transposable elements. As this expanded region contains predicted genes we hypothesize that transposable elements may be playing a significant role in the diversification of these two species following domestication of *A. oryzae*.

References

Machida M. et al. 2005. Genome sequencing and analysis of *Aspergillus oryzae*. Nature. 438(7071):1157-61.

ACKNOWLEDGEMENTS

This project was supported by the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service, grant number # 2004-35600-14172 and funds from USDA/ARS/SRRC, New Orleans, LA.