

Supplementary Tables and Figures

Suppl. Table S1 Comparison of genome assembly statistics of Fonio.

Parameters	CM05836 [20]	Niatia Genome
Platform	Illumina, 10X Genomics, Hi-C	Pacific Biosciences
Sequencing depth	321x	99x
Genome size, Mb	716.42	760.66
Contig number	29,115	3,329
Contig N50, kb	78	1,734
Contig L50	2,624	8
Protein coding genes	59,844	67,855
Mean gene length, bp	2,531	2,656
Mean CDS length, bp	1,055	1,224
Mean exon per gene	4.6	4.8
Mean exon length, bp	230	254
Mean intron length, bp	405	376

Suppl. Table S2. Statistics for the gene annotation

Nr of protein coding genes	67,855
Mean gene length (bp)	2668.8
Mean CDS length (bp)	1228.0
Mean nr exons per gene	4.8
Mean exon length (bp)	253.9
Mean intron length	550.6

Suppl. Table S3. Annotated non-coding RNA genes.

Target name	Count	Percentage
5_8S_rRNA	32	0.67
5S_rRNA	580	12.23
ACEA_U3	9	0.19
C4	1	0.02
ctRNA_pND324	1	0.02
enod40	6	0.13
Histone3	10	0.21
Intron_gpI	9	0.19
Intron_gpII	195	4.11
IRE_II	1	0.02
IsrR	11	0.23
LSU's	278	5.86
mir's	350	7.38
RNase_MRP	2	0.04
sno RNAs	1109	23.39
Spliceosomal snRNAs	375	7.91
SRP's	36	0.76
SSU_rRNA	249	5.25
TPP	2	0.04
tRNA	1478	31.17
tRNA-Sec	7	0.15
TOTAL	4741	

Suppl. Table S4. Orthologs for Suppression of Shattering1 genes

Species	# of SSH1 orthologues
<i>Digitaria exilis</i>	4
<i>Brachypodium distachyon</i>	2
<i>Cenchrus americanus</i>	2
<i>Hordeum vulgare</i>	1
<i>Oropetium thomaeum</i>	2
<i>Oryza sativa</i>	2
<i>Setaria italica</i>	2
<i>Sorghum bicolor</i>	3
<i>Zea mays</i>	2

Suppl. Table S5. Orthologs of Dwarf Gene-3

Species	# of dw3 orthologues
<i>Digitaria exilis</i>	2
<i>Brachypodium distachyon</i>	1
<i>Cenchrus americanus</i>	1
<i>Hordeum vulgare</i>	1
<i>Oropetium thomaeum</i>	1
<i>Oryza sativa</i>	1
<i>Setaria italica</i>	1
<i>Sorghum bicolor</i>	1
<i>Zea mays</i>	1

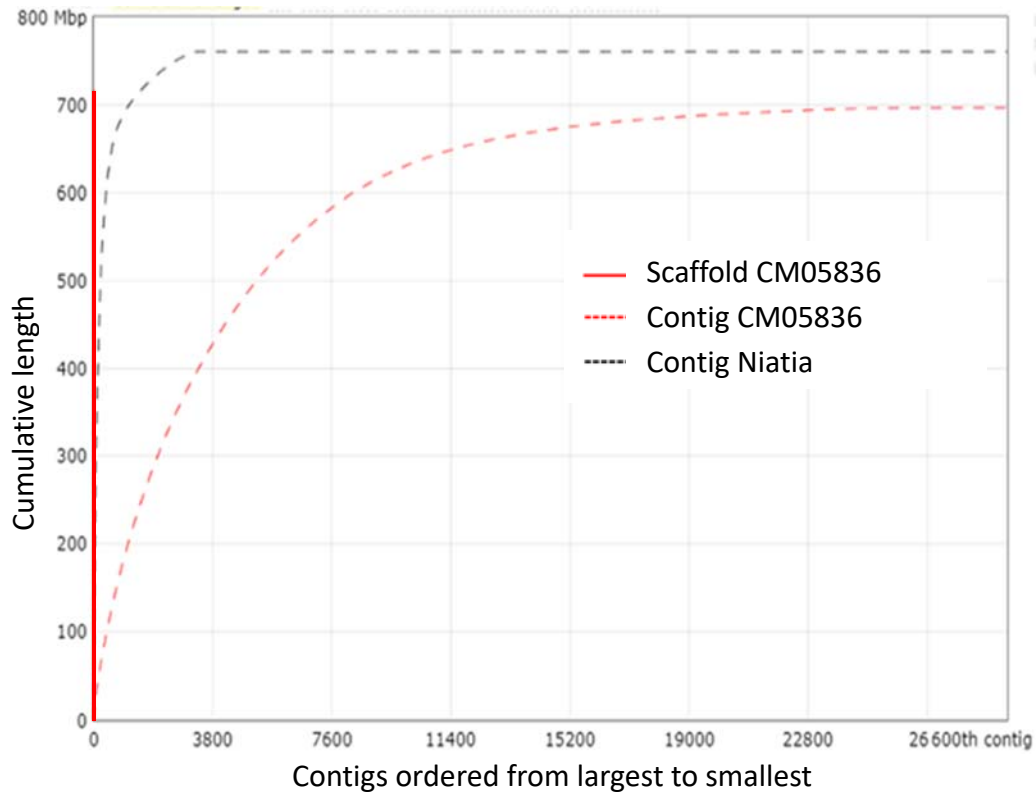
Suppl. Table S6. Orthologs of Grain Weight-2 genes

Species	# of GW2 orthologues
<i>Digitaria exilis</i>	2
<i>Brachypodium distachyon</i>	1
<i>Cenchrus americanus</i>	1
<i>Hordeum vulgare</i>	1
<i>Oropetium thomaeum</i>	1
<i>Oryza sativa</i>	1
<i>Setaria italica</i>	1
<i>Sorghum bicolor</i>	1
<i>Zea mays</i>	2

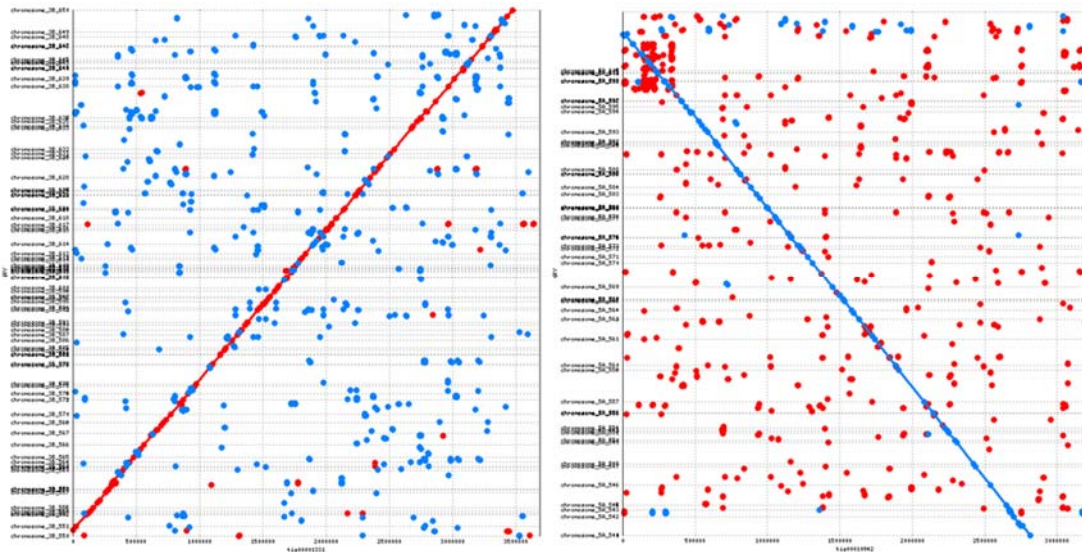
Suppl. Table S7. Passport data for accessions and samples used for diversity study (see Suppl. Tables Excel file)

Suppl. Table S8. SNP database used for diversity study (see Suppl. Tables Excel file)

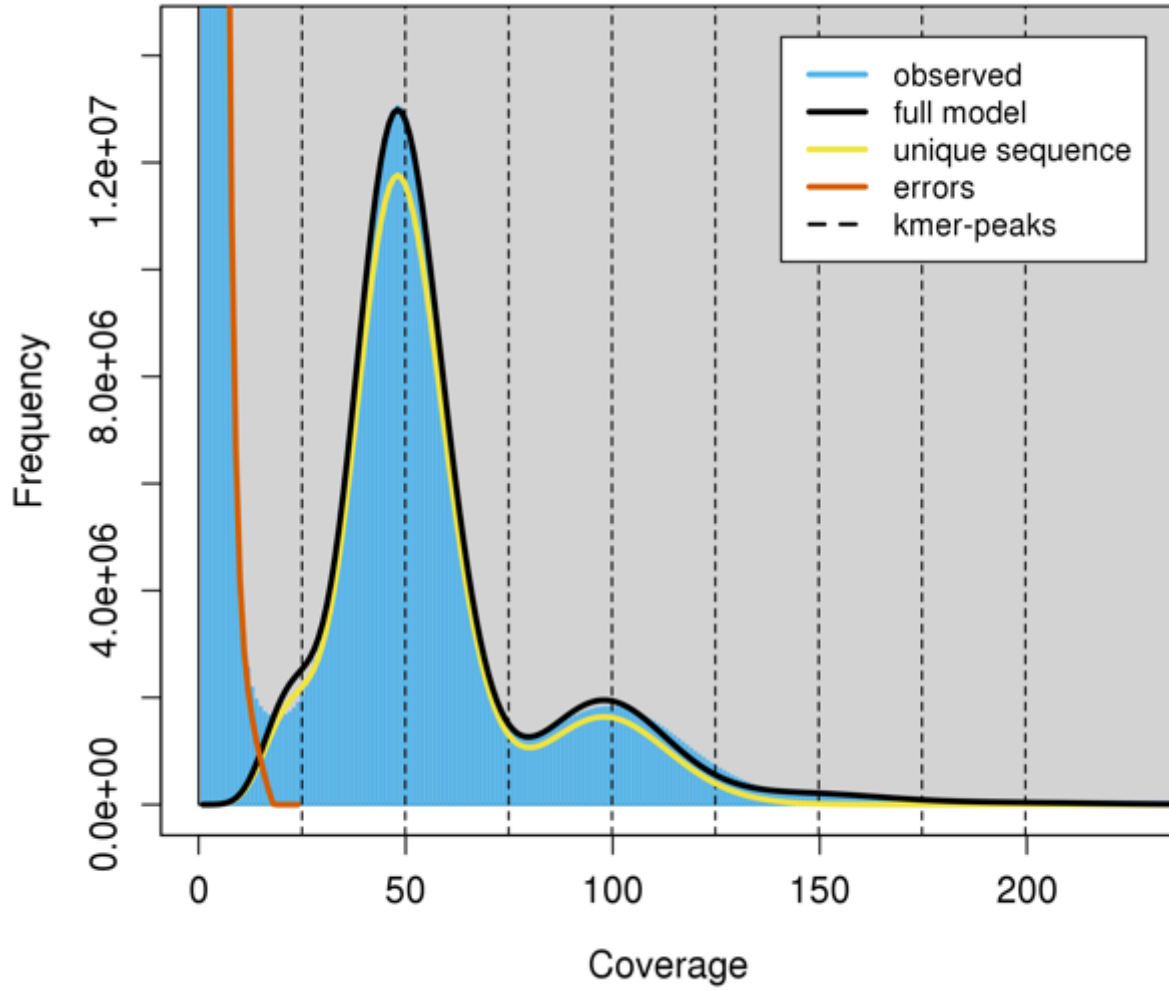
A.



B.

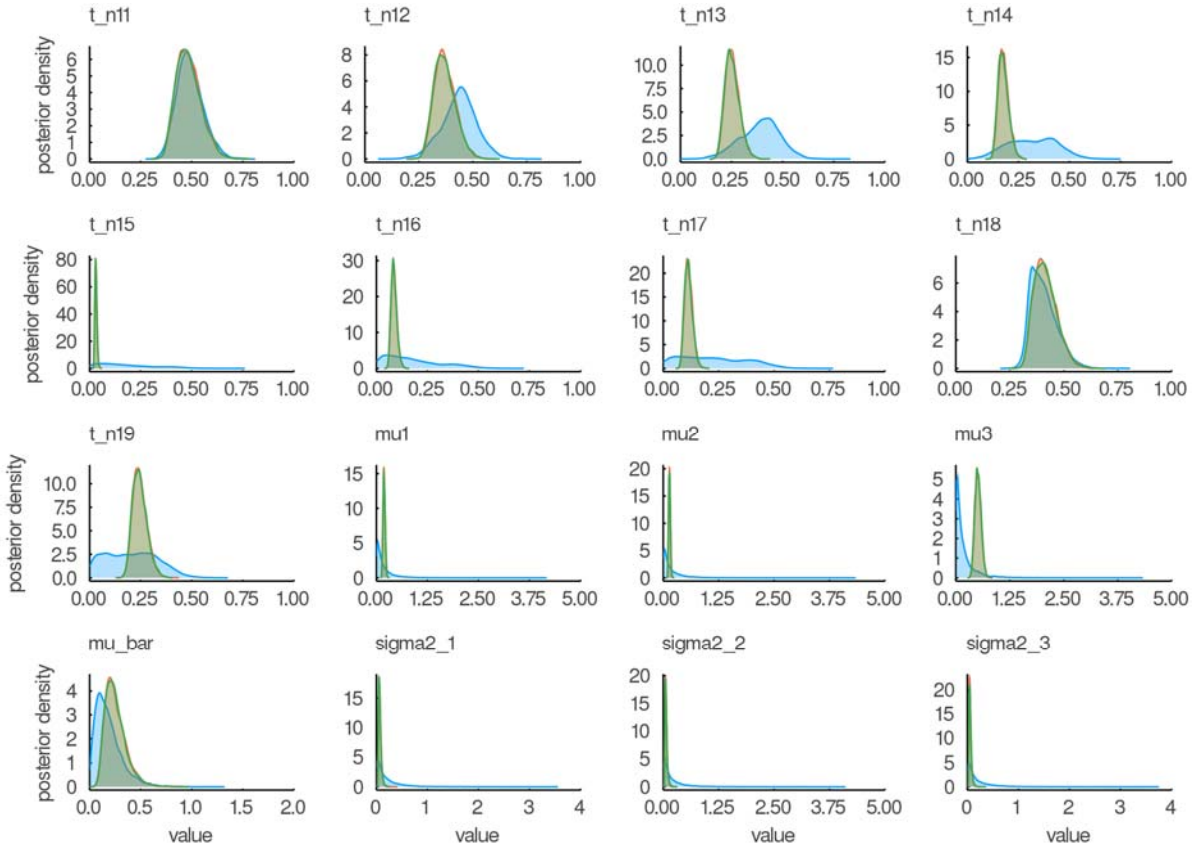


Suppl Figure S1. A. comparison of the contiguity of the Niatia Genome and CM05836 [20] genome. B. comparison of Contig tig00001331 corresponding to 100 consecutive segments anchored on the same chromosome 3B and tig00010942 corresponding to 65 consecutive segments on the chromosome 5A on the Abrouk et al. [20] genome.

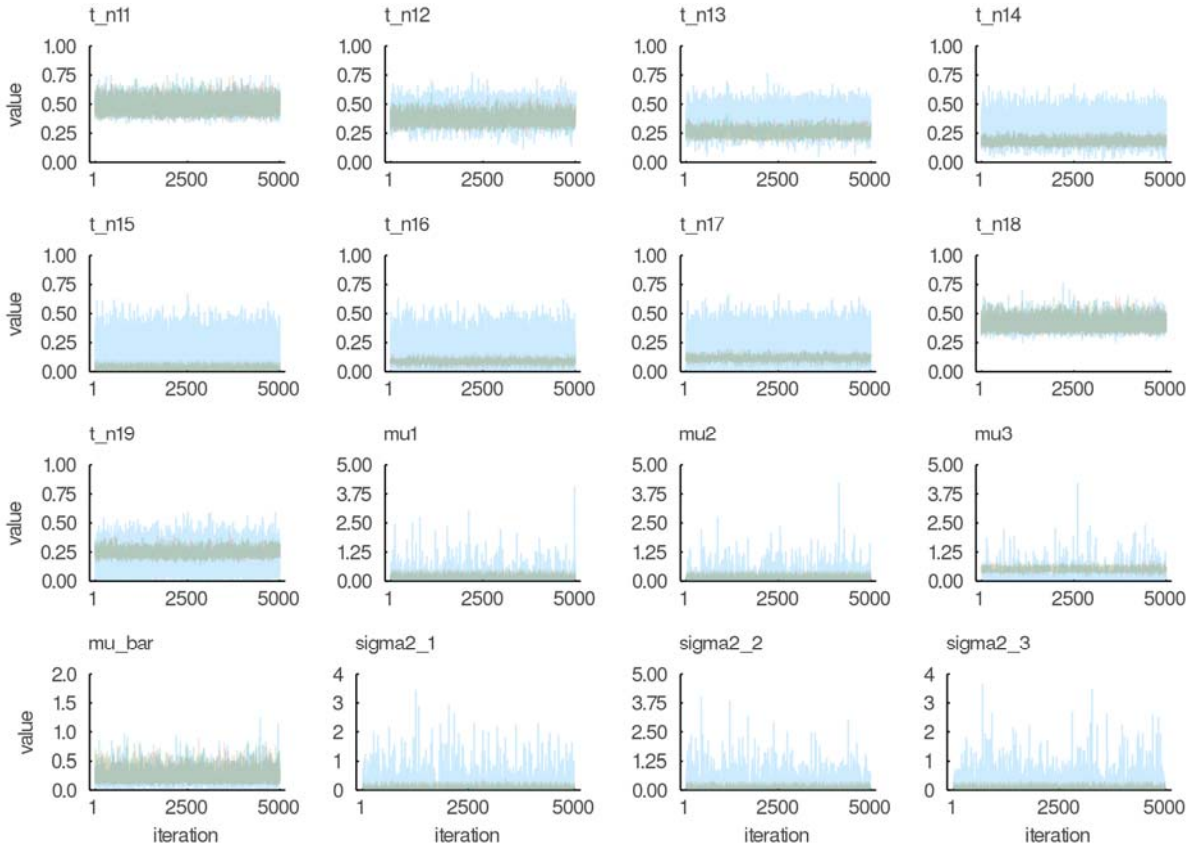


Suppl. Figure S2. The kmer distribution of raw Illumina reads at kmer value 33bp.

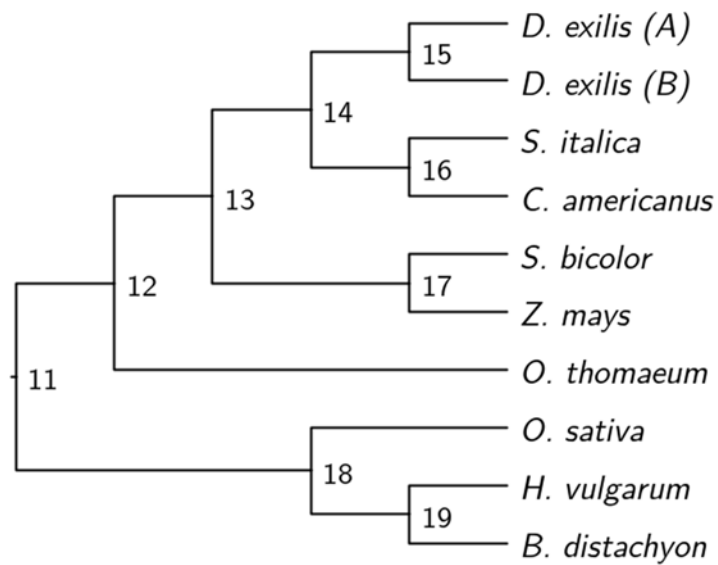
A



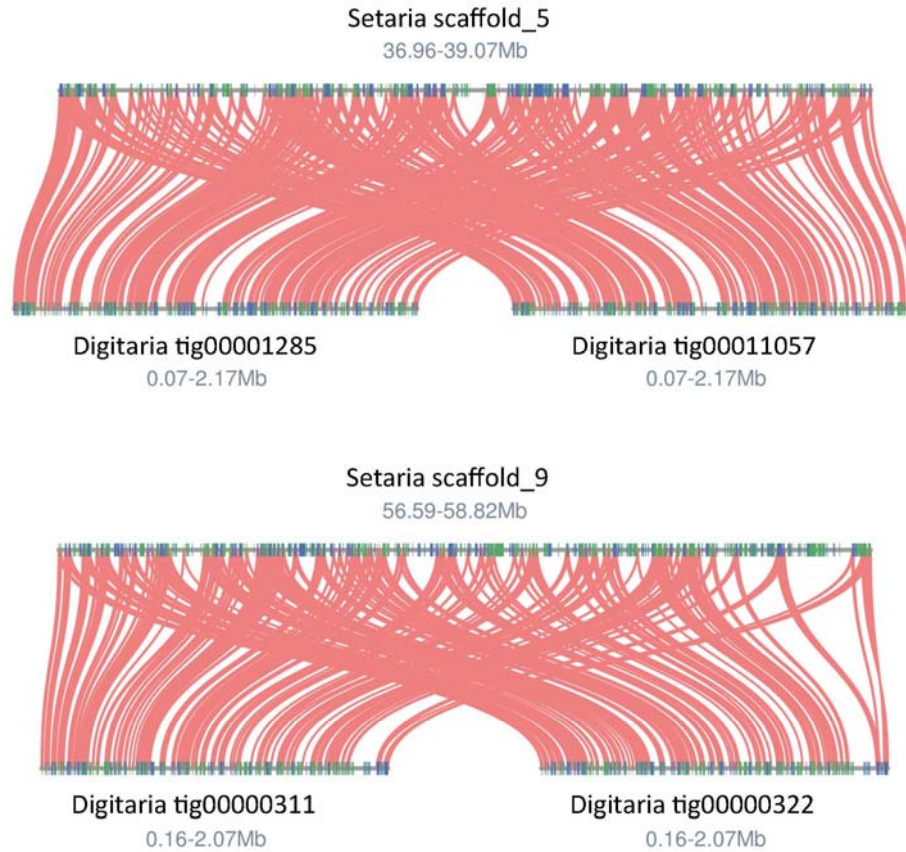
B



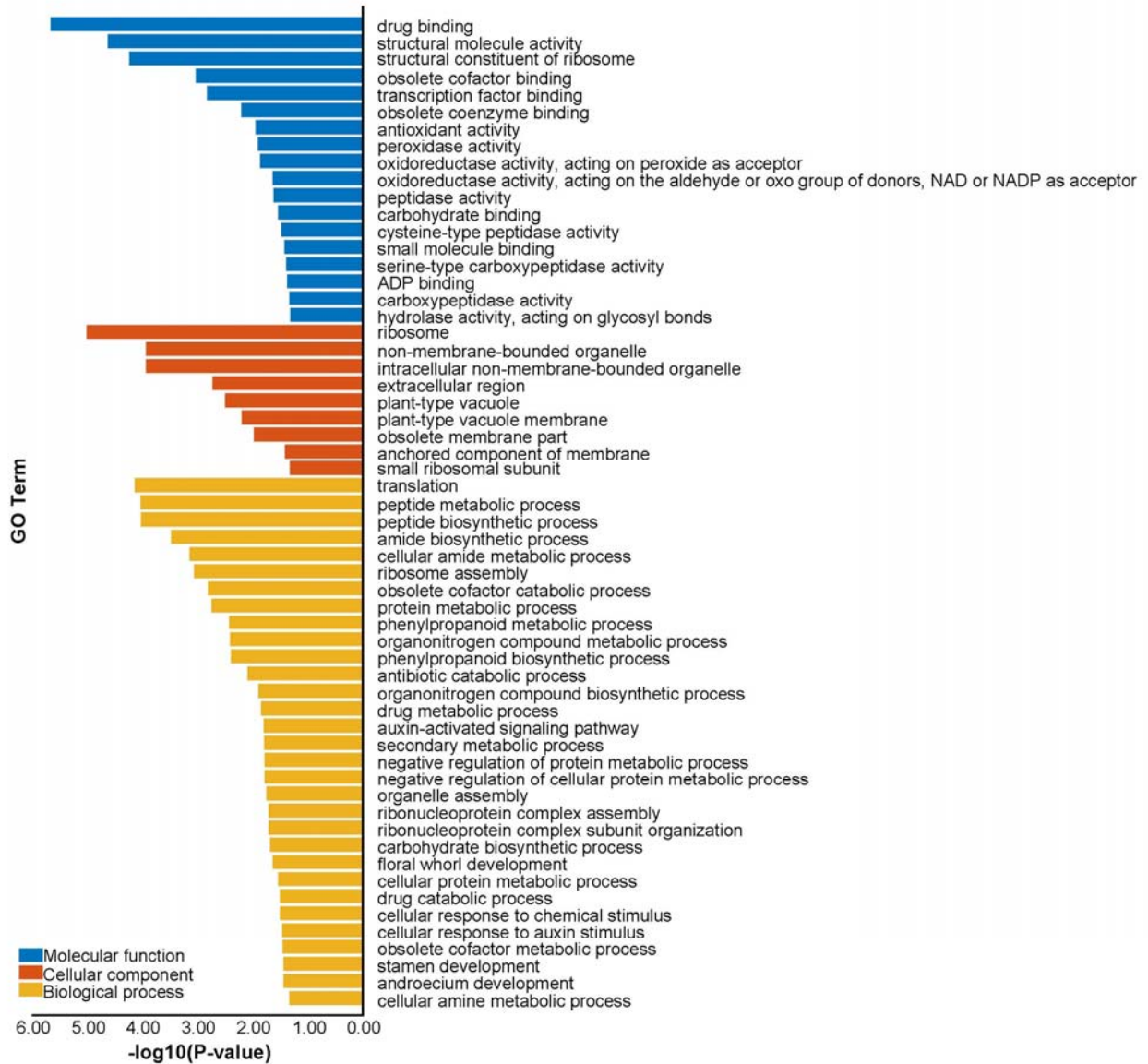
C



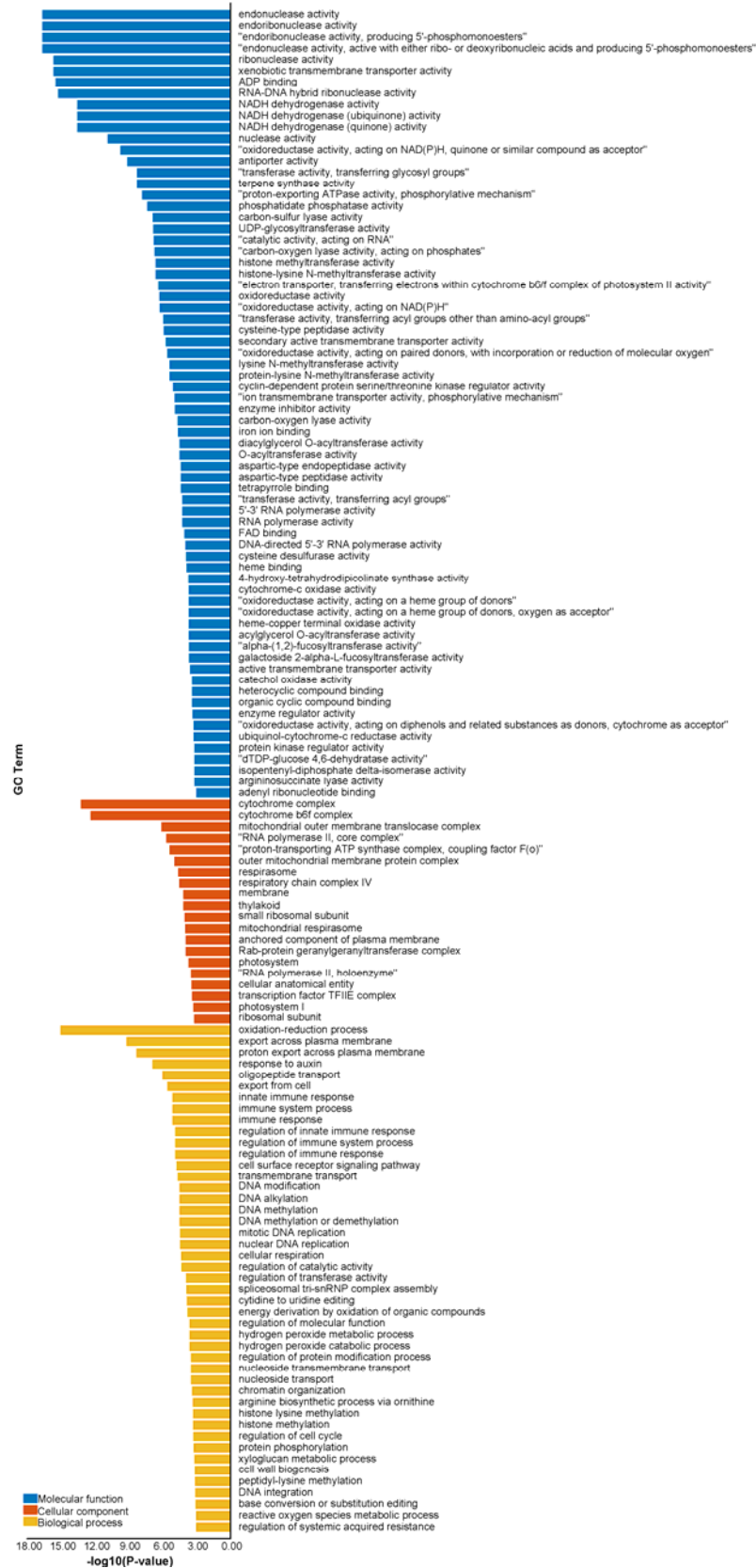
Suppl. Figure S3. A Marginal posterior distributions for two independent chains (green and orange) and induced marginal prior distributions (blue) for internal node ages (t_{n11} to t_{n19} , see panel C), overall mean substitution rate (μ), mean substitution rate for different codon positions (μ_1 , μ_2 and μ_3) and variance parameter of the uncorrelated relaxed clock (σ_1^2 , σ_2^2 and σ_3^2) for the three codon positions. **B** Trace plots for the MCMC chains associated with panel (A).



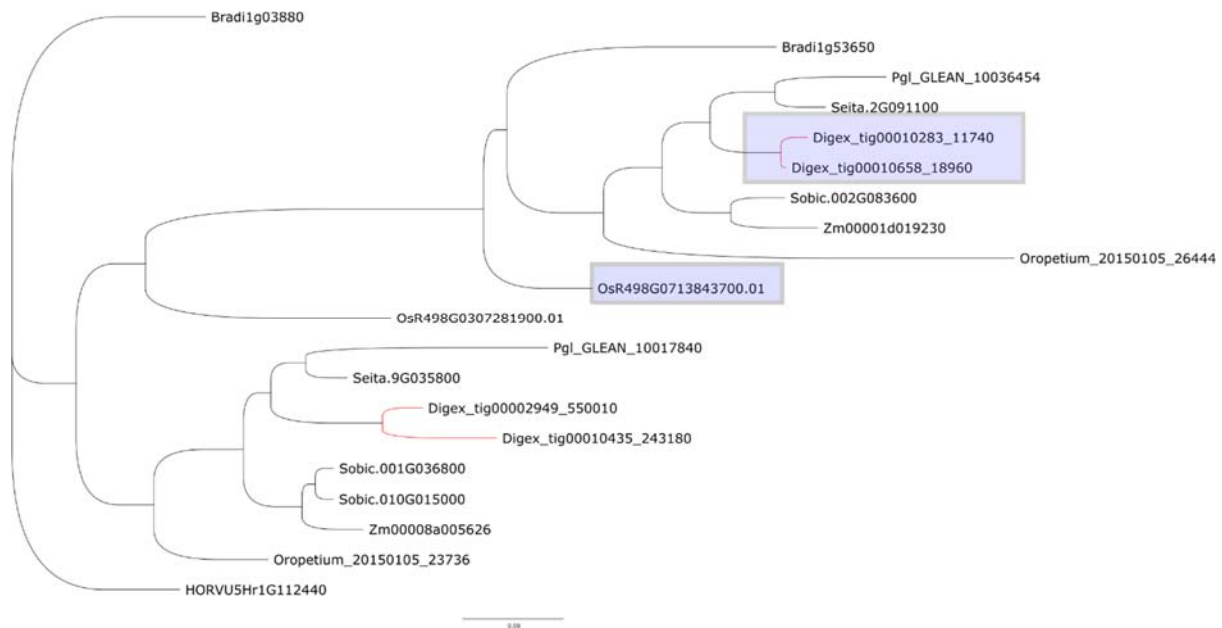
Suppl Figure S4 There are 10075 families that have two-copy in fonio and one copy in *Setaria talica* and 90% of two-copy families are located in syntenic blocks. The above four examples indicate the high degree of collinearity and synteny between *Setaria italica* and fonio.



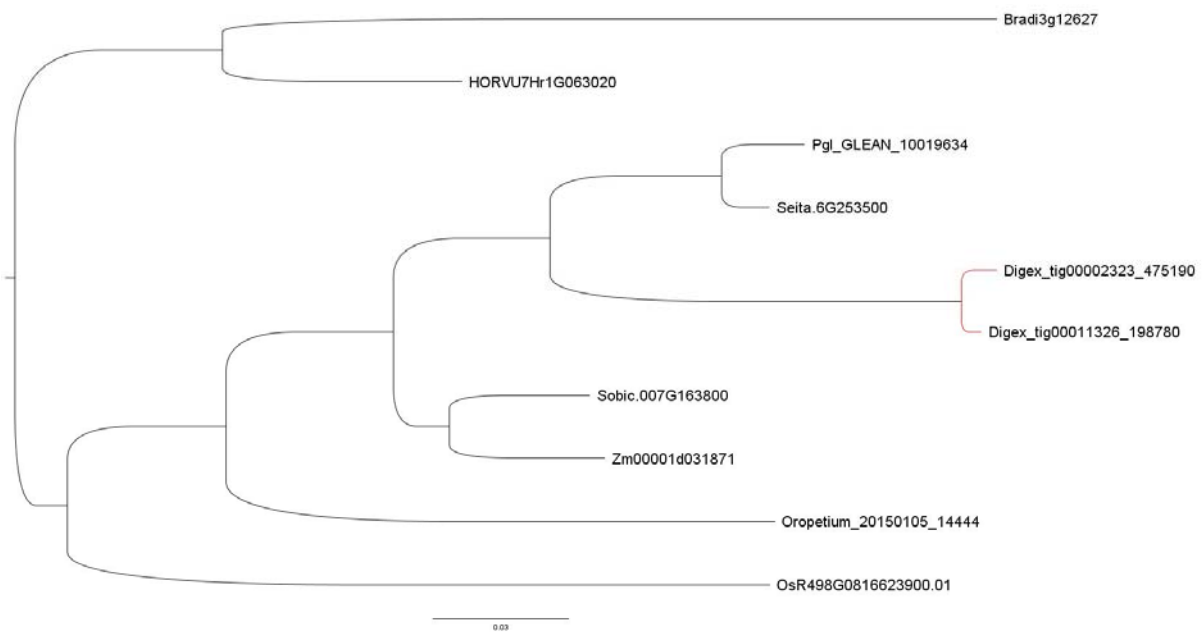
Suppl Figure S5. GO for of single copy, contracted genes in fonio.



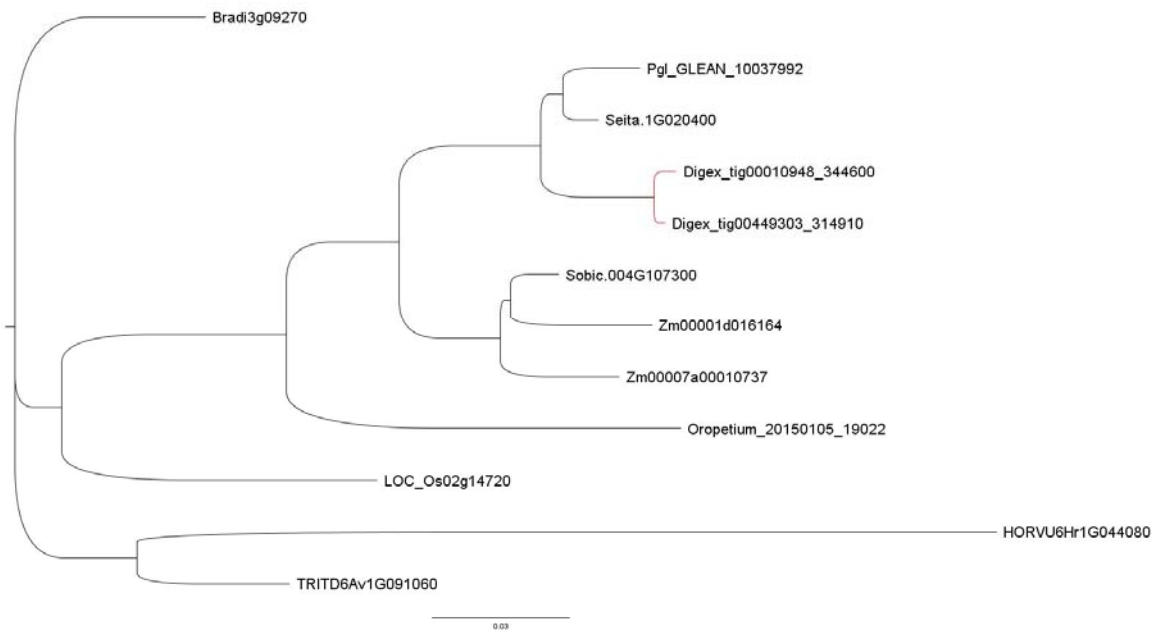
Suppl Figure S6. GO enrichment for expanded genes in *D. exilis* and relative to *O. sativa*



Suppl. Figure S7. Phylogenetic tree of the SSH-like genes from fonio and related species. The genes shaded in light blue are the family members most closely related to SSH-1 in *O. sativa* and *D. exilis*. Genes are named according to their PLAZA identifiers. Abbreviations for species names are as follow: Bradi (*Brachypodium distachyon*), pgl_GLEAN (*Cenchrus amercianus*), Digex (*Digitaria exilis*), Oropetium (*Oropetium thomaeium*), OsR (*Oryza sativa*), Seita (*Setaria italica*), Sobic (*Sorghum bicolor*), and Zm (*Zea mays*).



Suppl. Figure S8. Phylogenetic tree of the *dw3* gene family of fonio and related species.



Suppl. Figure S9. Gene family tree for GW2-A-like genes in fonio and related species. This figure also includes the genes from two additional Pooid species, barley (*Hordeum vulgare*) (HORV) and wheat, *Triticum turgidum* (TRITD).