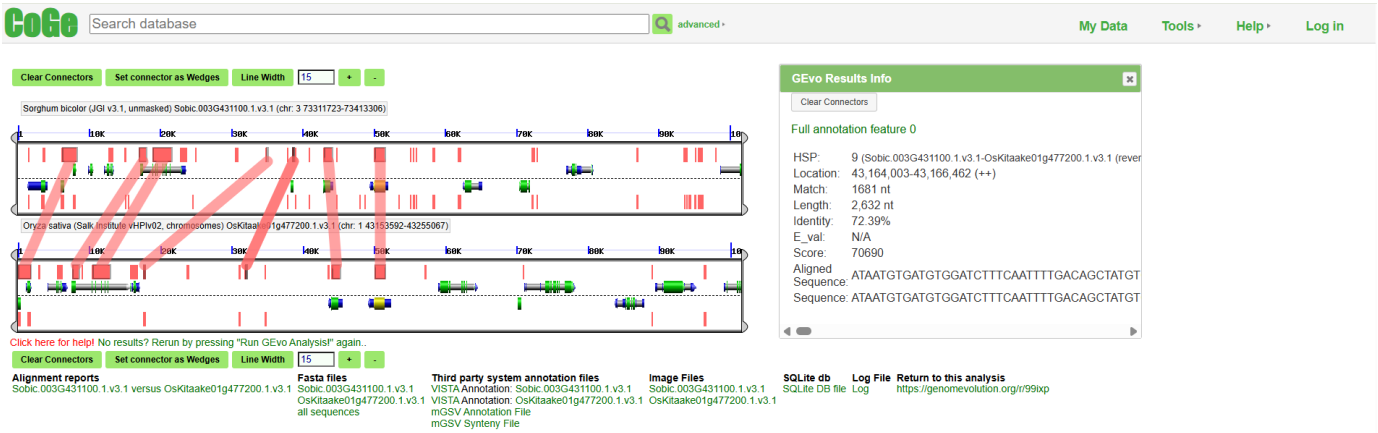
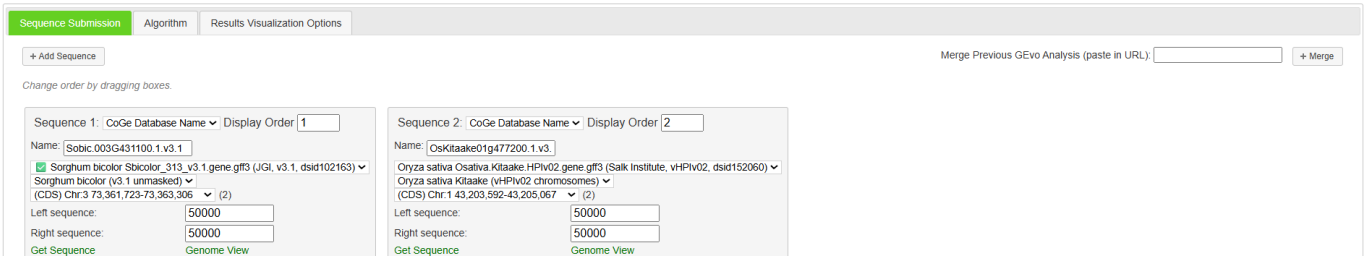
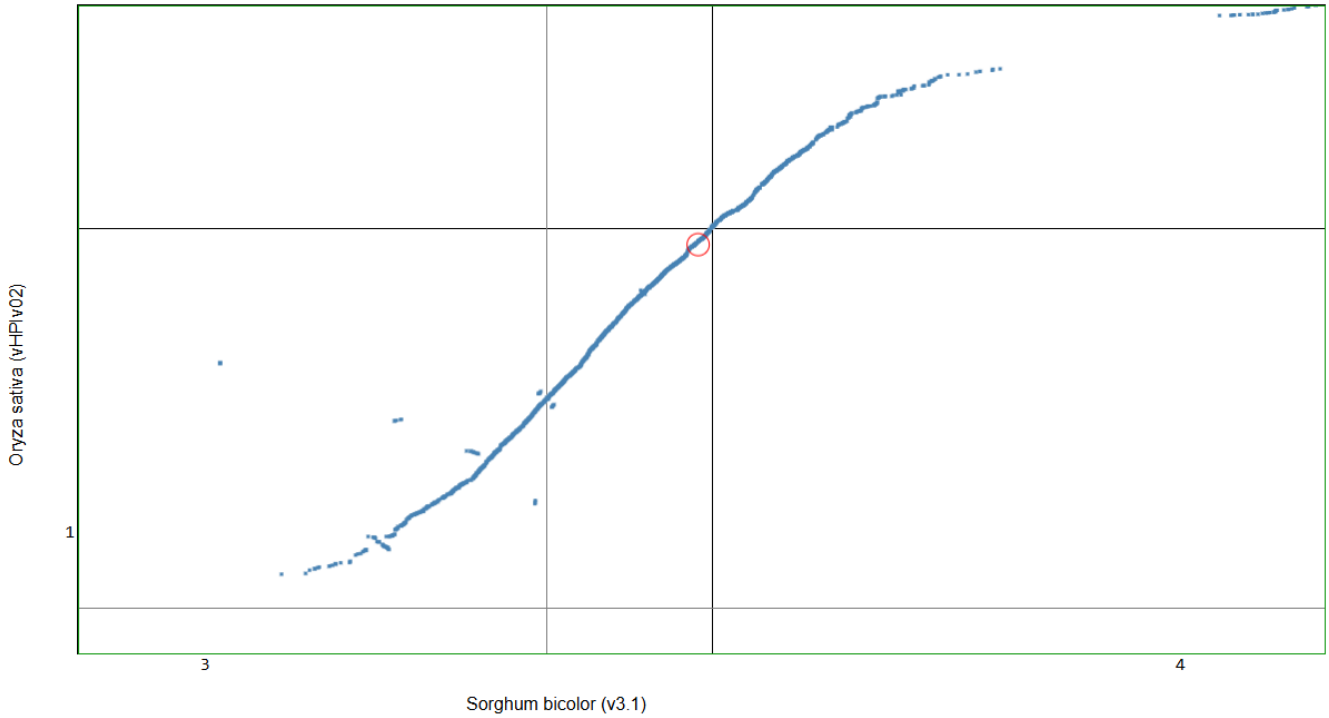
Joseph Ifeanyi Ulasi

Comparative Genomics of *Sorghum bicolor* and *Oryza sativa* (ssp. *Japonica*)







General Statistics

1. Which genomes did you choose?

- Genome 1: *Sorghum bicolor*
- Genome 2: *Oryza sativa* (ssp. *Japonica*)

2. What was your rationale?

I understanbd that sorghum and rice are both key model organisms in the grass family (*Poaceae*) but diverged from a common ancestor, hence, comparing both allows for understanding genome evolution and synteny conservation over a deeper over time, as well as identify highly conserved genomic regions that are likely under strong functional constraint.

3. Are these two genomes well assembled?

Yes, both are high-quality, chromosome-level assemblies. Sorghum bicolor v3.1: This is a reference-quality assembly where the sequence is organized into 10 chromosomes (pseudomolecules). Oryza sativa vHFM02: The rice genome was one of the first plant genomes sequenced and has been refined to a very high-quality, chromosome-level assembly.

4. How many contigs/scaffolds are there?

- Sorghum: the assembly consists of 10 scaffolds, corresponding to its 10 chromosomes.
- Rice: the assembly consists of 12 pseudomolecules, corresponding to its 12 chromosomes.

5. Are your species diploid, polyploid, or a combination of both?

- Sorghum bicolor: True diploid (2n=2x=20).
- Oryza sativa: True diploid (2n=2x=24).

6. Can you tell who sequenced that genome or what technology was used?

- Sorghum: Sequenced by the Joint Genome Institute (JGI) using a Sanger-based whole-genome shotgun approach.
- Rice: The original genome was sequenced by an international consortium using a BAC-by-BAC approach, and this version (vHFM02) from the Salk Institute represents a modern, high-quality reassembly.

Macrosynteny

Interpretation of a Macrosyntenic Dotplot (Sorghum vs. Rice)

The dotplot would have Sorghum chromosomes 1-10 on one axis and Rice chromosomes 1-12 on the other. The plot would show a complex pattern of multiple sorghum chromosomes corresponding to segments of multiple rice chromosomes and vice-versa. This is due to an ancient whole-genome duplication in the grass family ancestor, followed by extensive chromosome fission/fusion and rearrangement events in the lineages leading to modern sorghum and rice.

1. Do these two species have good macrosynteny? Is there clear evidence of polyploidy?

Yes, there is clear macrosynteny, but it is complex and reveals a shared ancient polyploidy. While there is no simple one-to-one relationship, I could see that one sorghum chromosome has syntenic blocks with 2-4 different rice chromosomes. This is evidence of the ancestral grass whole-genome duplication millions of years ago and the subsequent differential genome repackaging in the sorghum and rice lineages.

2. How similar do you think these species are? Do the syntenic dotplots match your expectations based on how close they are phylogenetically?

They are moderately related within the grass family, about 40-50 MYA divergence. The complex dotplot matches expectations perfectly. We expect to see conserved syntenic blocks broken up by numerous rearrangements over this evolutionary timescale.

3. Are there any inversions or large-scale chromosomal rearrangements?

Yes, many. The most prominent large-scale rearrangement is the well-documented fusion/fission history. A classic example is *Sorghum chromosome 3 region aligns with rice chromosome 1*, but the GEvo highlights show several inversions and gaps. Other chromosomes show synteny broken into smaller blocks, suggesting rearrangements and divergence.

4. Are these genomes well assembled? How does that affect your interpretation?

Yes, I believe that both are near chromosome-level assemblies. However, the complex pattern of ancient polyploidy and massive rearrangement would be impossible to interpret with fragmented genomes.

Microsynteny

The GEvo analysis is set up to compare a specific region around two genes:

- Sorghum: Locus Sobic.003G335100.1 on Chromosome 3 (from ~65.797 Mb to ~65.906 Mb). This is a ~109 kb region.
- Rice: Locus OsKitaka02g377000.1 on Chromosome 1 (from ~35.565 Mb to ~35.570 Mb). This is a ~5.1 kb region.

1. What chromosomal regions did you choose?

- Sorghum bicolor: A ~109 kilobase region on Chromosome 3 centered on the gene Sobic.003G335100.
- Oryza sativa: A ~5.1 kilobase region on Chromosome 1 centered on the gene OsKitaka02g377000.

2. Is there strong microsynteny? How much of the sequence or gene models in that region are conserved?

Yes, the GEvo screenshot shows several aligned genes with conserved order, although not perfect. Multiple genes are colinear, but there are also breaks and expansions. However, I think that sorghum region may contain the some set of genes as the rice region, but the intergenic spaces have been massively expanded by repetitive elements. The alignment shows ~72.39% sequence identity across aligned regions, meaning coding regions are moderately conserved, but intergenic regions diverged more.

3. Are there noticeable gaps, expansions, or rearrangements?

- Yes, the most dramatic finding is the ~20-fold expansion of the genomic region in sorghum compared to rice. This is almost certainly caused by the differential accumulation of repetitive DNA, particularly LTR retrotransposons, in the sorghum lineage after its divergence from rice. Sorghum has a larger genome than rice, and this expansion happens locally in specific regions like this one.

4. What is the sequence homology?

- **Coding Sequences (Exons):** The homology in the exons of the conserved genes (like Sobic.003G335100 and OsKitaka02g377000) would be moderately high, typically **75-85% identity** over 40-50 million years of divergence.
- **Non-Coding Sequences (Introns, Intergenic):** The homology in these regions would be very low, often falling to background levels (**<50% identity**), because they are under less evolutionary constraint and have been invaded by species-specific repetitive elements.