Supplementary Information

Transcriptional profiling of contrasting genotypes revealed key candidates and nucleotide variations for drought dissection in *Camellia sinensis* (L.) O. Kuntze

Rajni Parmar[1,2], Romit Seth[1,3], Pradeep Singh[1,3], Gopal Singh[1,2], Sanjay Kumar[1,2], Ram Kumar Sharma[1,2]*

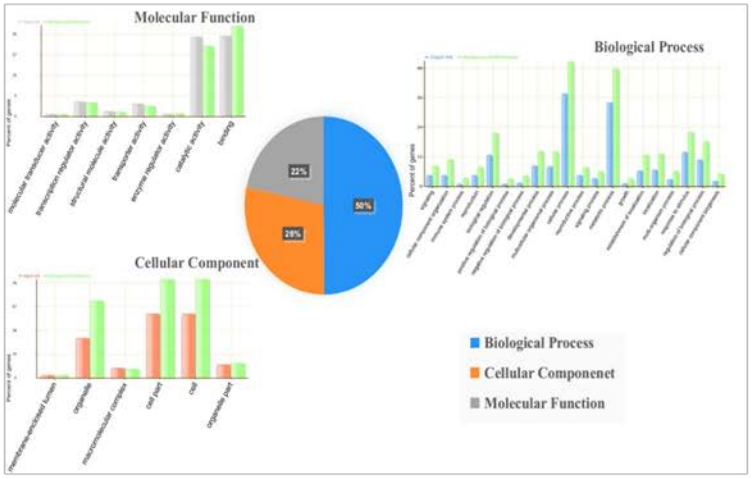
[1]Biotechnology Department, CSIR-Institute of Himalayan Bioresource Technology (CSIR-IHBT), Palampur, Himachal Pradesh, India

[2]Academy of Scientific and Innovative Research (AcSIR), CSIR-IHBT, Palampur, Himachal Pradesh, India

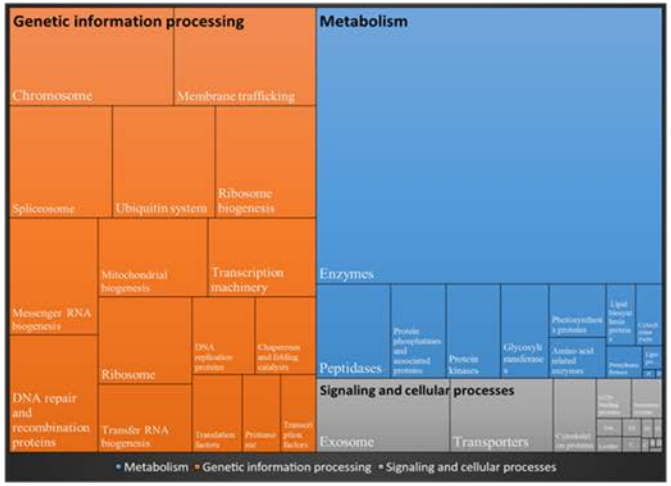
[3]Department of Biotechnology, Guru Nanak Dev University, Amritsar, 143005, India

***Correspondence:**
Dr Ram Kumar Sharma
Biotechnology Division
CSIR-Institute of Himalayan Bioresource Technology
Palampur (H.P)
India
Email: rksharma.ihbt@gmail.com, ramsharma@ihbt.res.in

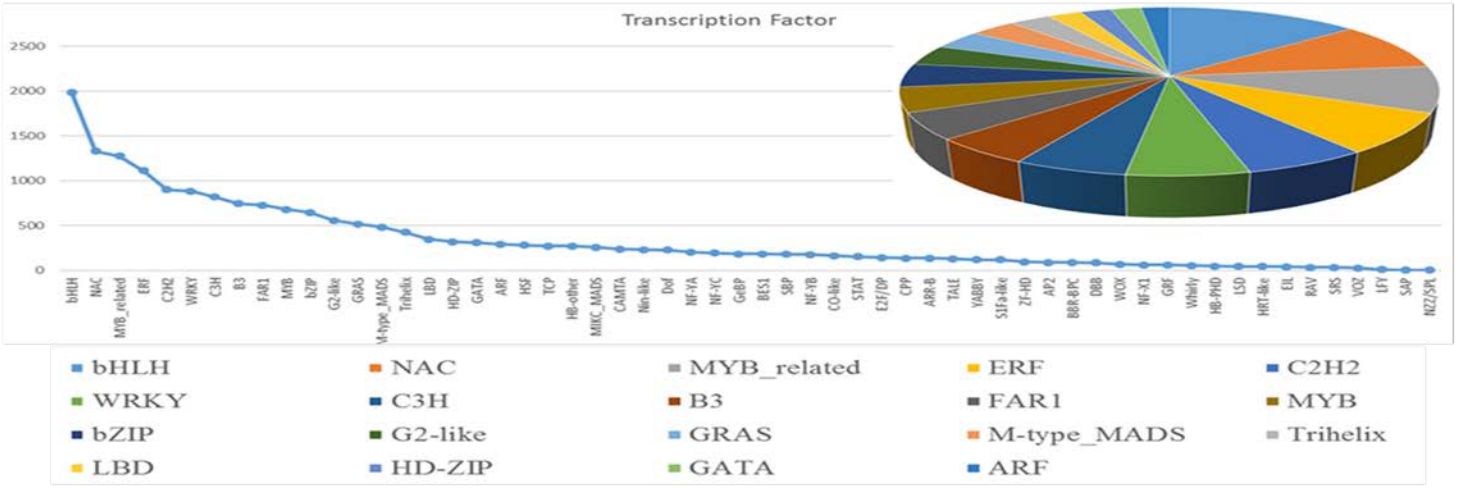
Figure S1



a



b



c

Figure S2

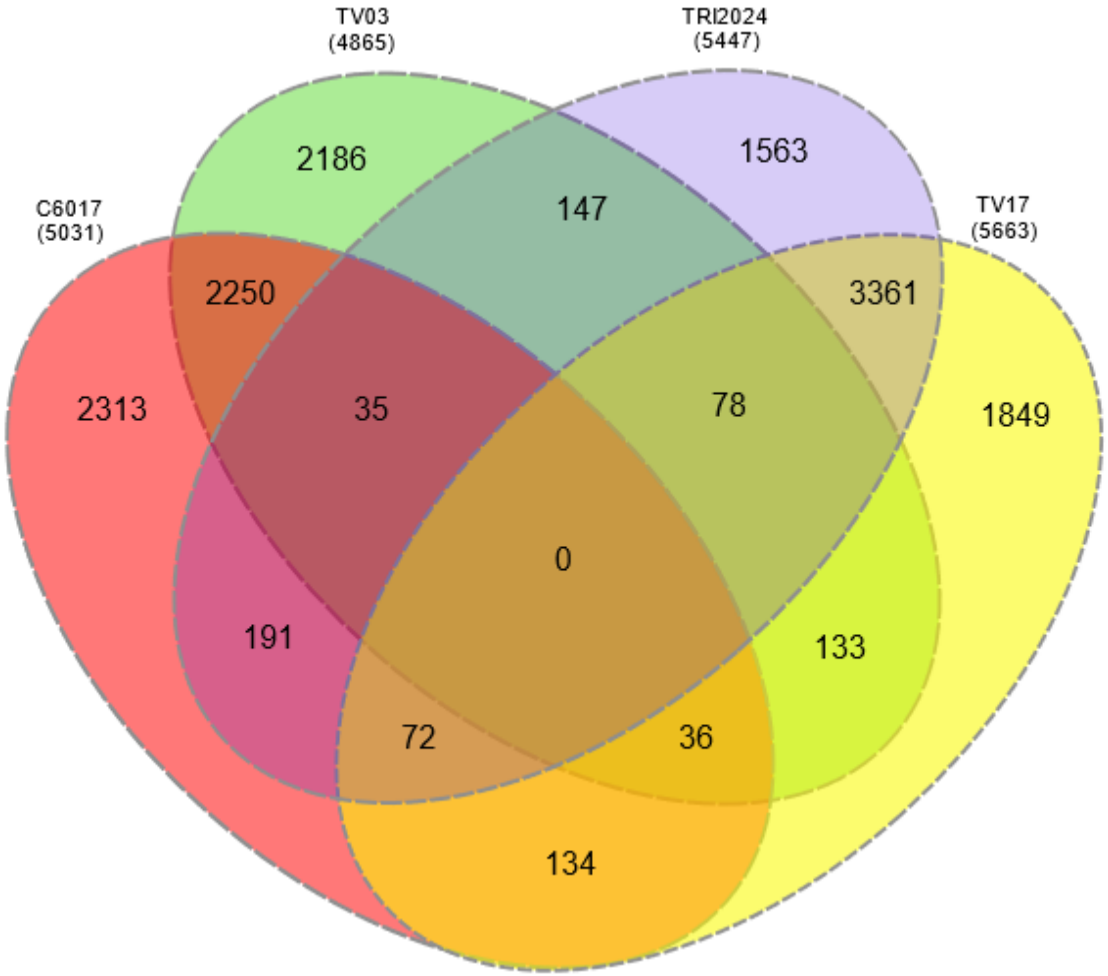

TV03
(4865)
TRI2024
(5447)
C6017
(5031)
TV17
(5663)
2186
1563
147
2250
3361
2313
35
78
1849
0
191
133
72
36
134

Figure S3a

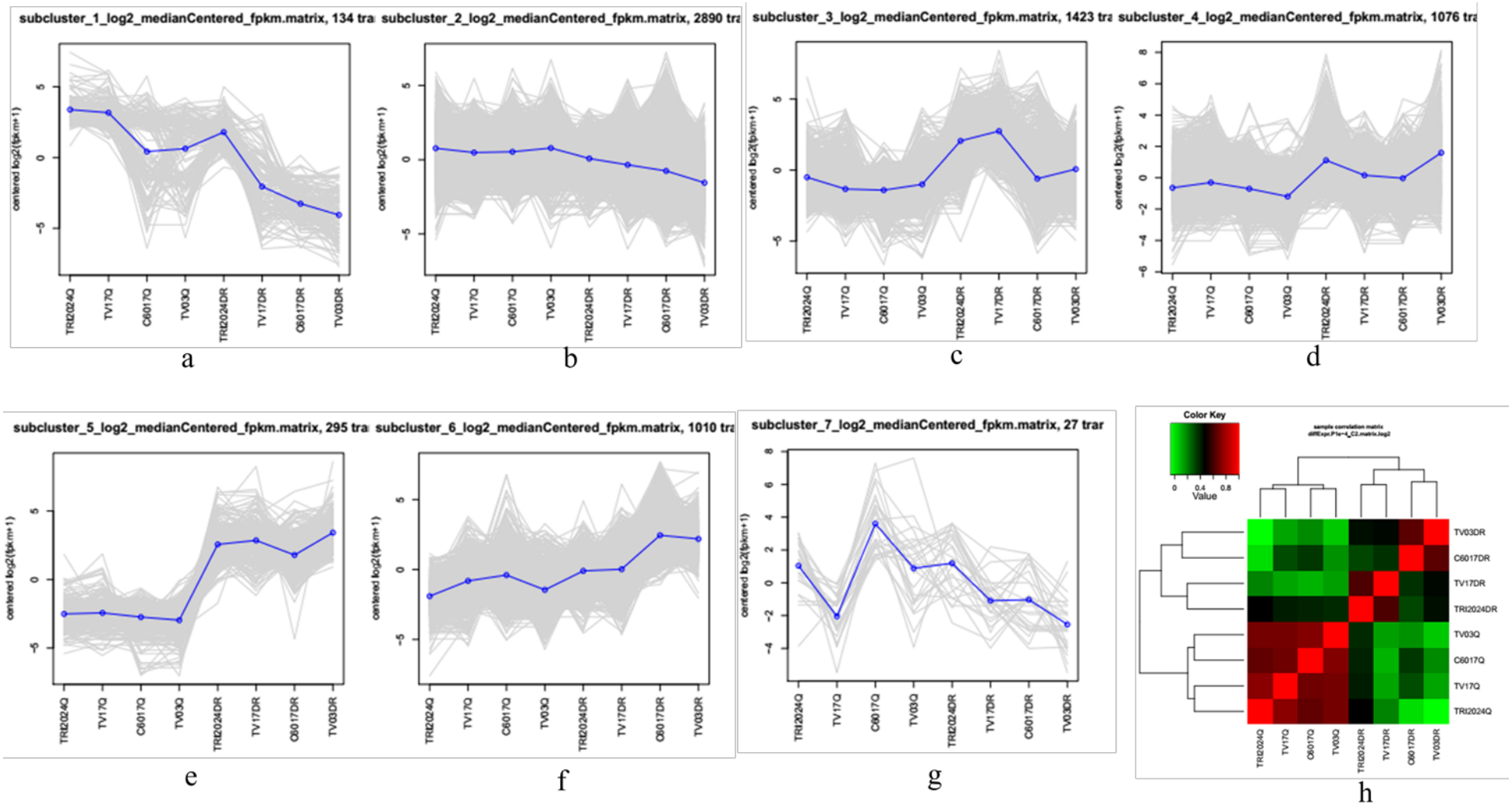

subcluster_1_log2_medianCentered_fpkm.matrix, 134 tra
subcluster_2_log2_medianCentered_fpkm.matrix, 2890 tra
subcluster_3_log2_medianCentered_fpkm.matrix, 1423 tra
subcluster_4_log2_medianCentered_fpkm.matrix, 1076 tra
subcluster_5_log2_medianCentered_fpkm.matrix, 295 tra
subcluster_6_log2_medianCentered_fpkm.matrix, 1010 tra
subcluster_7_log2_medianCentered_fpkm.matrix, 27 trar
centered log2(fpkm+1)
TRI2024Q
TV17Q
C6017Q
TV03Q
TRI2024DR
TV17DR
C6017DR
TV03DR
Color Key
Value
a
b
c
d
e
f
g
h

Figure S3b

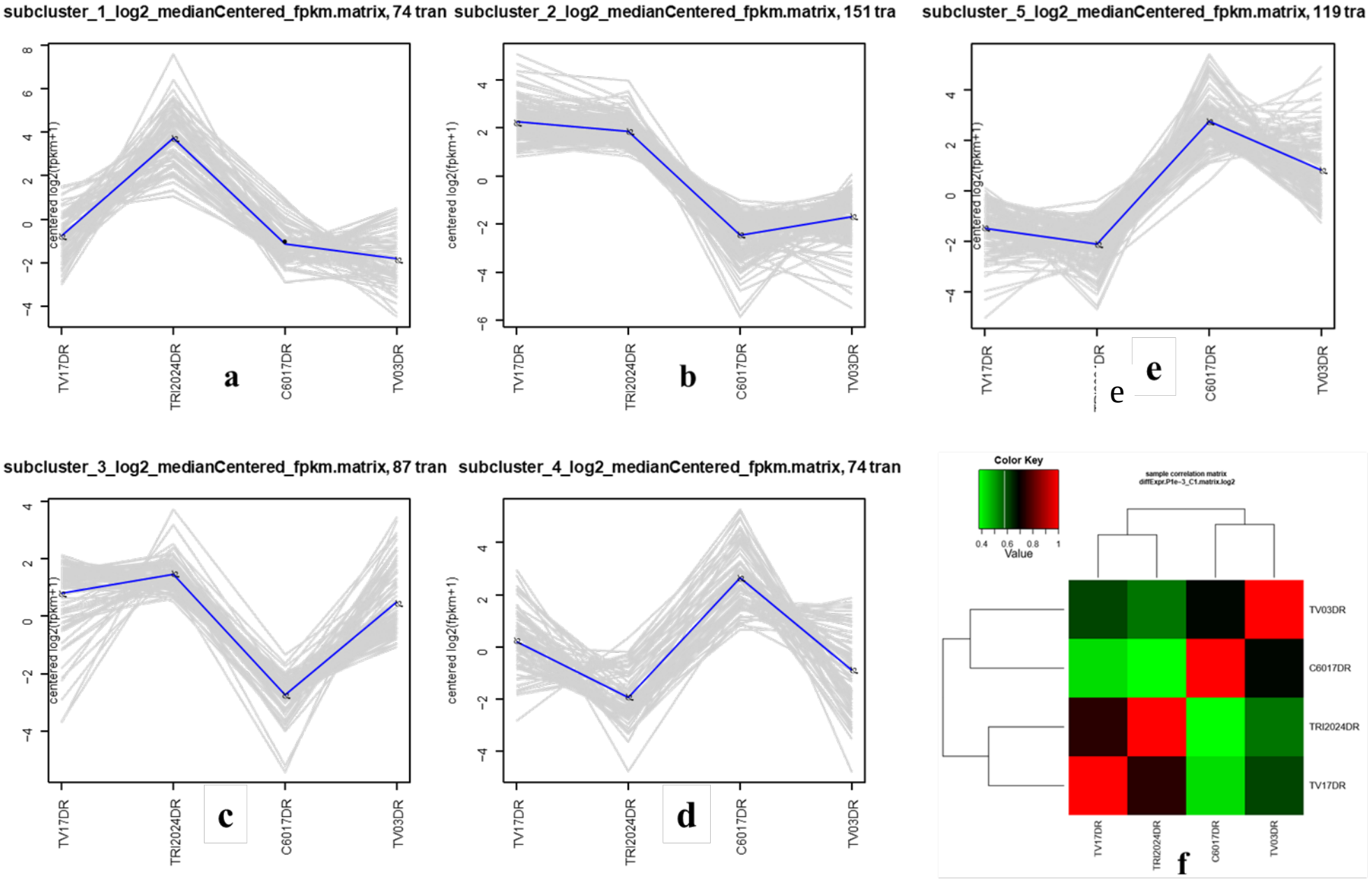

subcluster_1_log2_medianCentered_fpkm.matrix, 74 tran
subcluster_2_log2_medianCentered_fpkm.matrix, 151 tra
subcluster_5_log2_medianCentered_fpkm.matrix, 119 tra
subcluster_3_log2_medianCentered_fpkm.matrix, 87 tran
subcluster_4_log2_medianCentered_fpkm.matrix, 74 tran
centered log2(fpkm+1)
TV17DR
TRI2024DR
C6017DR
TV03DR
a
b
c
d
e
Color Key
Value
sample correlation matrix
diffExpr.P1e-3_C1.matrix.log2
f

Figure S4a (Molecular Function)

DT_G1

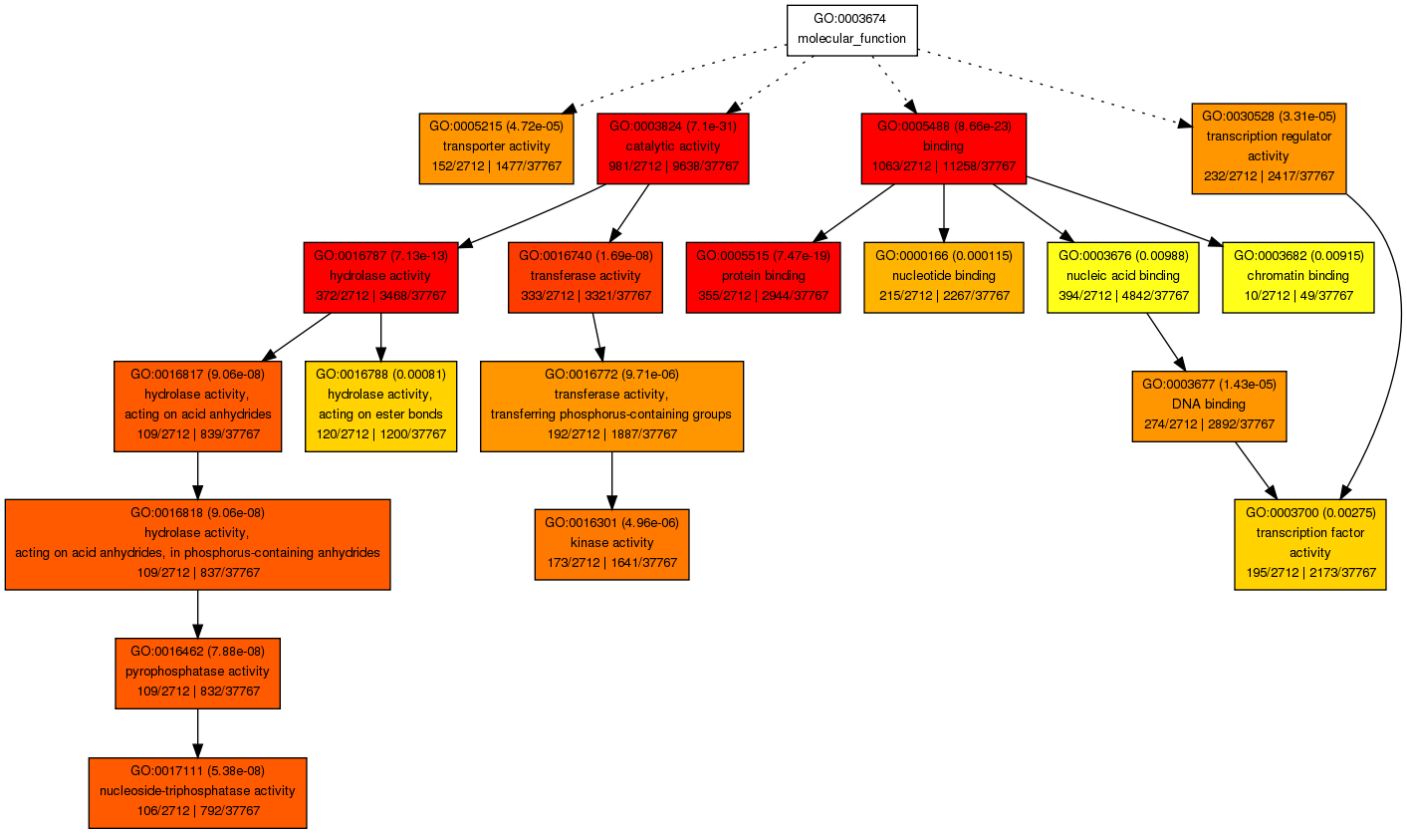


Figure S4b (Biological Function)

DT_G1

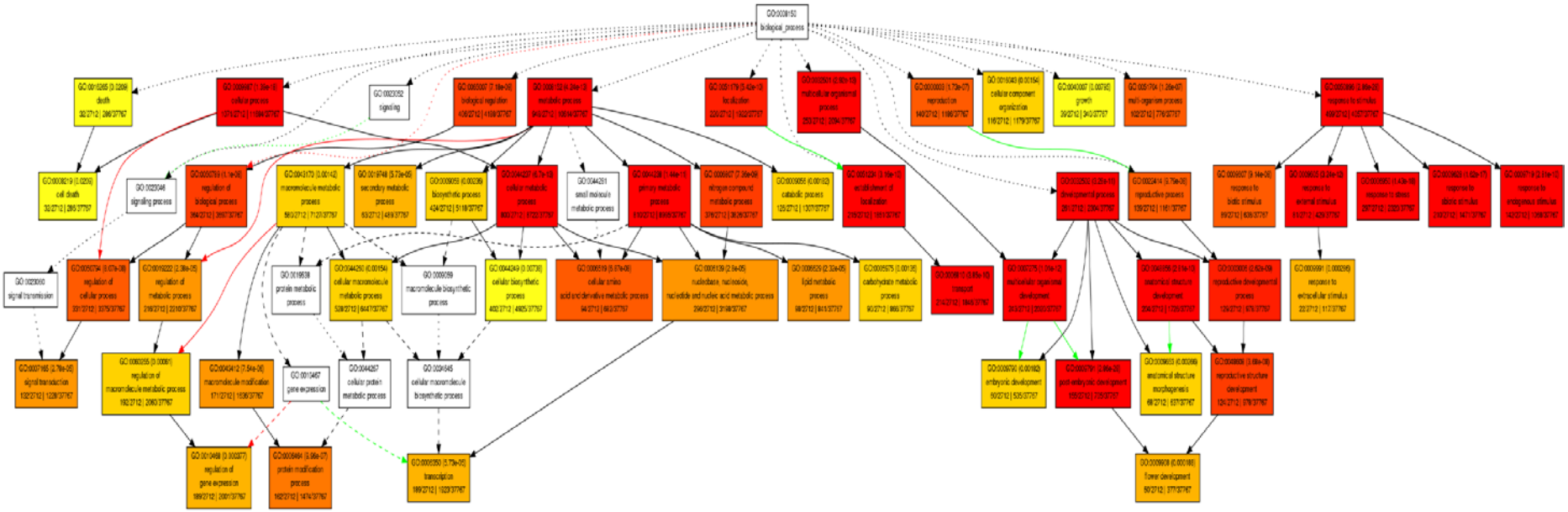


Figure S4c (Cellular Component)

DT_G1

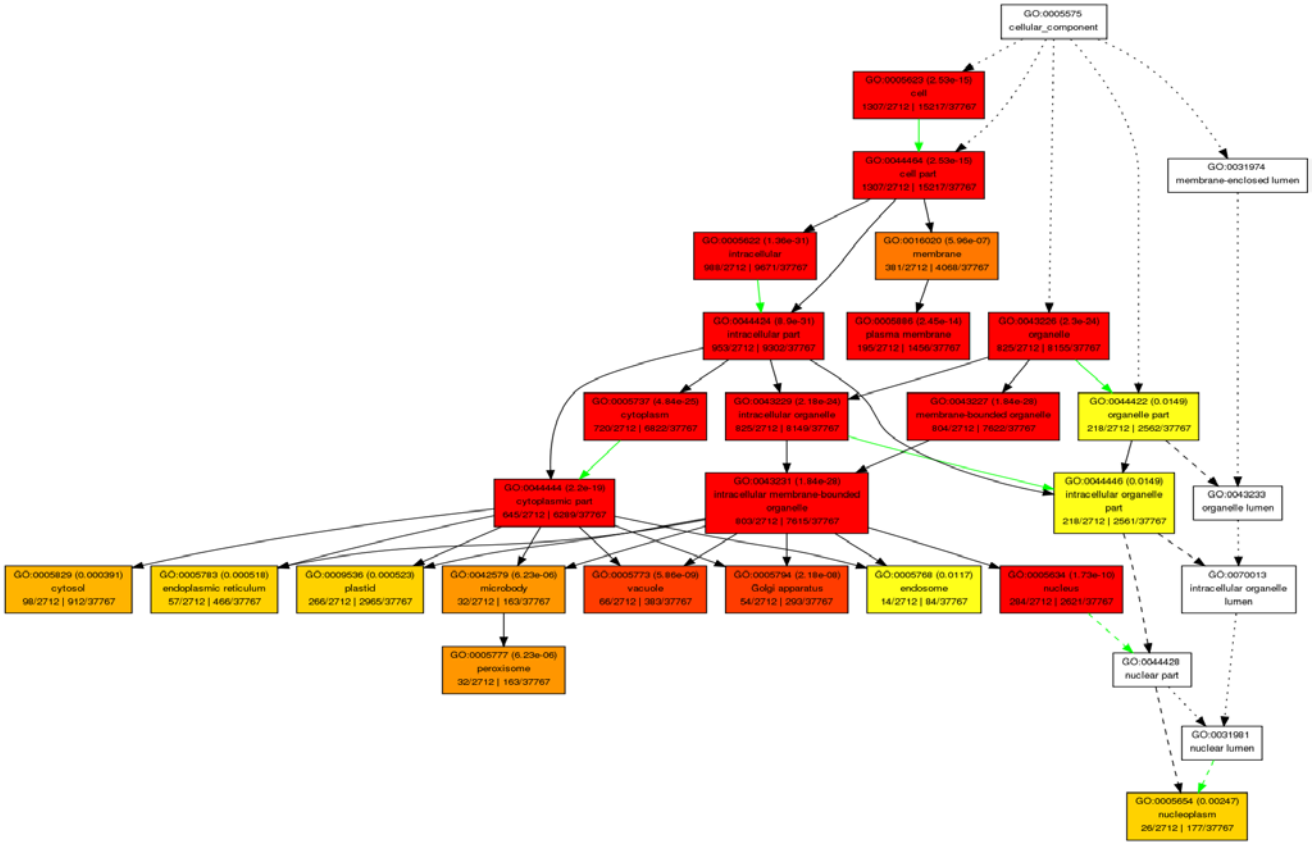


Figure S5a (Molecular Function)

DT_G2

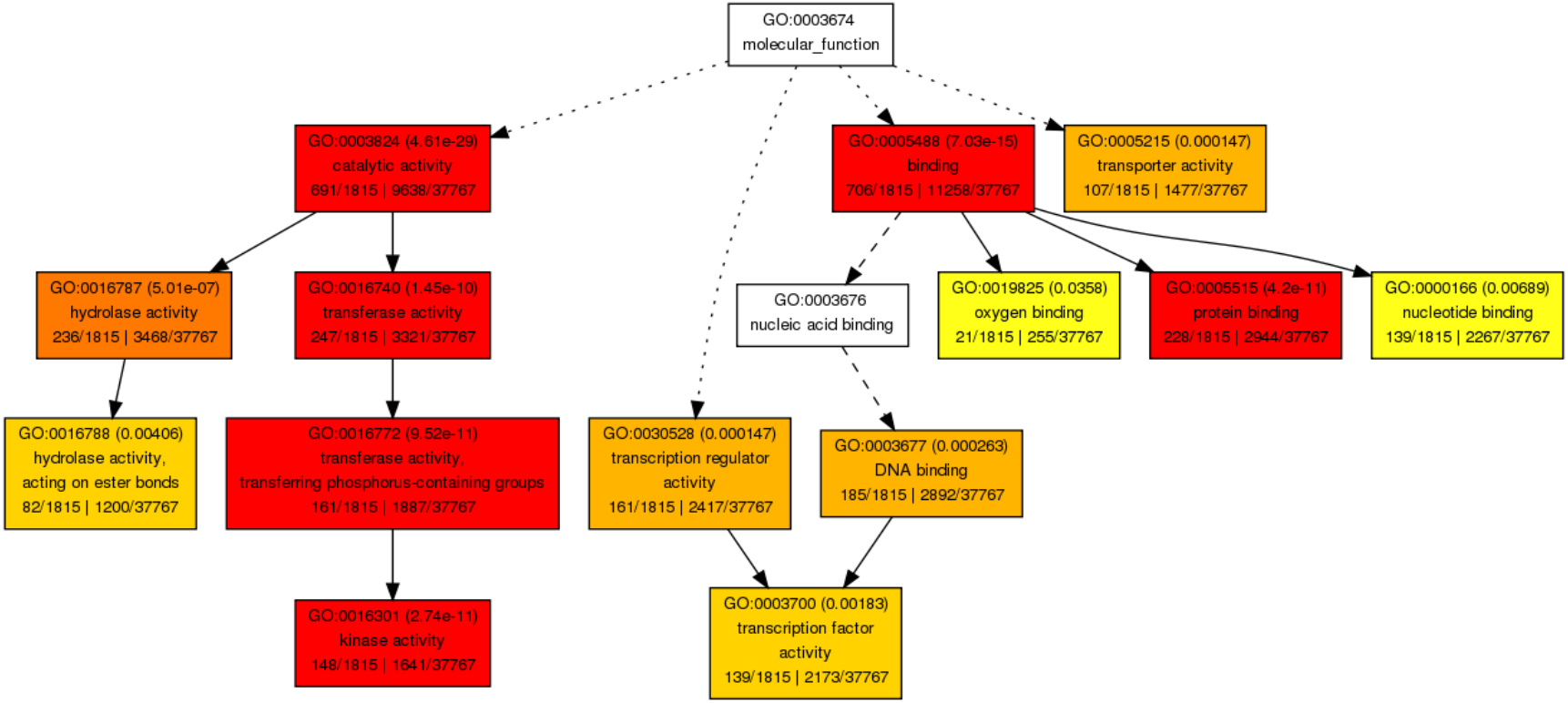


Figure S5b (Biological Function)

DT_G2

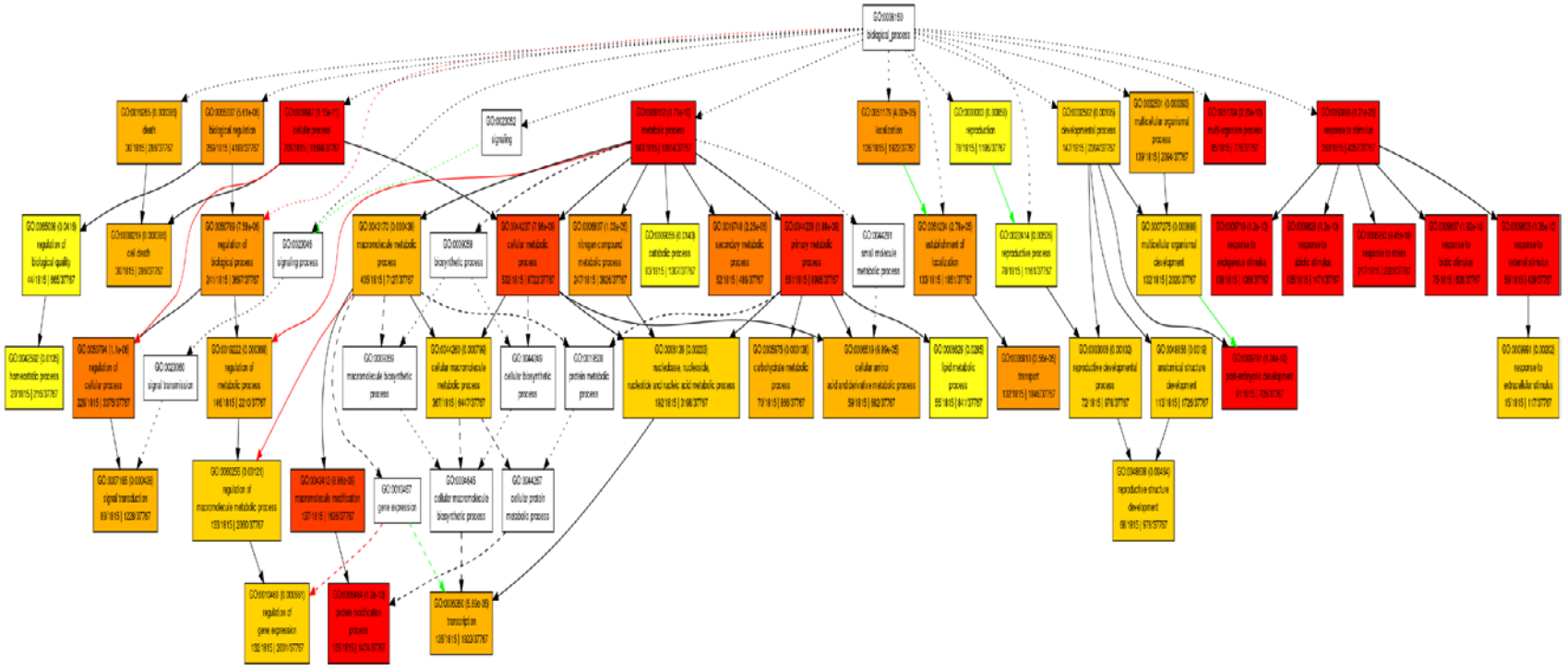


Figure S5c (Cellular Component)

DT_G2

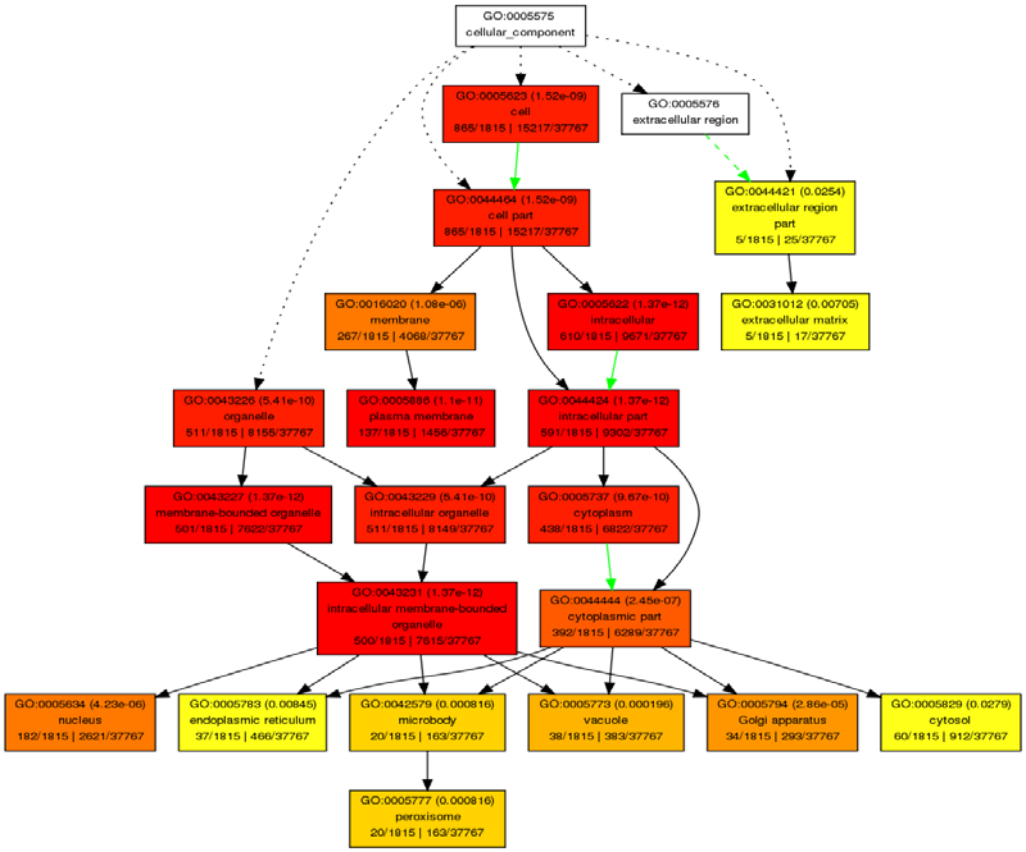


Figure S6a (Molecular Function)

DS_G1

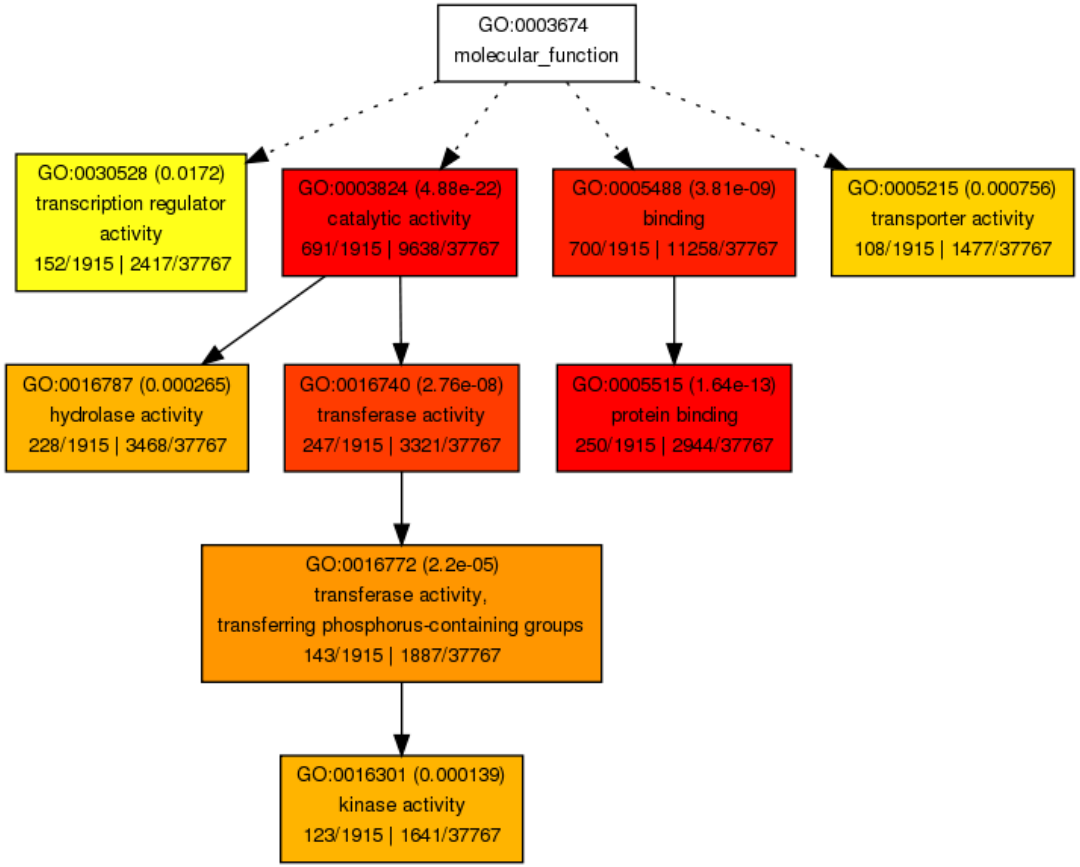


Figure S6b (Biological Function)

DS_G1

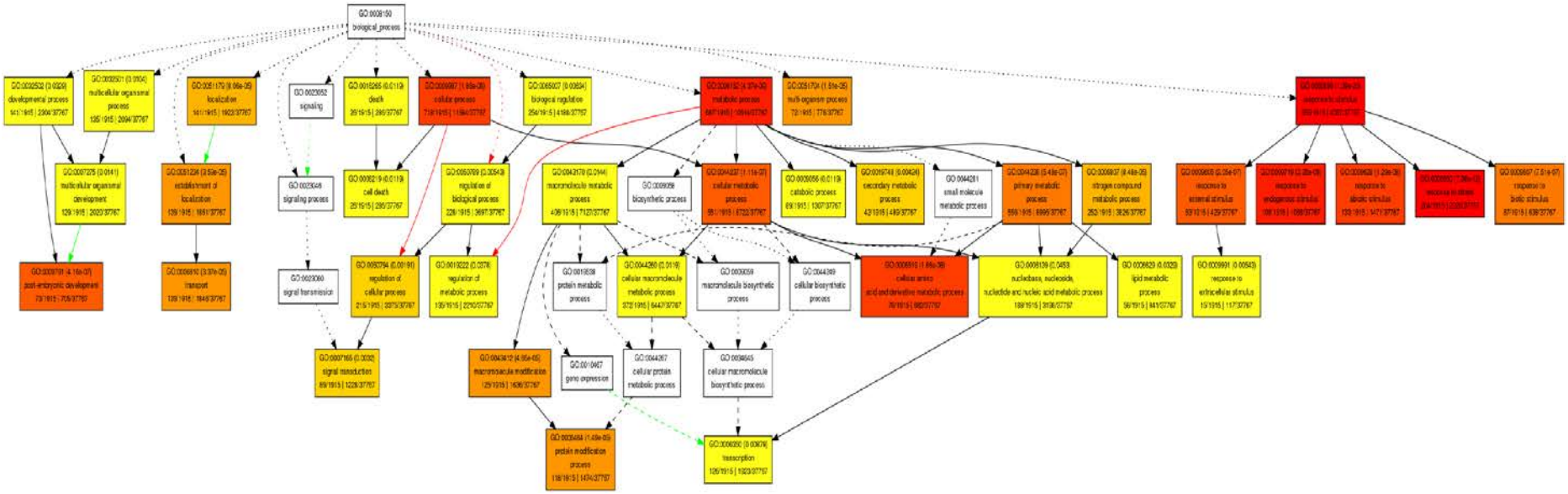


Figure S6c (Cellular Component)

DS_G1

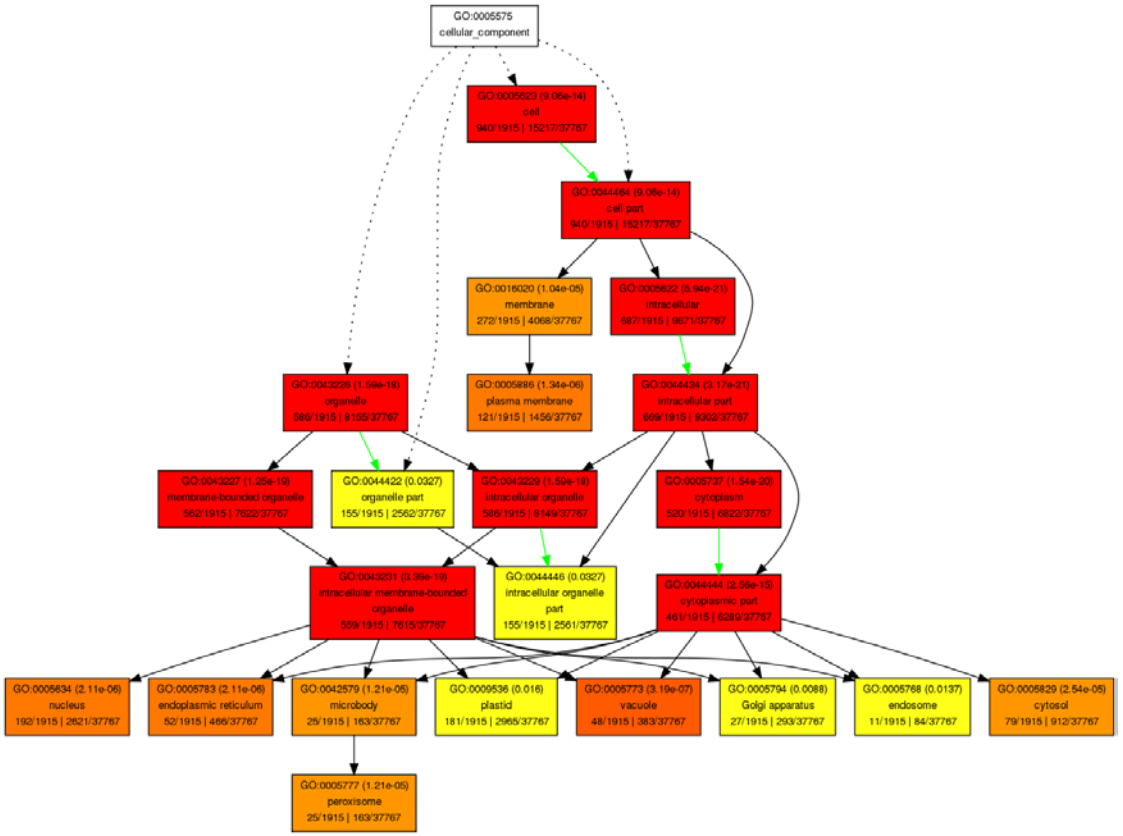


Figure S7a (Molecular Function)

DS_G2

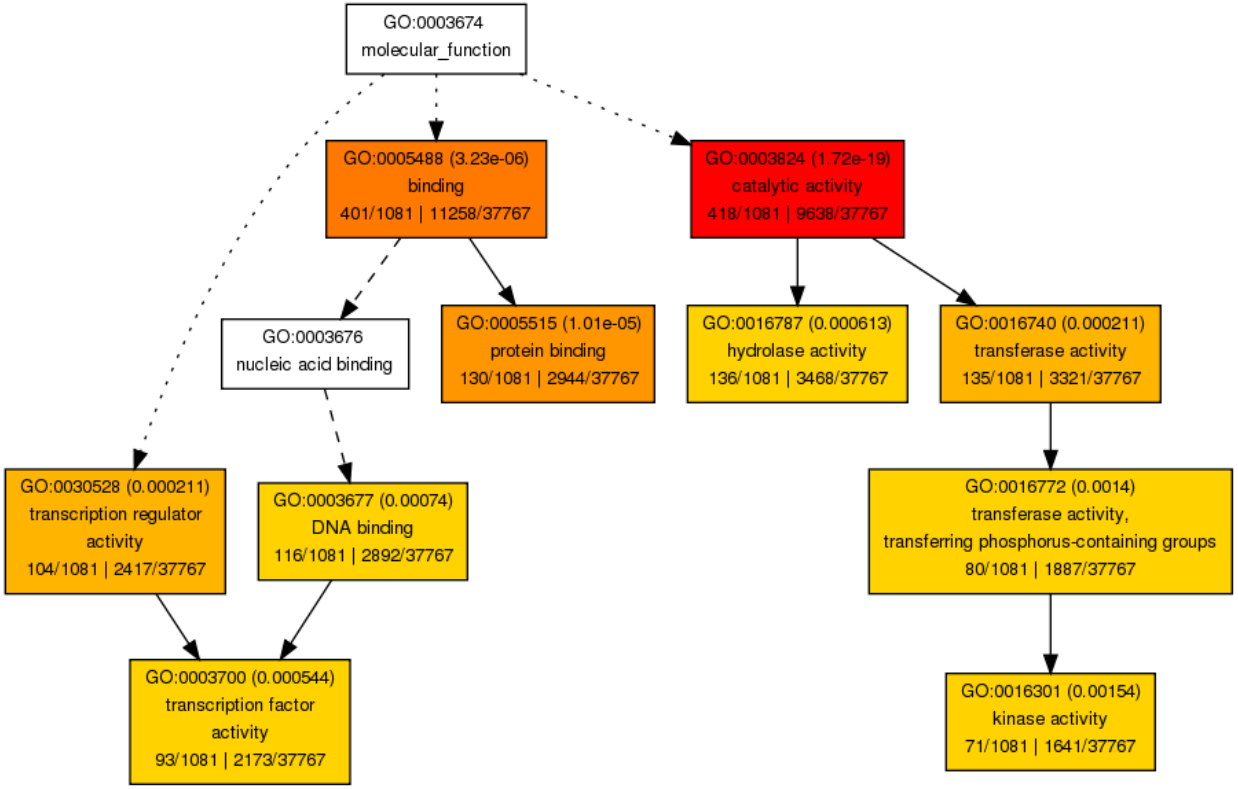


Figure S7b (Biological Function)

DS_G2

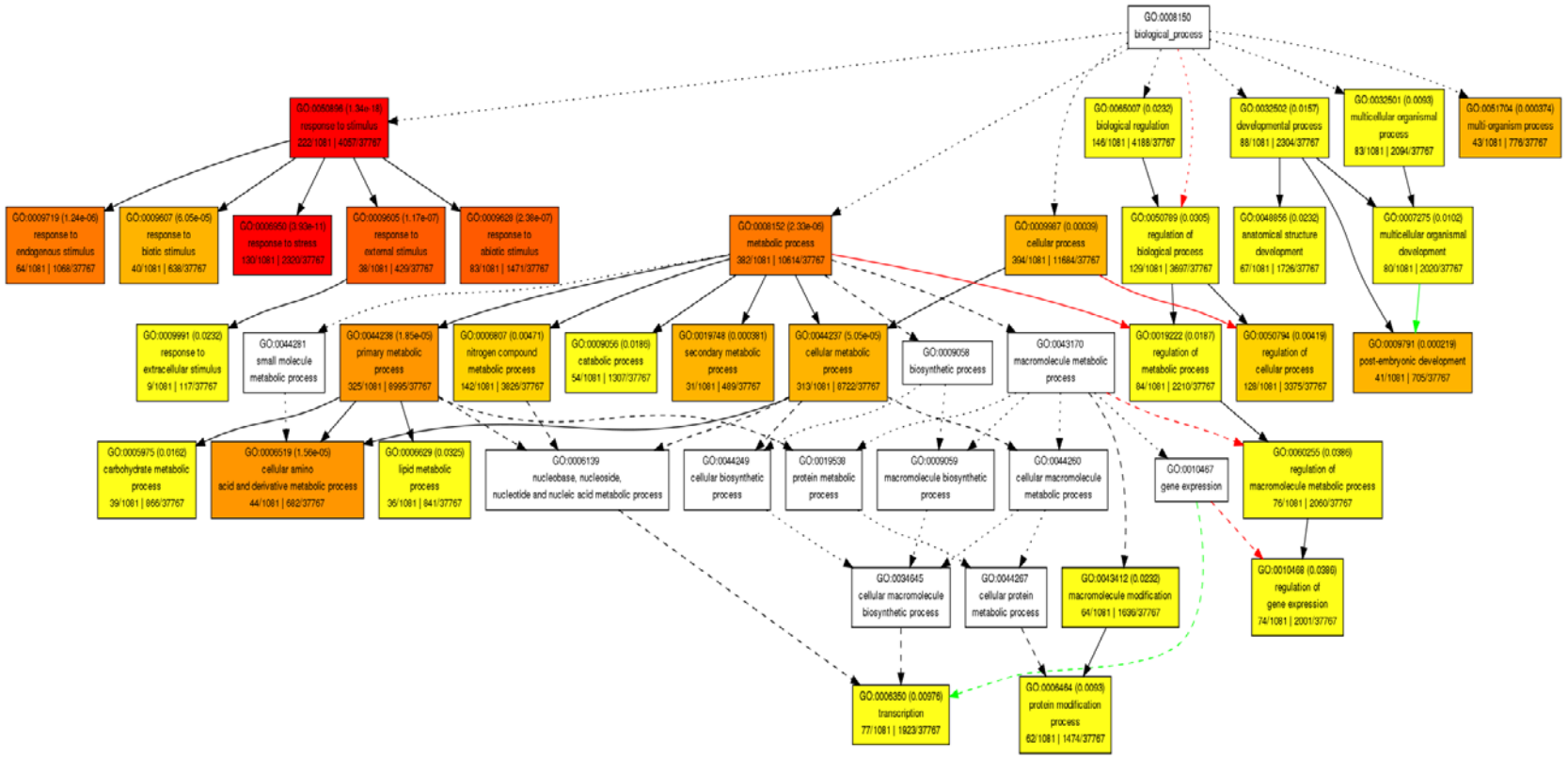


Figure S7a (Cellular Component)

DS_G2

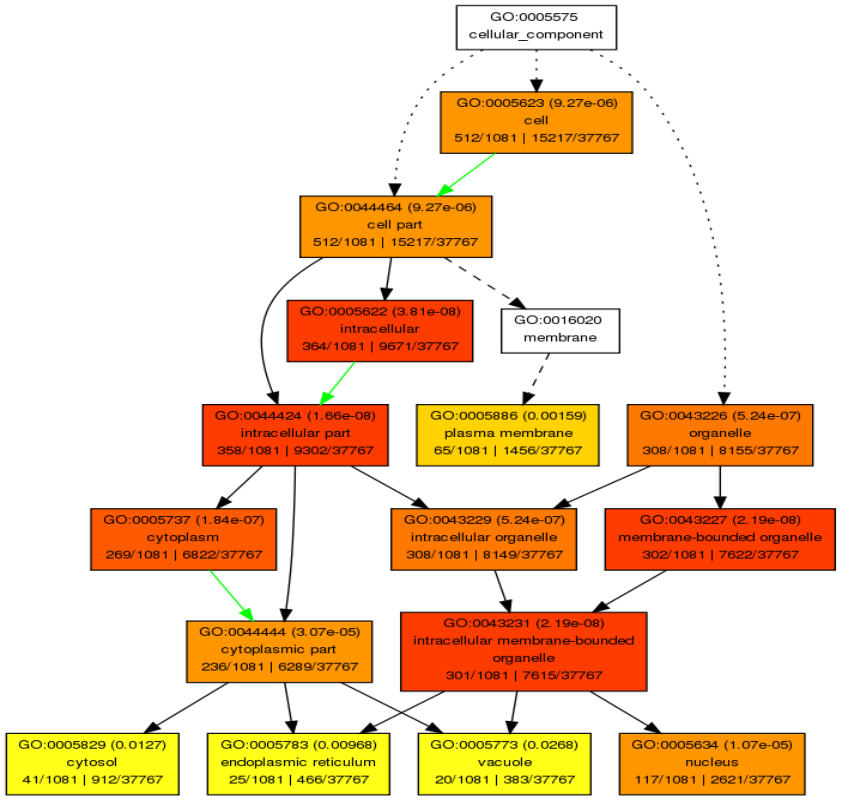
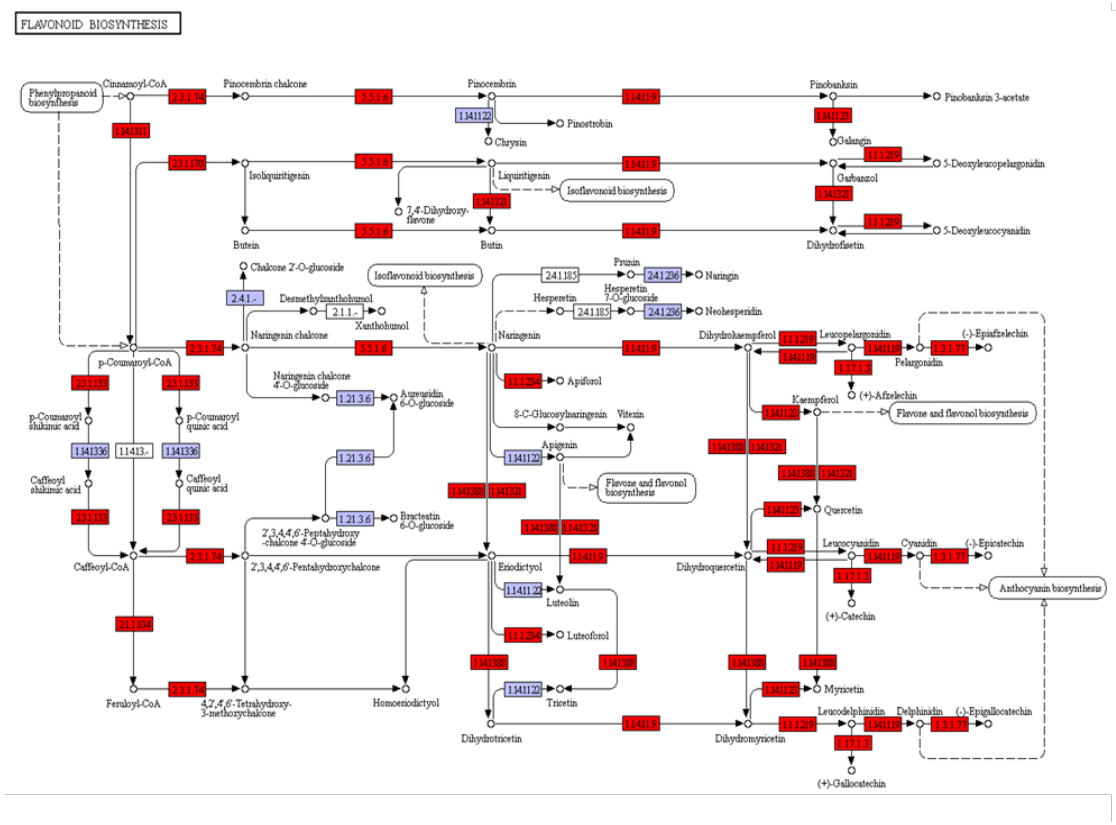


Figure S8

DT_G1_T and DT_G2_T



DS_G1_T and DS_G2_T

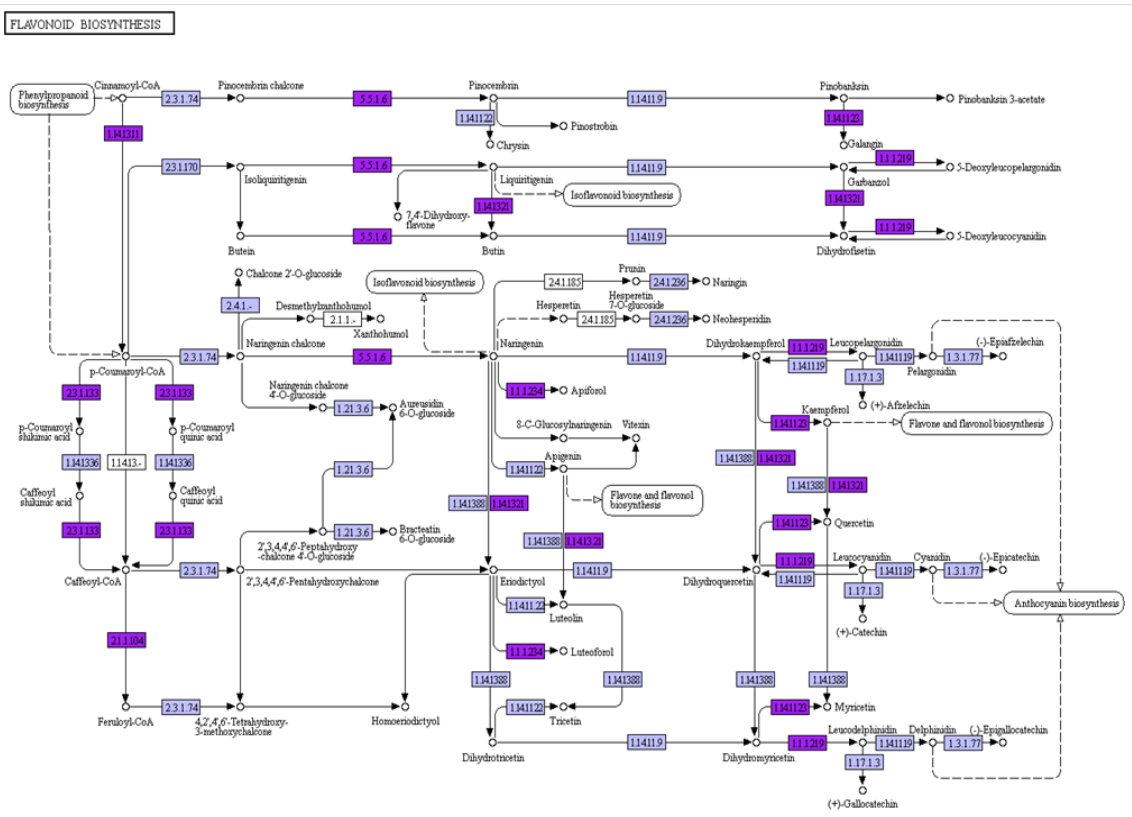
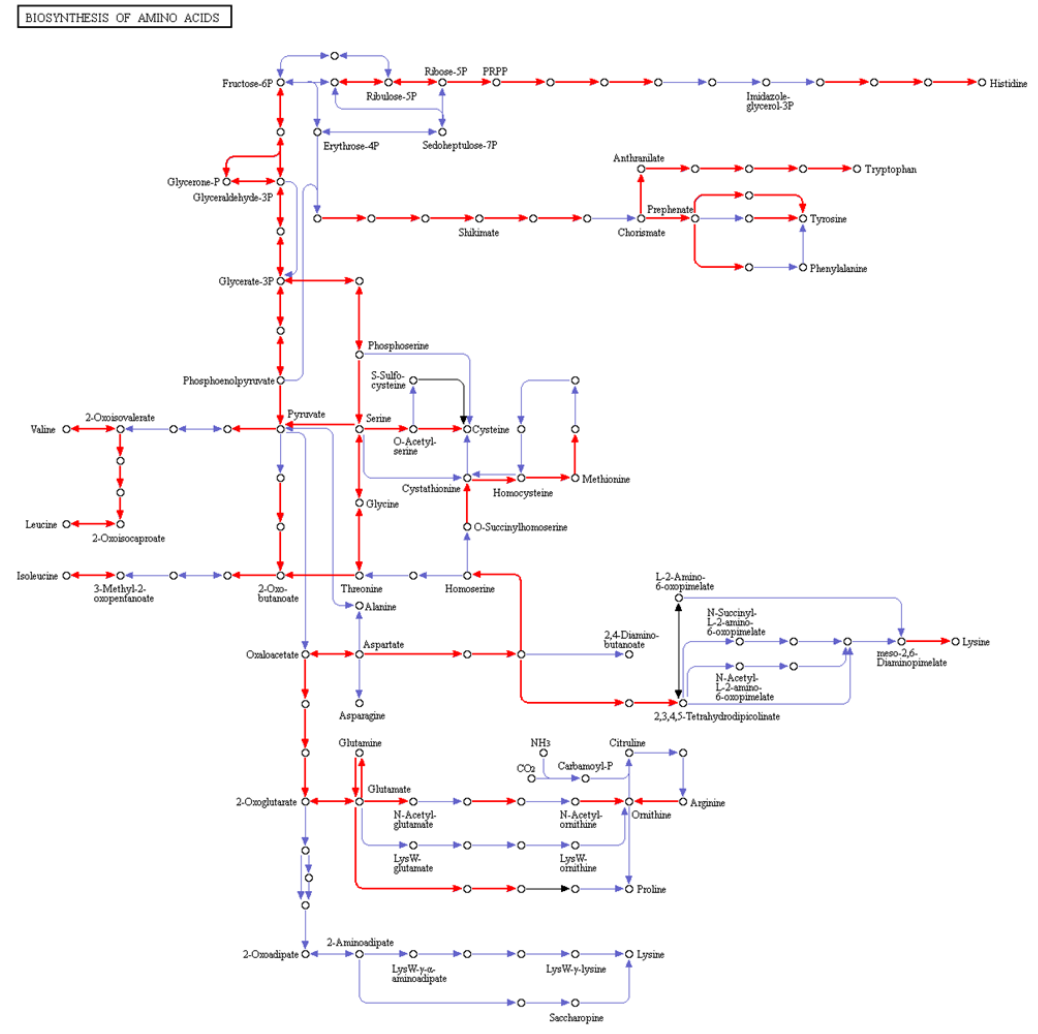


Figure S9

DT_G1_T and DT_G2_T



DS_G1_T and DS_G2_T

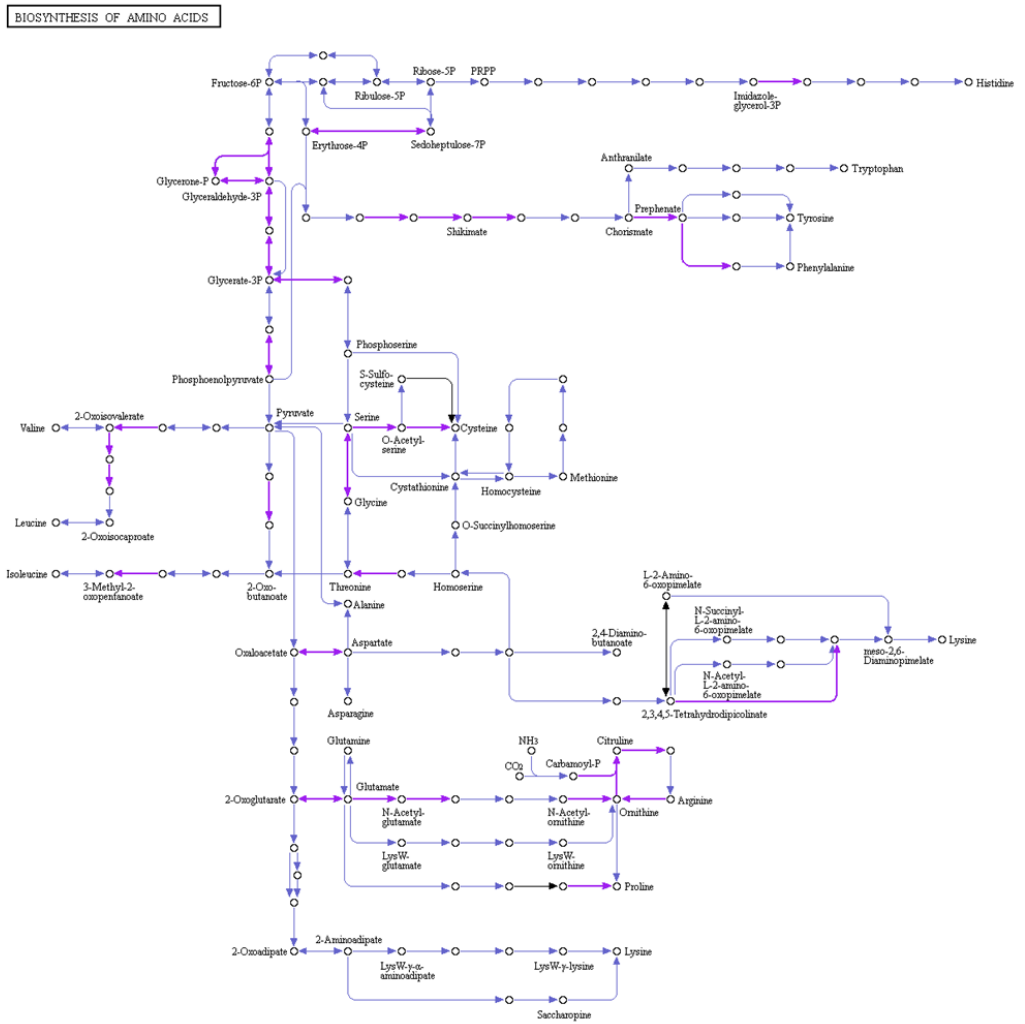
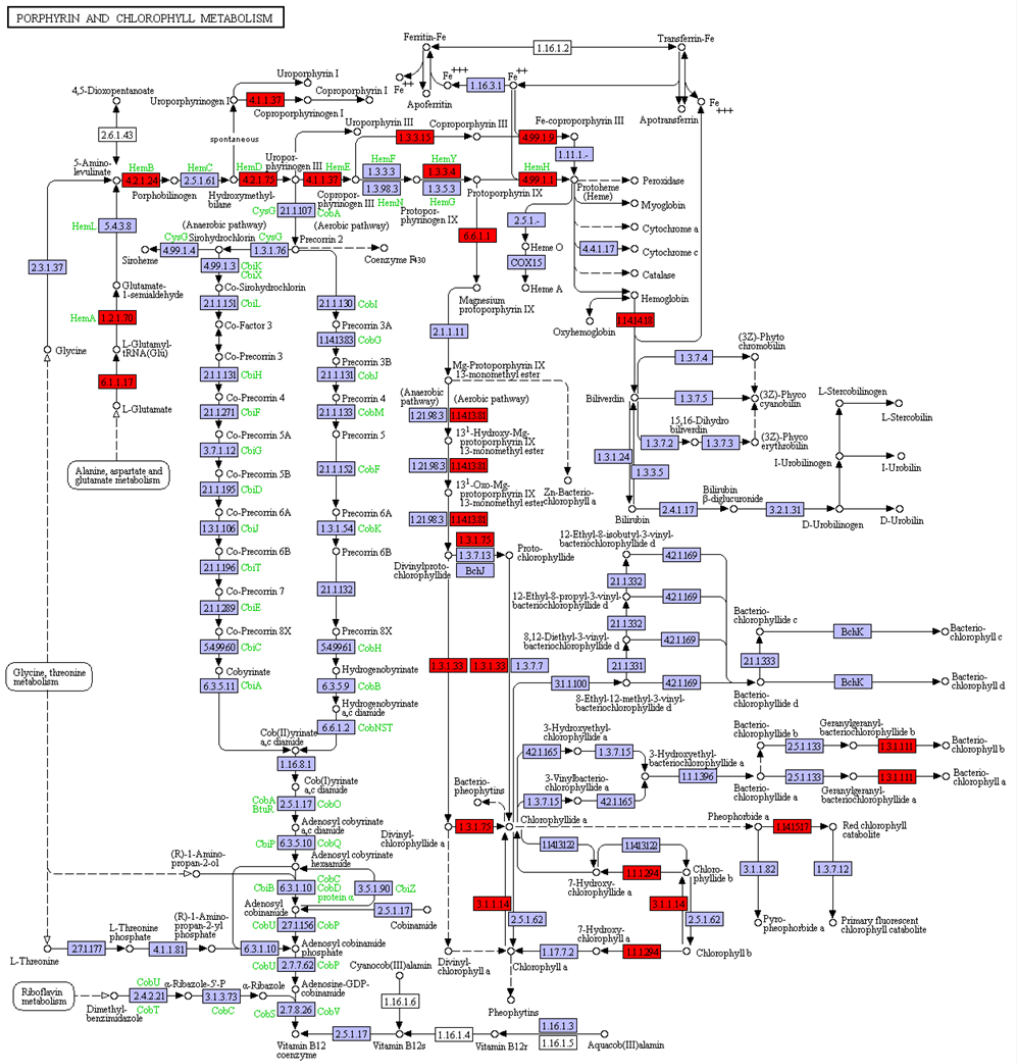


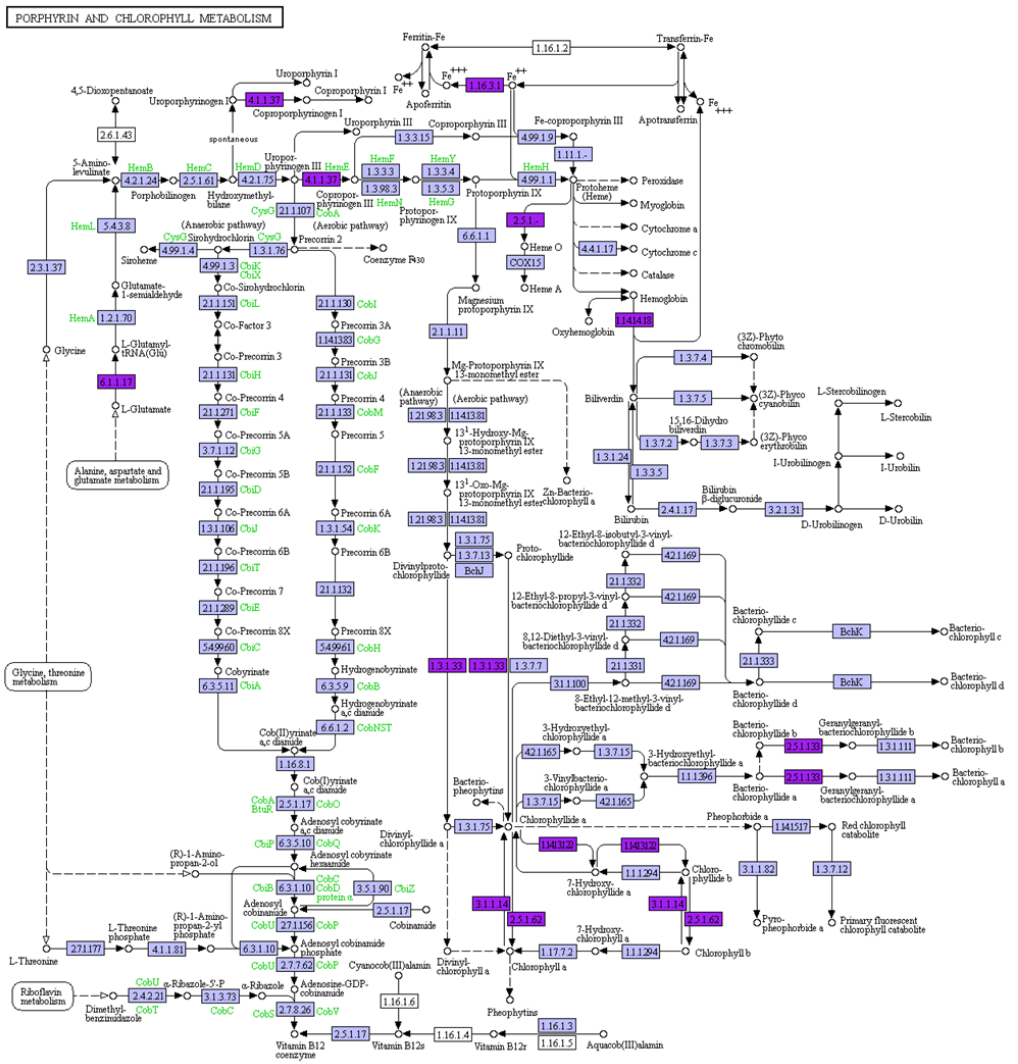
Figure S10

DT_G1_T and DT_G2_T



a

DS_G1_T and DS_G2_T



b

Figure S11

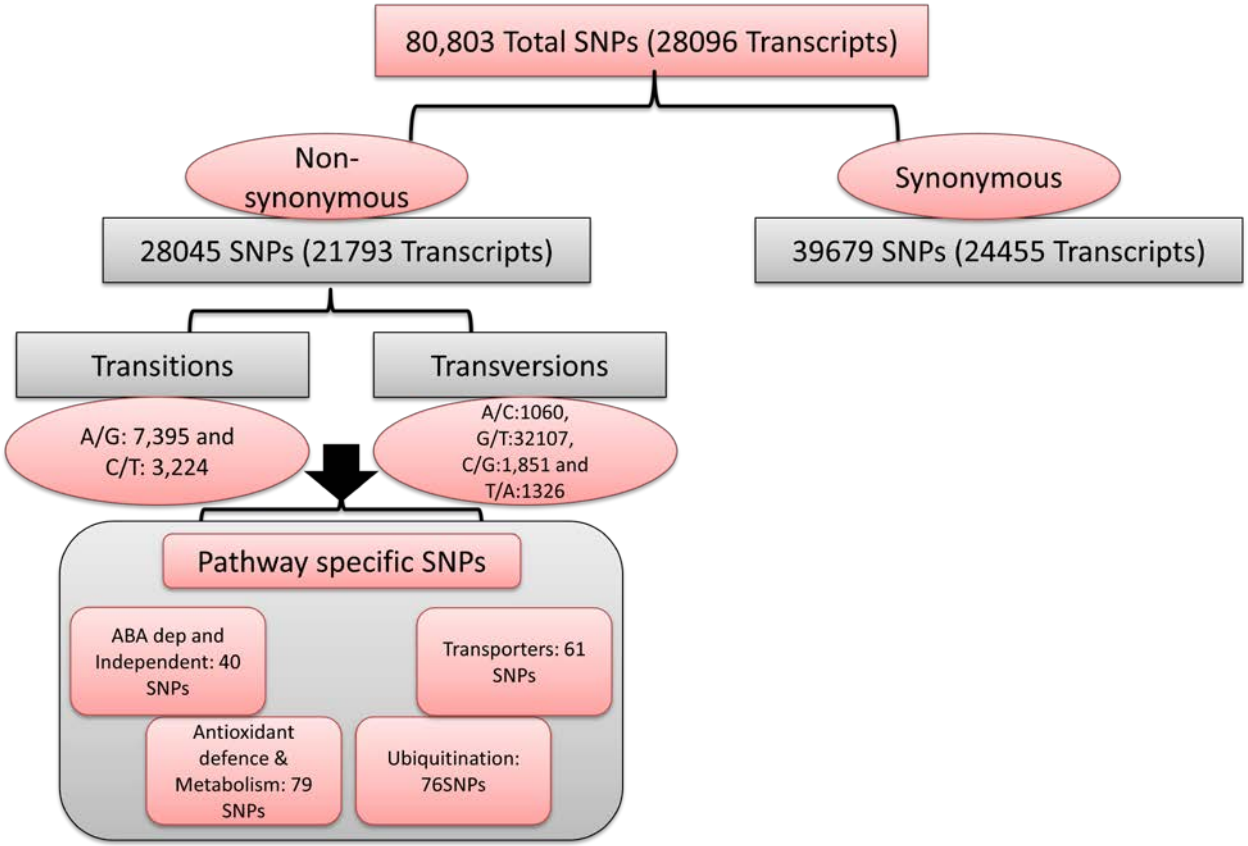

80,803 Total SNPs (28096 Transcripts)
Non-synonymous
Synonymous
28045 SNPs (21793 Transcripts)
39679 SNPs (24455 Transcripts)
Transitions
Transversions
A/G: 7,395 and C/T: 3,224
A/C:1060, G/T:32107, C/G:1,851 and T/A:1326
Pathway specific SNPs
ABA dep and Independent: 40 SNPs
Transporters: 61 SNPs
Antioxidant defence & Metabolism: 79 SNPs
Ubiquitination: 76SNPs

Figure S12

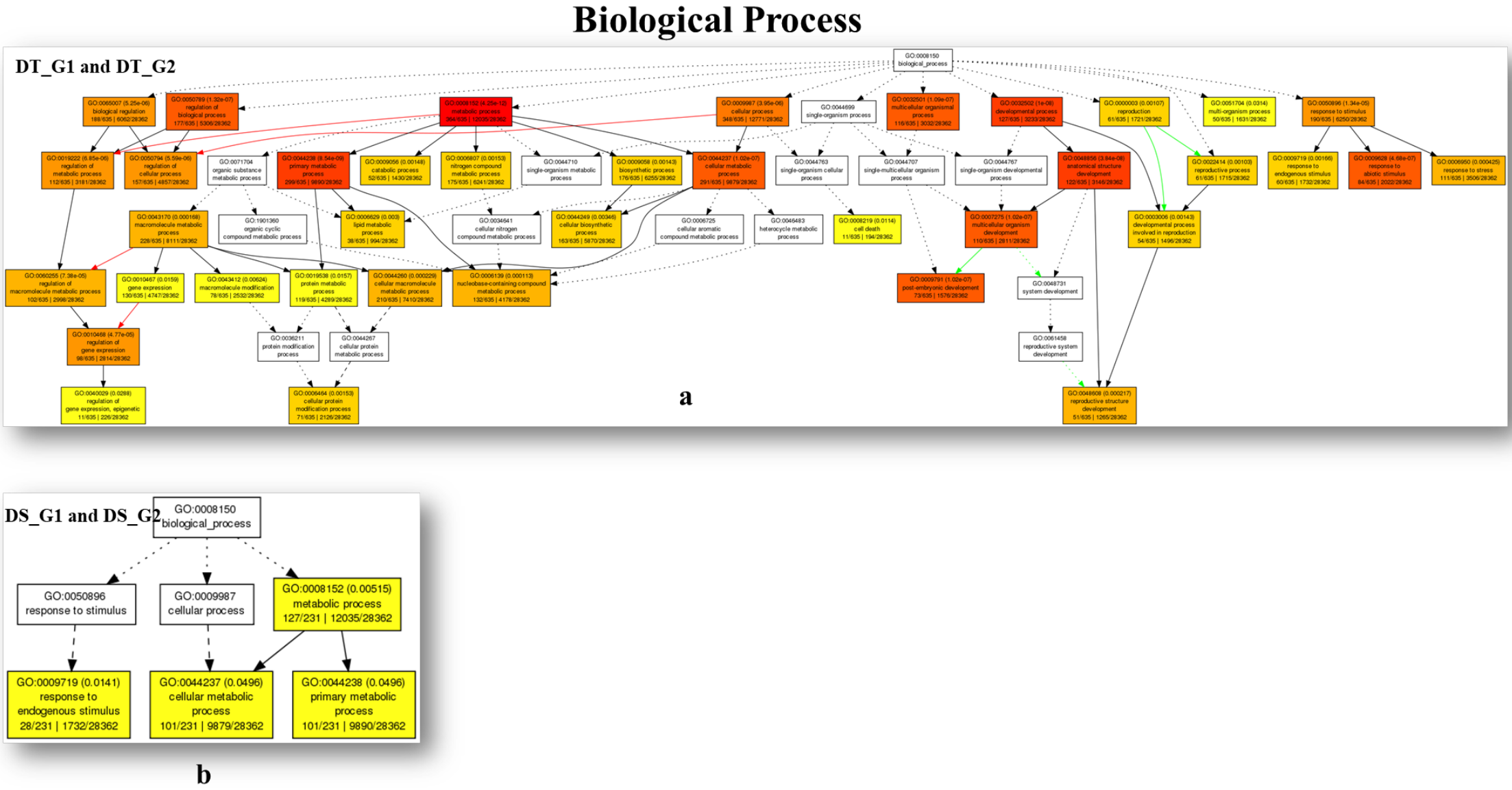
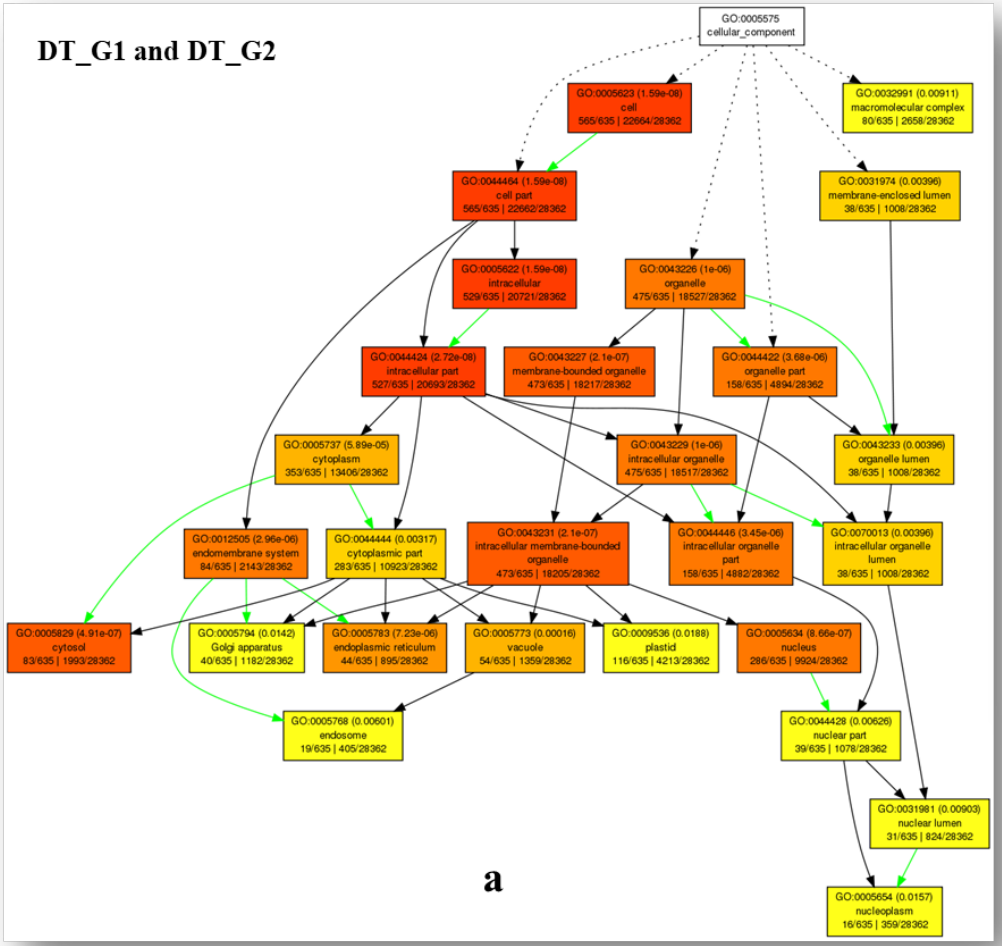

Biological Process
DT_G1 and DT_G2
GO:0008150
biological_process
GO:0065007 (5.25e-06)
biological regulation
188/635 | 6062/28362
GO:0050789 (1.32e-07)
regulation of
biological process
177/635 | 5306/28362
GO:0008152 (4.25e-12)
metabolic process
364/635 | 12035/28362
GO:0009987 (3.95e-06)
cellular process
348/635 | 12771/28362
GO:0044699
single-organism process
GO:0032501 (1.09e-07)
multicellular organismal
process
116/635 | 3032/28362
GO:0032502 (1e-08)
developmental process
127/635 | 3233/28362
GO:0000003 (0.00107)
reproduction
61/635 | 1721/28362
GO:0051704 (0.0314)
multi-organism process
50/635 | 1631/28362
GO:0050896 (1.34e-05)
response to stimulus
190/635 | 6250/28362
GO:0019222 (6.85e-06)
regulation of
metabolic process
112/635 | 3181/28362
GO:0050794 (5.59e-06)
regulation of
cellular process
157/635 | 4857/28362
GO:0071704
organic substance
metabolic process
GO:0044238 (8.54e-09)
primary metabolic
process
299/635 | 9890/28362
GO:0009056 (0.00148)
catabolic process
52/635 | 1430/28362
GO:0006807 (0.00153)
nitrogen compound
metabolic process
175/635 | 6241/28362
GO:0044710
single-organism metabolic
process
GO:0009058 (0.00143)
biosynthetic process
176/635 | 6255/28362
GO:0044237 (1.02e-07)
cellular metabolic
process
291/635 | 9879/28362
GO:0044763
single-organism cellular
process
GO:0044707
single-multicellular organism
process
GO:0044767
single-organism developmental
process
GO:0048856 (3.84e-08)
anatomical structure
development
122/635 | 3146/28362
GO:0022414 (0.00103)
reproductive process
61/635 | 1715/28362
GO:0009719 (0.00166)
response to
endogenous stimulus
60/635 | 1732/28362
GO:0009628 (4.68e-07)
response to
abiotic stimulus
84/635 | 2022/28362
GO:0006950 (0.000425)
response to stress
111/635 | 3506/28362
GO:0043170 (0.000168)
macromolecule metabolic
process
228/635 | 8111/28362
GO:1901360
organic cyclic
compound metabolic process
GO:0006629 (0.003)
lipid metabolic
process
38/635 | 994/28362
GO:0034641
cellular nitrogen
compound metabolic process
GO:0044249 (0.00346)
cellular biosynthetic
process
163/635 | 5870/28362
GO:0006725
cellular aromatic
compound metabolic process
GO:0046483
heterocycle metabolic
process
GO:0008219 (0.0114)
cell death
11/635 | 194/28362
GO:0007275 (1.02e-07)
multicellular organism
development
110/635 | 2811/28362
GO:0003006 (0.00143)
developmental process
involved in reproduction
54/635 | 1496/28362
GO:0060255 (7.38e-05)
regulation of
macromolecule metabolic process
102/635 | 2998/28362
GO:0010467 (0.0159)
gene expression
130/635 | 4747/28362
GO:0043412 (0.00624)
macromolecule modification
79/635 | 2532/28362
GO:0019538 (0.0157)
protein metabolic
process
119/635 | 4289/28362
GO:0044260 (0.000229)
cellular macromolecule
metabolic process
210/635 | 7410/28362
GO:0006139 (0.000113)
nucleobase-containing compound
metabolic process
132/635 | 4178/28362
GO:0009791 (1.02e-07)
post-embryonic development
73/635 | 1576/28362
GO:0048731
system development
GO:0010468 (4.77e-05)
regulation of
gene expression
98/635 | 2914/28362
GO:0036211
protein modification
process
GO:0044267
cellular protein
metabolic process
GO:0061458
reproductive system
development
GO:0040029 (0.0288)
regulation of
gene expression, epigenetic
11/635 | 226/28362
GO:0006464 (0.00153)
cellular protein
modification process
71/635 | 2126/28362
GO:0048608 (0.000217)
reproductive structure
development
51/635 | 1265/28362
a
DS_G1 and DS_G2
GO:0008150
biological_process
GO:0050896
response to stimulus
GO:0009987
cellular process
GO:0008152 (0.00515)
metabolic process
127/231 | 12035/28362
GO:0009719 (0.0141)
response to
endogenous stimulus
28/231 | 1732/28362
GO:0044237 (0.0496)
cellular metabolic
process
101/231 | 9879/28362
GO:0044238 (0.0496)
primary metabolic
process
101/231 | 9890/28362
b

Figure S13

Cellular Component



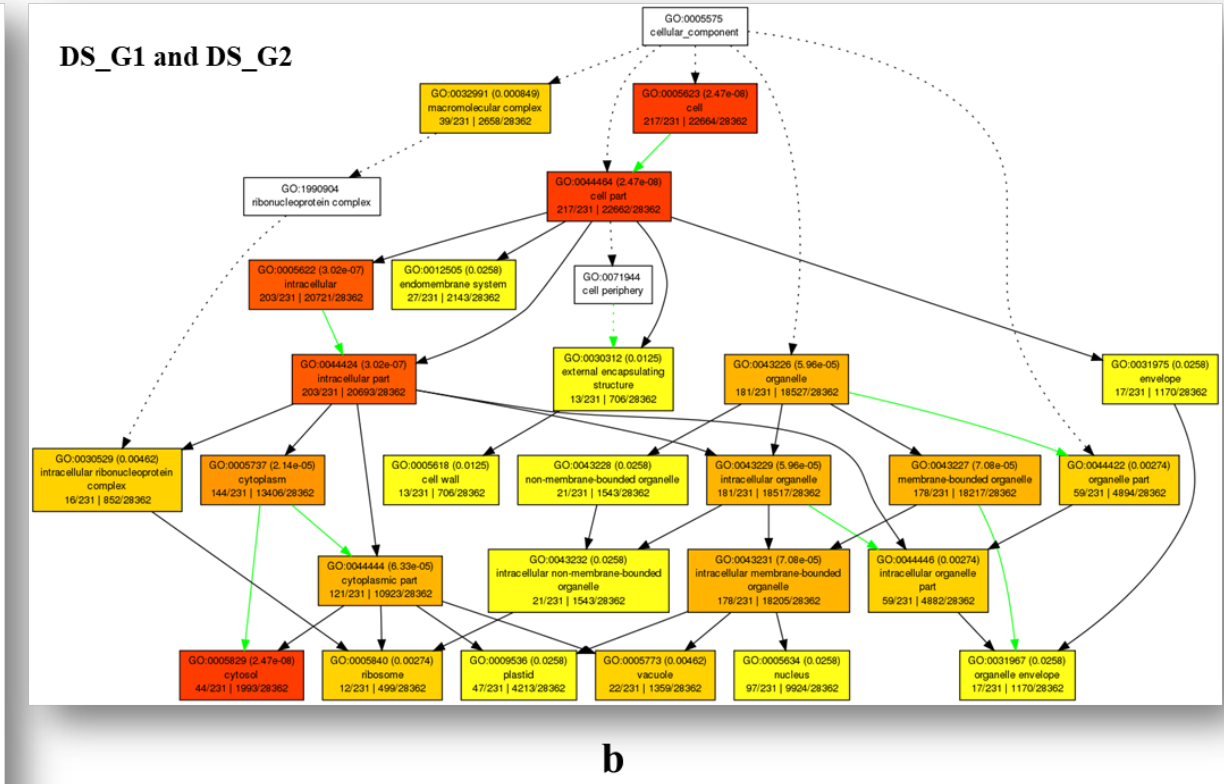
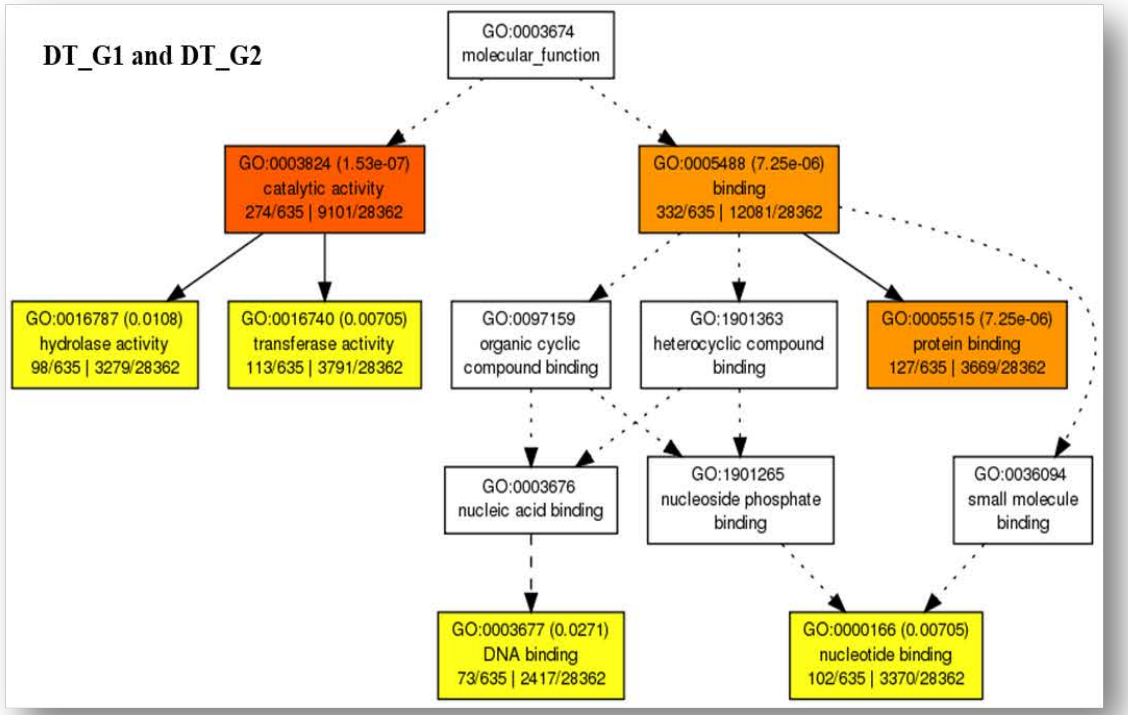
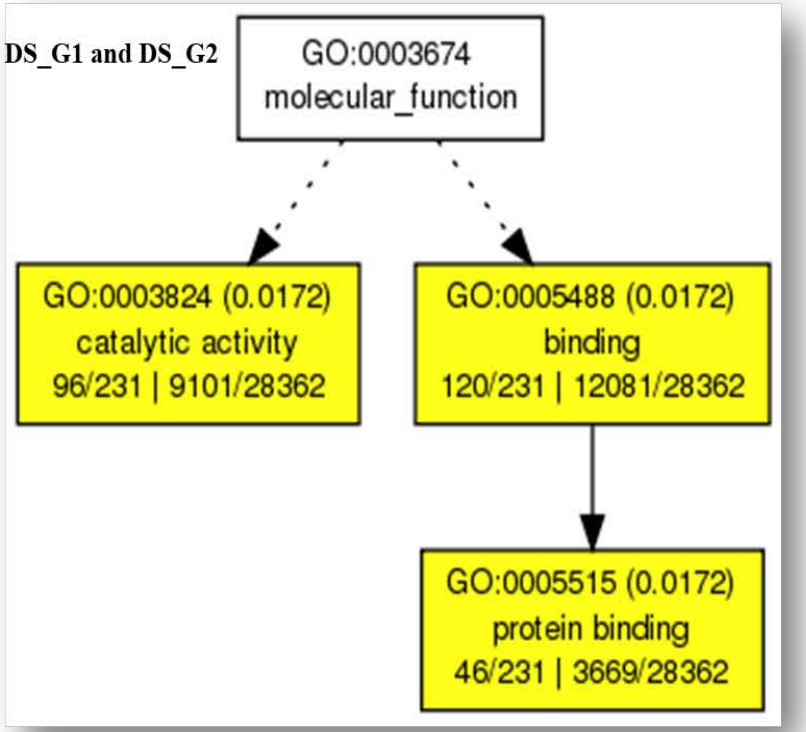


Figure S14

Molecular Function



a



b

Figure S15

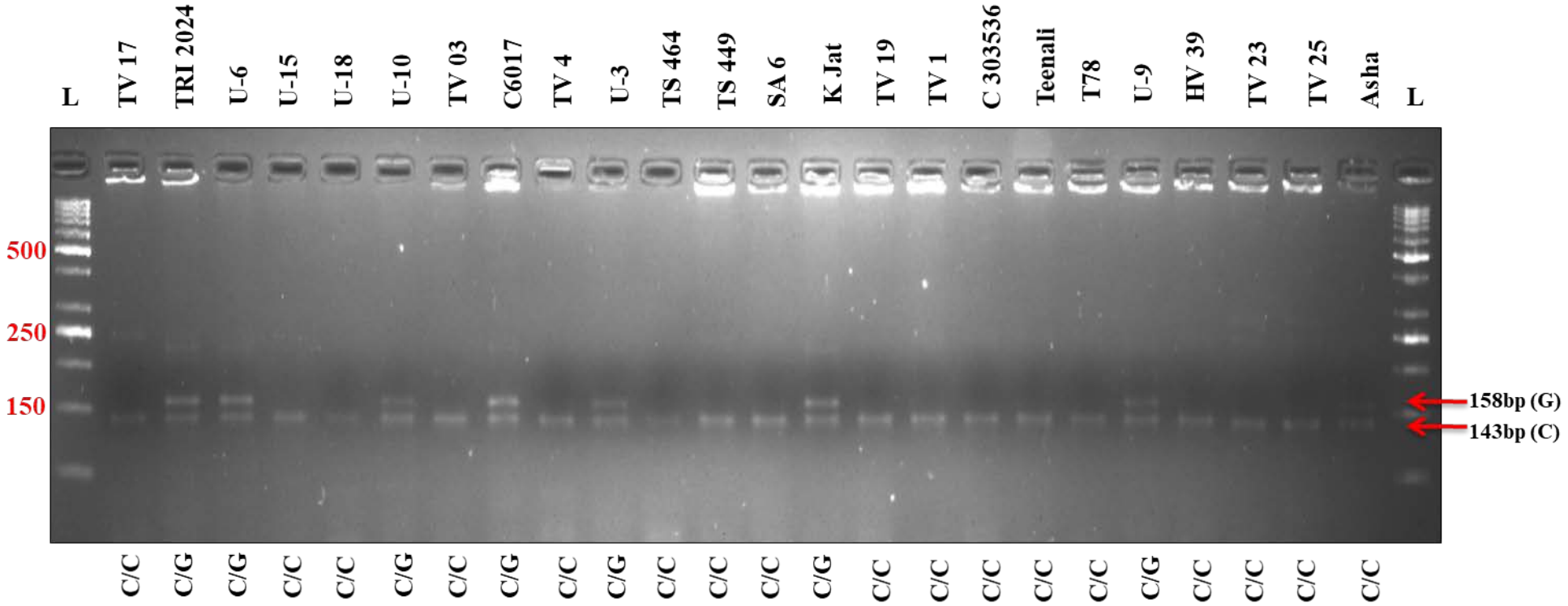

L
TV 17
TRI 2024
U-6
U-15
U-18
U-10
TV 03
C6017
TV 4
U-3
TS 464
TS 449
SA 6
K Jat
TV 19
TV 1
C 303536
Teenali
T78
U-9
HV 39
TV 23
TV 25
Asha
L
500
250
150
158bp (G)
143bp (C)
C/C
C/G
C/G
C/C
C/C
C/G
C/C
C/G
C/C
C/G
C/C
C/C
C/C
C/G
C/C
C/C
C/C
C/C
C/C
C/G
C/C
C/C
C/C
C/C

Figure S15

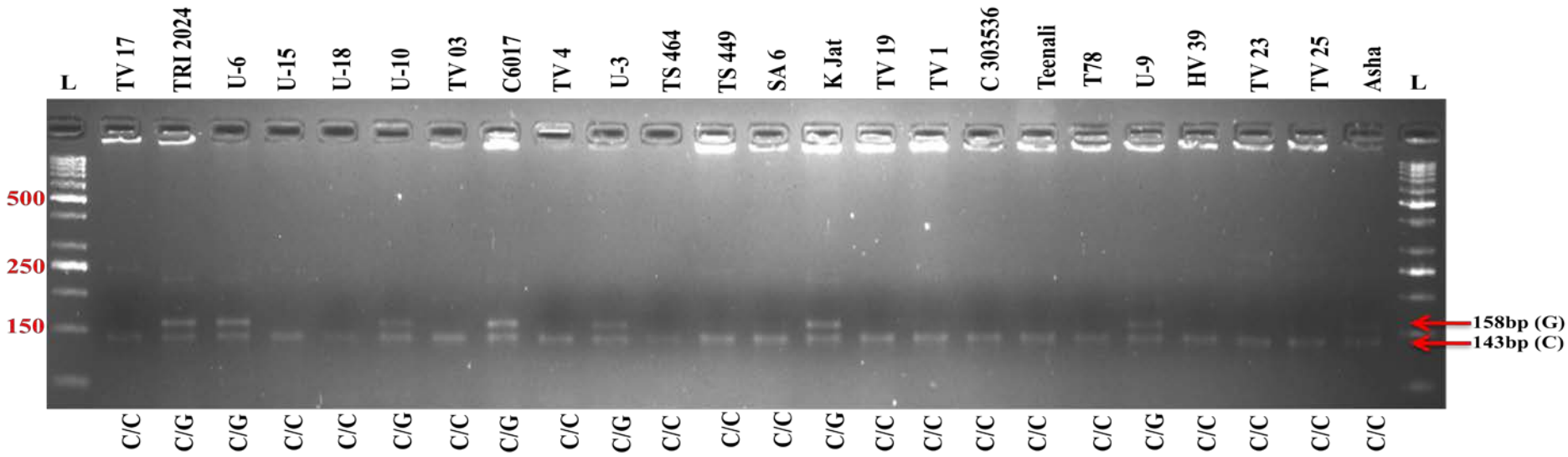

L
TV 17
TRI 2024
U-6
U-15
U-18
U-10
TV 03
C6017
TV 4
U-3
TS 464
TS 449
SA 6
K Jat
TV 19
TV 1
C 303536
Teenali
T78
U-9
HV 39
TV 23
TV 25
Asha
L
500
250
150
158bp (G)
143bp (C)
C/C
C/G
C/G
C/C
C/C
C/G
C/C
C/G
C/C
C/G
C/C
C/C
C/C
C/G
C/C
C/C
C/C
C/C
C/C
C/G
C/C
C/C
C/C
C/C

Table S1: **Summary of *de novo* assembly of Tea transcriptome.**

Quality Filtering	
Raw Reads	140205740
Filtered Reads	123602034
TRINITY Assembly Statistics	
No. of assembled transcripts	67,093
Minimum sequence length	301
Maximum sequence length	17,990
N50 length	1501
Average sequence length	1086
Clustering	
No. of transcripts	54,508
No. of genes	35,379
No. of Isoforms	14,077

Table S2: Summary of functional annotation of transcripts with public databases.

Databases	Transcripts Annotated	%Annotation
nr Blast	52,484	78.20%
Swiss-Prot	41,475	61.08%
TFIIdb	27,204	40.54%
TAIR 10	47,962	71.40%
KEGG	13,454	20.05%

Table S9: Detail of genotypes used for SNP validation using Tetra ARM PCR.

S. No.	Genotypes	Varietal Type
1	TV 17	Assam Hybrid
2	TRI 2024	Assam
3	U 6	Assam
4	U 15	China
5	U 18	Cambod
6	U 10	China
7	TV 03	Assam
8	C 6017	Cambod
9	TV 04	Assam
10	U 3	Assam
11	TS 464	Assam Hybrid
12	TS 449	Assam Hybrid
13	SA 6	Assam
14	K Jat	China Hybrid
15	TV 19	Cambod
16	TV 1	Assam China Hybrid
17	C 303536	China
18	Teenali	Assam
19	T 78	China
20	U 9	Assam
21	HV 39	China
22	TV 23	Cambod
23	TV 25	Cambod
24	K Asha	China Hybrid

Supplementary Legends

Supplementary figures

Figure S1: (a) Overall GO annotation of transcripts categorized into Cellular Component, Biological Processes and Molecular Functions. (b) Tree map representation of transcripts annotated in KEGG Pathway: Area of box represents the number of transcripts mapped to the pathway. (c) Plant TF represented in the form of scattered plot.

Figure S2: Venn diagram representing overall DE transcripts in *de novo* assembly

Figure S3a: (a-g) Clustering of pair-wise DE transcripts in Control and Treatments (FC ≥ 2; p-value < 1e-4), (h) Correlation plot between Control and Treatments.

Figure S3c: (a-e) Clustering of pair-wise DE transcripts in Tolerant and Sensitive (FC ≥ 2; p-value < 1e-4), (f) Correlation plot between Tolerant and Sensitive.

Figure S4: GO enrichment analysis of up-regulated transcripts in DT_G1. (a) Molecular Function; (b) Biological processes (C) Cellular component.

Figure S5: GO enrichment analysis of up-regulated transcripts in DT_G2. (a) Molecular Function; (b) Biological processes (C) Cellular component.

Figure S6: GO enrichment analysis of up-regulated transcripts in DS_G1. (a) Molecular Function; (b) Biological processes (C) Cellular component.

Figure S7: GO enrichment analysis of up-regulated transcripts in DS_G2. (a) Molecular Function; (b) Biological processes (C) Cellular component.

Figure S8: KEGG enrichment analysis[56] of up-regulated transcripts in Flavonoid pathway in (a) Tolerant genotypes (DT_G1_T & DT_G2_T) and (b) Sensitive genotypes (DS_G1_T & DS_G2_T) (www.kegg.jp/kegg/kegg1.html).

Figure S9: KEGG enrichment analysis[56] of up-regulated transcripts in Biosynthesis of amino acids in (a) Tolerant genotypes (DT_G1_T & DT_G2_T) and (b) Sensitive genotypes (DS_G1_T & DS_G2_T) (www.kegg.jp/kegg/kegg1.html).

Figure S10: KEGG enrichment analysis[56] of up-regulated transcripts in porphyrin and chlorophyll metabolism in (a) Tolerant genotypes (DT_G1_T & DT_G2_T) and (b) Sensitive genotypes (DS_G1_T & DS_G2_T) (www.kegg.jp/kegg/kegg1.html).

Figure S11: Flowchart of total SNPs identified from the transcriptome data and

Figure S12: GO enrichment of transcripts containing SNPs; Biological processes (a) Tolerant genotypes (DT_G1_T & DT_G2_T) and (b) Sensitive genotypes (DS_G1_T & DS_G2_T)

Figure S13: GO enrichment of transcripts containing SNPs; Cellular Component (a) Tolerant genotypes (DT_G1_T & DT_G2_T) and (b) Sensitive genotypes (DS_G1_T & DS_G2_T)

Figure S14: GO enrichment of transcripts containing SNPs; Molecular Functions (a) Tolerant genotypes (DT_G1_T & DT_G2_T) and (b) Sensitive genotypes (DS_G1_T & DS_G2_T)

Figure S15: Validation of SNPs using tetra primer ARM-PCR.

Supplementary Tables

Supplementary Table S1: Summary of *de novo* assembly of Tea transcriptome.

Supplementary Table S2: Summary of functional annotation of transcripts with public databases.

Supplementary Table S3: Functional annotation of assembled transcripts with multifarious databases and tea reference genome localisation of transcripts; GO annotation; KEGG annotation.

Supplementary Table S4: Differential Gene Expression extracted from *de novo* assembly using EdgeR tools. DGE between C 6017 vs TRI 2024; TRI 2024 vs TV 03; C 6017 vs TV 17; TV 03 vs TV 17.

Supplementary Table S5: Tea reference genome based differential gene expression using tuxedo pipeline.

Supplementary Table S6a: Shortlisted differentially expressed transcripts of targeted pathways in response to drought stress among Control and Treated plants.

Supplementary Table S6b: Shortlisted differentially expressed transcripts of targeted pathways in response to drought stress among Tolerant and Sensitive genotypes.

Supplementary Table S7: List of primer sequences of selected transcripts used for quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR) analysis.

Supplementary Table S8: Non-synonymous SNPs identification in CDS region of denovo assembled transcripts.

Supplementary Table S9: Detail of genotypes used for SNP validation using Tetra ARM PCR**.**

Supplementary Table S10: Synthesized and validated 37 SNPs, identified in drought stress responsive transcripts.