ORIGINAL ARTICLE   **OPEN ACCESS**

Study of orthologous gene clusters among *Aspergillus* species

J. H. Kahodariya*[1], H. P. Gajera[1], B. A. Golakiya[1], V. V. Kothari[1], J. V. Kheni[1], J. M. Sanghani[1], H. R. Ramani[2], N. H. Garaniya[3]

[1]Department of Biotechnology, Junagadh Agricultural University, Junagadh- 362001
[2]Main Cotton Research station, Navsari Agricultural University, Surat- 395007
[3] College of Agriculture, Navsari Agricultural University, Bharuch- 392012
*Email: jayesh8286@gmail.com

ABSTRACT

Orthologs or orthologous genes are clusters of genes in different species that originated by vertical descent from a single gene in the last common ancestor. To study the orthologous genes among different Aspergillus isolates we carried out whole genome sequencing of most aflatoxigenic A. flavus JAU2 and highly potent biocontrol atoxigenic A. flavus NRRL 21882. The protein gene FASTA file containing 14414 predicted protein sequences for AF JAU2 and 10154 for AF NRRL 21882 were used to predict the orthologous gene clusters among six species of Aspergillus viz A. oryzae, A. clavatus, A. flavus, A. niger, A. flavus JAU2 and A. flavus NRRL 21882. The analysis showed 12902 clusters based on the protein sequences from the 6 species. The numbers in the Venn diagram represent the number of orthologous clusters that AF JAU2 and AF NRRL 21882 shares with the four other species. All six species shared 5159 clusters and 210 clusters were found unique in all six species. The diagram showed that 5159 gene clusters were shared by all six species, suggesting their conservation in the lineage after speciation. The unique clusters were also found 41 in AF JAU2, two in AF NRRL 21882 and A. flavus, 115 in A. niger, 24 in A. clavatus and 21 in A. oryzae. So highest numbers of 115 unique genes were present in A. niger and only two unique gene clusters were present in AF NRRL 21882 and A. flavus. All A. flavus isolates viz. AF JAU2, AF NRRL 21882 and A. flavus shared 271 gene clusters whereas AF JAU2 and AF NRRL 21882 shared 109 orthologous genes.
Keywords*: Aspergillus, orthologous genes, whole genome, protein genes*

INTRODUCTION

Orthologs or orthologous genes are clusters of genes in different species that originated by vertical descent from a single gene in the last common ancestor [1]. Comparative analysis of the organization of orthologous clusters is important for understanding the rules of genome structure. The tools for establishing homologies between genes or their products are becoming increasingly important to transfer knowledge from well-studied model organisms to other organisms [2]. One of the simplest but most useful methods of genome wide orthologous comparison is to display the different and overlapping orthologous clusters in a Venn diagram, which in our case provides circles or other shapes representing each species with overlapping regions that illustrate the genes or gene clusters that are unique to or shared between each species. Genome wide analysis of orthologous clusters is an important component of comparative genomics studies. Identifying the overlap among orthologous clusters can enable us to elucidate the function and evolution of proteins across multiple species [3].

Recently a web platform named OrthoVenn (http: //probes.pw.usda.gov/OrthoVenn or http://aegilops.wheat. ucdavis.edu/OrthoVenn) is used for the comparison and analysis of genome wide orthologous clusters across multiple species. In OrthoVenn, users can select protein sequence data for genome wide comparisons from 272 species in the database, including vertebrates, metazoa, protists, fungi, plants and bacteria. OrthoVenn also allows user-defined species to be uploaded as customized protein sequences. An efficient and interactive graphics tool is employed to provide a Venn diagram view of the genome wide comparison of orthologous clusters based on the protein sequence data selected from up to six species. The intersection of orthologous clusters is analyzed by GO Slim annotation and UniProt search. In the output of OrthoVenn, each orthologous cluster provides sequence analysis data, single copy

gene cluster identification, protein similarity comparisons, and the phylogenetic relationships among clustered genes. OrthoVenn also provides key word search and BLAST functions for finding clusters of specific interest to the user. In addition, OrthoVenn allows the user to create Venn diagrams from orthologous cluster files generated by other software. Genome wide analysis of orthologous clusters is an important component of comparative genomics studies. Identifying the overlap among orthologous clusters can enable us to elucidate the function and evolution of proteins across multiple species. OrthoVenn provides coverage of vertebrates, metazoa, protists, fungi, plants and bacteria for the comparison of orthologous clusters and also supports uploading of customized protein sequences from userdefined species. An interactive Venn diagram, summary counts, and functional summaries of the disjunction and intersection of clusters shared between species are displayed as part of the OrthoVenn result. OrthoVenn is an efficient and user-friendly web server freely accessible at http: //probes.pw.usda.gov/OrthoVenn or http://aegilops. wheat.ucdavis.edu/OrthoVenn [3].

MATERIALS AND METHODS
A total 19 isolates of microscopic fungi consisting of *Aspergillus* species were obtained from the groundnut seeds, cake and soil samples collected from farmer's field of different places of Saurastra region, Gujarat. Morphological, Biochemical and molecular characteristics of *Aspergillus* species showed AF JAU2 as highly aflatoxigenic and AF NRRL 21882 as most potent atoxigenic biocontrol isolate. The atoxigenic *Aspergillus flavus* NRRL 21882 strain was procured from The ARS culture collection (NRRL), United state department of Agriculture (USDA), Peoria (Illinois), USA. All the laboratory works have been carried out at the Department of Biotechnology, Junagadh Agricultural University, Junagadh (Gujarat).

The whole genome sequence was carried out of these two isolates using the Illumina NextSeq platform with two pair- end library of 150 bp. The trimming of contaminant or ambiguous sequences such as adapters and ambiguous nucleotides other than A, G, C and T were carried out using CLC genomic workbench v9.5.4 and Bioedit v7.2.5 followed by *de novo* genome assembly with minimum conting length 500 bp. After trimming the contaminants the genome size and number of contigs for both the genomes found to be 40.27 Mbp and 2367 respectively for AF JAU2 and 37.54 Mbp and 398 respectively for AF NRRL 21882. The whole genome sequence of these isolates were annotated using Augustus, GenMarkES and Glimmer tools by using https://www.gensas.org/ as a web server. The genome annotation results is shown in table 1.

Table 1. Genome annotation results of AF JAU2 and AF NRRL 21882.

	AF JAU2	AF NRRL 21882
CDS	42194	32384
Exon	42194	32384
mRNA	14598	10186
Gene	14598	10186
Protein	14414	10154

The final number of protein genes were 14414 for AF JAU2 and 10154 for AF NRRL 21882. So we compared the protein genes of these two isolates with other four *Aspergillus* isolates viz. *A. flavus, A. clavatus, A. niger and A. oryzae.* The protein gene sequences of other four isolates were already exists in http://www.bioinfogenome.net/OrthoVenn. The online tool OrthoVenn available at website http://www.bioinfogenome.net/OrthoVenn/ is used for developing Venn diagram. We supplied protein FASTA file to find our protein coding orthologous gene clusters among different *Aspergillus* species such as *A. oryzae, A. clavatus, A. flavus, A. niger,* AF JAU2 and AF NRRL 21882. The OrthoVenn generated Venn diagram showing numbers of orthologous gene clusters among six isolates (Fig.1 A and 1B).

RESULTS AND DISCUSSION
The protein FASTA file containing 14414 predicted protein sequences for AF JAU2 and 10154 for AF NRRL 21882 were used to predict the orthologous gene clusters among six species of *Aspergillus viz A. oryzae, A. clavatus, A. flavus, A. niger, A. flavus* JAU2 and *A. flavus* NRRL 21882. The number of protein sequences of all *Aspergillus* isolates is depicted in Fig. 1 A. There are multiple methods for orthology prediction. Here the popular heuristic approach named OrthoMCL [4] used to identify ortholog groups. The OrthoMCL performs an all-against-all BLASTP alignment, identifies putative orthology and inparalogy relationships with the Inparanoid algorithm [5] and generates disjoint clusters of closely related proteins with the Markov Clustering Algorithm (MCL) [6].
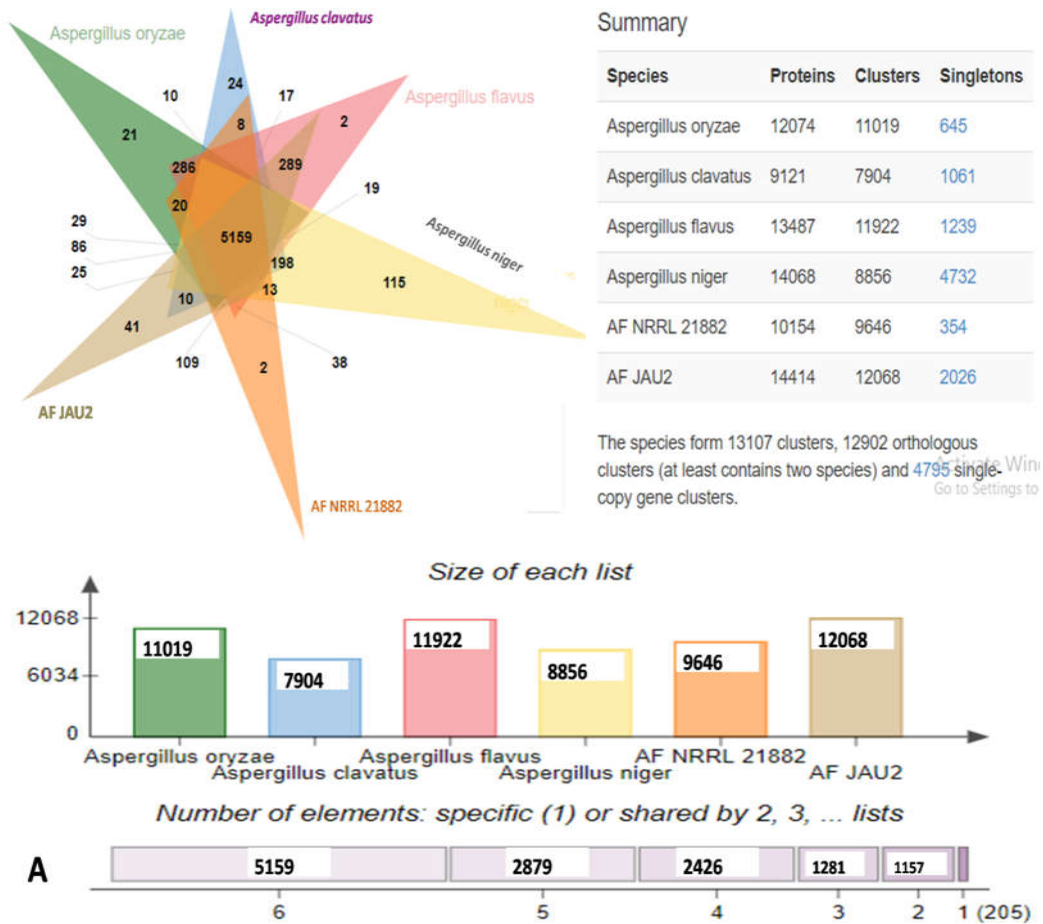
Fig. 1 A. Venn diagram showing the distribution of shared gene families (orthologous clusters) among AF JAU2, AF NRRL 21882, *A. oryzae, A. clavatus, A. flavus* and *A. niger*. The cluster number in each component is listed.
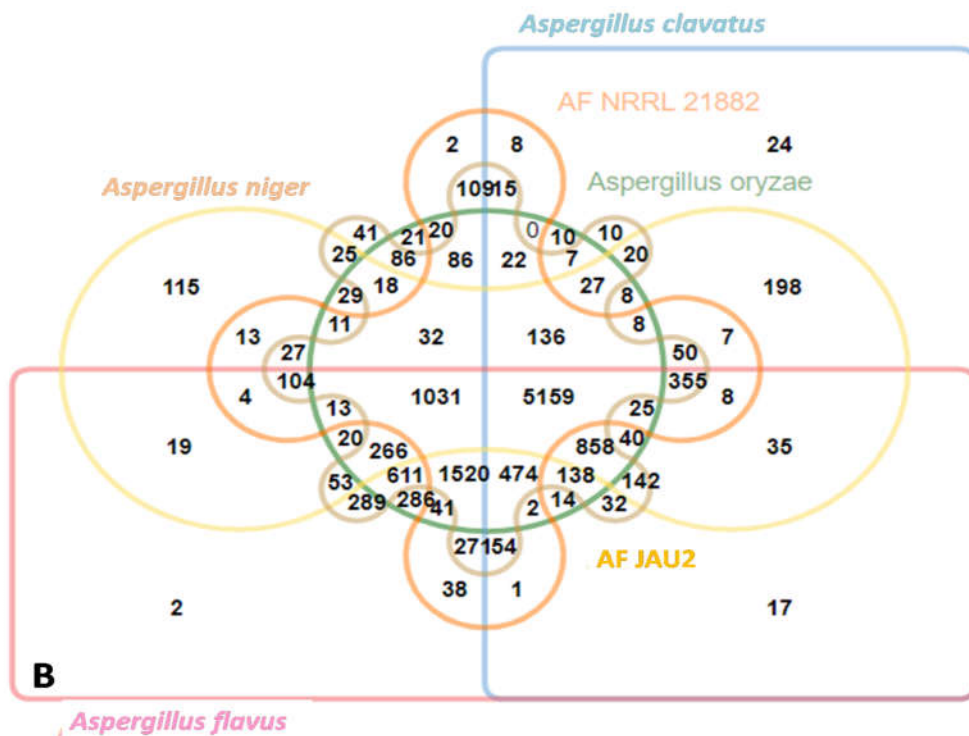

Fig. 1 B. Circular Venn diagram showing same results

The analysis showed 12902 clusters based on the protein sequences from the 6 species (Fig. 1A and 1B). The numbers in the Venn diagram represent the number of orthologous clusters that AF JAU2 *and* AF NRRL 21882 shares with the four other species. All six species shared 5690 clusters and 210 clusters were found unique in all six species. The diagram showed that 5159 gene clusters were shared by all six species, suggesting their conservation in the lineage after speciation. The unique clusters were also found 41 in AF JAU2, two in AF NRRL 21882 and *A .flavus*, 115 *in A. niger*, 24 in *A. clavatus* and 21 in *A. oryzae*. So highest numbers of 115 unique genes were present in *A. niger* and only two unique gene clusters were present in AF NRRL 21882 and *A. flavus*. These specific gene clusters are likely gene clusters within multiple genes or in- paralog clusters. The presence of in-paralog clusters suggests that there might be a lineage specific gene expansion in these gene families in *Aspergillus* species (Wang *et al.,* 2015). Draft genome sequencing *of A. rabiei* for secretome analysis in comparisons with closely related other fungal genomes. *i.e. Cochliobolus heterostrophus*, *Pyrenophora tritici-repentis* and *Stagonospora nodorum*. OrthoMCL analysis showed that 6,432 (60.7 %) of *A. rabiei* predicted proteins had orthologs in these three fungal species, while 693 (6.5 %) predicted proteins were unique [7]. All *A. flavus* isolates *viz.* AF JAU2, AF NRRL 21882 and *A. flavus* shared 271 gene clusters whereas AF JAU2 and AF NRRL 21882 shared 109 orthologous genes.

CONCLUSION
The complete analysis protein gene sequences of six *Aspergillus* isolates revealed the correlation of orthologous genes among the genus. The species form 13107 clustes, 12902 orthologous clusters (at least contains two species) AND 4795 single- copy gene coustes. These clusters are likely gene clusters within multiple genes or in-paralog clusters suggesting that there might be a lineage specific gene expansion in these gene families in AF JAU2 and AF NRRL 21882.

REFERENCES
1. Fitch, W.M. (1970). Distinguishing homologous from analogous proteins. Syst. Zool., 19: 99–113.
2. Altenhoff, A.M. & Dessimoz, C. (2012). Inferring orthology and paralogy. Methods Mol. Biol., 855: 259–279.
3. Wang, Y., Coleman-Derr, D., Chen, G. & Gu, Y. Q. (2015). OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic acid Res., 43: 78- 84.
4. Li, L., Stoeckert, C.J. Jr. & Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res., 13: 2178–2189.
5. Remm, M., Storm, C.E. & Sonnhammer, E.L. (2001) Automatic clustering of orthologs and in paralogs from pairwise species comparisons. J. Mol. Biol., 314: 1041–1052.
6. Enright, A.J., Van Dongen, S. & Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res., 30: 1575–1584.
7. Verma, S., Gazara, R., Nizam, S., Parween, S., Chattopadhyay, D. & Verma, P. (2016). Draft genome sequencing and secretome analysis of fungal phytopathogen *Ascochyta rabiei* provides insight into the necrotrophic effector repertoire. Nature Sci. Reports. 6: 24638.