

# Gene Family Expansion, Gains and Losses Jul2021

oasisia 19  
osedax 20  
riftia 24  
paraescarpia 23  
lamellibrachia 14  
owenia 21  
capitella 5

## Expansions

orthogroups\_annotations\_expanded\_oasisia.sh

```
1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
   ul2021/expansions/oasisia
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
   l2021/expansions/oasisia
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=72:00:0
8  #$ -l highmem
9
10 cut -f 1,7 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
   grep 0alv | cut -f 1 > gene_families_expanded_oasisia.txt #famil
   ies expanded in oasisia
11 fgrep -f gene_families_expanded_oasisia.txt ../../Orthogroups.cs
   v > gene_families_expanded_oasisia.csv
12
13 cut -f 1,19 gene_families_expanded_oasisia.csv > orthogroups_gen
```

```

e_IDs_expanded_oasisia.txt
14 sed 's/0alv|//g' orthogroups_gene_IDs_expanded_oasisia.txt > orthogroups_gene_IDs_expanded_oasisia_OK.txt
15
16 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_number" > orthogroups_annotations_expanded_oasisia.csv
17 while read line; do
18     genes=$(cut -f 2 <<< "$line")
19     echo $genes
20     orthogroup_ID=$(cut -f 1 <<< "$line")
21     echo $orthogroup_ID
22     if [[ "$genes" == OALV* ]]
23     then
24         IFS=', '      # space is set as delimiter
25         read -ra ADDR <<< "$genes"  # str is read into an array as tokens separated by IFS
26         for gene in "${ADDR[@]}"; do
27             annotations=$(cut -f 13,14,15,18,20,24 ../../oasisia_annotation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
28             echo $orthogroup_ID$'\t'"oasisia"$'\t'$annotations >> orthogroups_annotations_expanded_oasisia.csv
29         done
30     else
31         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'"" >> orthogroups_annotations_expanded_oasisia.csv
32     fi
33 done < orthogroups_gene_IDs_expanded_oasisia_OK.txt
34
35 echo "Orthogroup"$'\t'"Panther_annotation" > orthogroups_mostAbundantAnnotation_expanded_oasisia.csv
36 while read line; do

```

```

37     orthogroup_ID=$(cut -f 1 <<< "$line")
38     annotation=$(fgrep $orthogroup_ID orthogroups_annotat
expanded_oasisia.csv | cut -f 7 | sort | uniq -c | sort -r | awk
'{$1=""; print $0}' | head -1)
39     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA
bundantAnnotation_expanded_oasisia.csv
40 done < gene_families_expanded_oasisia.csv

```

#### orthogroups\_annotat ions\_expanded\_osedax.sh

```

1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
ul2021/expansions/osedax
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
l2021/expansions/osedax
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=72:00:0
8  #$ -l highmem
9
10 cut -f 1,7 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
grep Ofra | cut -f 1 > gene_families_expanded_osedax.txt #famili
es expanded in oasisia
11 fgrep -f gene_families_expanded_osedax.txt ../../Orthogroups.csv
> gene_families_expanded_osedax.csv
12
13 cut -f 1,20 gene_families_expanded_osedax.csv > orthogroups_gene
_IDs_expanded_osedax.txt
14 sed 's/Ofra|//g' orthogroups_gene_IDs_expanded_osedax.txt > orth
ogroups_gene_IDs_expanded_osedax_OK.txt
15
16 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t'
"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu

```

```

mber" > orthogroups_annotations_expanded_osedax.csv
17 while read line; do
18     genes=$(cut -f 2 <<< "$line")
19     echo $genes
20     orthogroup_ID=$(cut -f 1 <<< "$line")
21     echo $orthogroup_ID
22     if [[ "$genes" == OFRA* ]]
23     then
24         IFS=', '      # space is set as delimiter
25         read -ra ADDR <<< "$genes"  # str is read into an array a
s tokens separated by IFS
26         for gene in "${ADDR[@]}"; do
27             annotations=$(cut -f 13,14,15,18,20,24 ../osedax_annot
ation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
28             echo $orthogroup_ID$'\t'"osedax"$'\t'$annotations >> ort
hogroups_annotations_expanded_osedax.csv
29         done
30     else
31         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_expanded_osedax.csv
32     fi
33 done < orthogroups_gene_IDs_expanded_osedax_OK.txt
34
35 echo "Orthogroup"$'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_expanded_osedax.csv
36 while read line; do
37     orthogroup_ID=$(cut -f 1 <<< "$line")
38     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_exp
anded_osedax.csv | cut -f 7 | sort | uniq -c | sort -r | awk
'{$1=""; print $0}' | head -1)
39     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA
bundantAnnotation_expanded_osedax.csv

```

```
40 done < gene_families_expanded_osedax.csv
```

#### orthogroups\_annotations\_expanded\_riftia.sh

```
1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
   ul2021/expansions/riftia
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
   l2021/expansions/riftia
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=72:00:0
8  #$ -l highmem
9
10 cut -f 1,7 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
   grep Rpac | cut -f 1 > gene_families_expanded_riftia.txt #famili
   es expanded in riftia
11 fgrep -f gene_families_expanded_riftia.txt ../../Orthogroups.csv
   > gene_families_expanded_riftia.csv
12
13 cut -f 1,24 gene_families_expanded_riftia.csv > orthogroups_gene
   _IDs_expanded_riftia.txt
14 sed 's/Rpac|//g' orthogroups_gene_IDs_expanded_riftia.txt > orth
   ogroups_gene_IDs_expanded_riftia_OK.txt
15
16 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t'
   '"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
   mber" > orthogroups_annotations_expanded_riftia.csv
17 while read line; do
18     genes=$(cut -f 2 <<< "$line")
19     echo $genes
20     orthogroup_ID=$(cut -f 1 <<< "$line")
```

```

21     echo $orthogroup_ID
22     if [[ "$genes" == RPAC* ]]
23     then
24         IFS=', '      # space is set as delimiter
25         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
26         for gene in "${ADDR[@]}"; do
27             annotations=$(cut -f 13,14,15,18,20,24 ../../riftia_annot
ation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
28             echo $orthogroup_ID$'\t'"riftia"$'\t'$annotations >> ort
hogroups_annotations_expanded_riftia.csv
29         done
30     else
31         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_expanded_riftia.csv
32     fi
33 done < orthogroups_gene_IDs_expanded_riftia_OK.txt
34
35 echo "Orthogroup"$'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_expanded_riftia.csv
36 while read line; do
37     orthogroup_ID=$(cut -f 1 <<< "$line")
38     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_exp
anded_riftia.csv | cut -f 7 | sort | uniq -c | sort -r | awk
'{$1=""; print $0}' | head -1)
39     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA
bundantAnnotation_expanded_riftia.csv
40 done < gene_families_expanded_riftia.csv

```

orthogroups\_annotations\_expanded\_lamellibrachia.sh

```

1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J

```

```

ul2021/expansions/lamellibrachia
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
   l2021/expansions/lamellibrachia
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=72:00:0
8  #$ -l highmem
9
10 cut -f 1,7 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
   grep Lluy | cut -f 1 > gene_families_expanded_lamellibrachia.txt
   #families expanded in riftia
11 fgrep -f gene_families_expanded_lamellibrachia.txt ../../Orthogr
   oups.csv > gene_families_expanded_lamellibrachia.csv
12
13 cut -f 1,14 gene_families_expanded_lamellibrachia.csv > orthogro
   ups_gene_IDs_expanded_lamellibrachia.txt
14 sed 's/Lluy|//g' orthogroups_gene_IDs_expanded_lamellibrachia.tx
   t > orthogroups_gene_IDs_expanded_lamellibrachia_OK.txt
15
16 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t'
   "'GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
   mber" > orthogroups_annotations_expanded_lamellibrachia.csv
17 while read line; do
18     genes=$(cut -f 2 <<< "$line")
19     echo $genes
20     orthogroup_ID=$(cut -f 1 <<< "$line")
21     echo $orthogroup_ID
22     if [[ "$genes" == FUN* ]]
23     then
24         IFS=', ' # space is set as delimiter
25         read -ra ADDR <<< "$genes" # str is read into an array a

```

s tokens separated by IFS

```
26     for gene in "${ADDR[@]"}; do
27         annotations=$(cut -f 13,14,15,18,20,24 ../../lamellibrach
ia_annotation_Feb2021_TrinoPantherK0_OK.xls | fgrep $gene)
28         echo $orthogroup_ID$'\t'"lamellibrachia"$'\t'$annotation
s >> orthogroups_annotations_expanded_lamellibrachia.csv
29     done
30     else
31         echo $orthogroup_ID$'\t'"'"$'\t'"'"$'\t'"'"$'\t'"'"$'\t'"'"$'\t
'"'"$'\t'"'" >> orthogroups_annotations_expanded_lamellibrachia.cs
v
32     fi
33 done < orthogroups_gene_IDs_expanded_lamellibrachia_OK.txt
34
35 echo "Orthogroup"$'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_expanded_lamellibrachia.csv
36 while read line; do
37     orthogroup_ID=$(cut -f 1 <<< "$line")
38     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_exp
anded_lamellibrachia.csv | cut -f 7 | sort | uniq -c | sort -r
| awk '{ $1="" ; print $0 }' | head -1)
39     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA
bundantAnnotation_expanded_lamellibrachia.csv
40 done < gene_families_expanded_lamellibrachia.csv
```

orthogroups\_annotations\_expanded\_paraescarpia.sh

```
1 #!/bin/bash
2 #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
ul2021/expansions/paraescarpia
3 #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
l2021/expansions/paraescarpia
4 #$ -j y
5 #$ -pe smp 1
```



```

6  #$ -l h_vmem=100G
7  #$ -l h_rt=72:00:0
8  #$ -l highmem
9
10 cut -f 1,7 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
    grep Pech | cut -f 1 > gene_families_expanded_paraescarpia.txt #
    families expanded in riftia
11 fgrep -f gene_families_expanded_paraescarpia.txt ../../Orthogrou
    ps.csv > gene_families_expanded_paraescarpia.csv
12
13 cut -f 1,23 gene_families_expanded_paraescarpia.csv > orthogroup
    s_gene_IDs_expanded_paraescarpia.txt
14 sed 's/Pech|//g' orthogroups_gene_IDs_expanded_paraescarpia.txt
    > orthogroups_gene_IDs_expanded_paraescarpia_OK.txt
15
16 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\
    t'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
    mber" > orthogroups_annotations_expanded_paraescarpia.csv
17 while read line; do
18     genes=$(cut -f 2 <<< "$line")
19     echo $genes
20     orthogroup_ID=$(cut -f 1 <<< "$line")
21     echo $orthogroup_ID
22     if [[ "$genes" == nbis* ]]
23     then
24         IFS=', '      # space is set as delimiter
25         read -ra ADDR <<< "$genes" # str is read into an array a
            s tokens separated by IFS
26         for gene in "${ADDR[@]}; do
27             annotations=$(cut -f 13,14,15,18,20,24 ../../paraescarpia
                _annotation_Jun2021_TrinoPantherK0.xls | fgrep $gene)
28             echo $orthogroup_ID$'\t'"paraescarpia"$'\t'"$annotations

```

```

>> orthogroups_annotations_expanded_paraescarpia.csv
29     done
30     else
31         echo $orthogroup_ID$'\t'$'\t'$'\t'$'\t'$'\t'$'\t
        '$'\t'$'\t' >> orthogroups_annotations_expanded_paraescarpia.csv
32     fi
33 done < orthogroups_gene_IDs_expanded_paraescarpia_0K.txt
34
35 echo "Orthogroup"$'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_expanded_paraescarpia.csv
36 while read line; do
37     orthogroup_ID=$(cut -f 1 <<< "$line")
38     annotation=$(fgrep $orthogroup_ID orthogroups_annotati
ons_expanded_paraescarpia.csv | cut -f 7 | sort | uniq -c | sort -r |
awk '{ $1="" ; print $0 }' | head -1)
39     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA
bundantAnnotation_expanded_paraescarpia.csv
40 done < gene_families_expanded_paraescarpia.csv

```

## Gains

orthogroups\_annotations\_originated\_oasisia.sh

```

1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
l2021/gains/oasisia
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
l2021/gains/oasisia
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=72:00:0
8  #$ -l highmem

```

```

9
10 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
    grep -w Oalv | cut -f 1 > gene_families_originated_oasisia.txt #
    families originated in oasisia
11 fgrep -f gene_families_originated_oasisia.txt ../../Orthogroups.
    csv > gene_families_originated_oasisia.csv
12
13 cut -f 1,19 gene_families_originated_oasisia.csv > orthogroups_g
    ene_IDs_originated_oasisia.txt
14 sed 's/Oalv|//g' orthogroups_gene_IDs_originated_oasisia.txt > o
    rthogroups_gene_IDs_originated_oasisia_OK.txt
15
16 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t'
    t'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
    mber" > orthogroups_annotations_originated_oasisia.csv
17 while read line; do
18     genes=$(cut -f 2 <<< "$line")
19     echo $genes
20     orthogroup_ID=$(cut -f 1 <<< "$line")
21     echo $orthogroup_ID
22     if [[ "$genes" == OALV* ]]
23     then
24         IFS=', '      # space is set as delimiter
25         read -ra ADDR <<< "$genes"    # str is read into an array a
    s tokens separated by IFS
26         for gene in "${ADDR[@]}; do
27             annotations=$(cut -f 13,14,15,18,20,24 ../../oasisia_anno
    tation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
28             echo $orthogroup_ID$'\t'"oasisia"$'\t'$annotations >> or
    thogroups_annotations_originated_oasisia.csv
29         done
30     else

```

```

31     echo $orthogroup_ID$'\t'"$'\t'"$'\t'"$'\t'"$'\t'"$'\t'
    ""$'\t'" >> orthogroups_annotations_originated_oasisia.csv
32     fi
33 done < orthogroups_gene_IDs_originated_oasisia_OK.txt
34
35 echo "Orthogroup"$'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_originated_oasisia.csv
36 while read line; do
37     orthogroup_ID=$(cut -f 1 <<< "$line")
38     annotation=$(fgrep $orthogroup_ID orthogroups_annotati
ons_originated_oasisia.csv | cut -f 7 | sort | uniq -c | sort -r | awk
'{$1=""; print $0}' | head -1)
39     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA
bundantAnnotation_originated_oasisia.csv
40 done < gene_families_originated_oasisia.csv

```

#### orthogroups\_annotations\_originated\_osedax.sh

```

1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
ul2021/gains/osedax
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
l2021/gains/osedax
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=72:00:0
8  #$ -l highmem
9
10 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
    grep -w Ofra | cut -f 1 > gene_families_originated_osedax.txt #f
amilies originated in oasisia

```

```

11 fgrep -f gene_families_originated_osedax.txt ../../Orthogroups.c
sv > gene_families_originated_osedax.csv
12
13 cut -f 1,20 gene_families_originated_osedax.csv > orthogroups_ge
ne_IDs_originated_osedax.txt
14 sed 's/Ofra|//g' orthogroups_gene_IDs_originated_osedax.txt > or
thogroups_gene_IDs_originated_osedax_OK.txt
15
16 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t'
t'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
mber" > orthogroups_annotations_originated_osedax.csv
17 while read line; do
18     genes=$(cut -f 2 <<< "$line")
19     echo $genes
20     orthogroup_ID=$(cut -f 1 <<< "$line")
21     echo $orthogroup_ID
22     if [[ "$genes" == OFRA* ]]
23     then
24         IFS=', '      # space is set as delimiter
25         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
26         for gene in "${ADDR[@]}; do
27             annotations=$(cut -f 13,14,15,18,20,24 ../../osedax_annot
ation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
28             echo $orthogroup_ID$'\t'"osedax"$'\t'$annotations >> ort
hogroups_annotations_originated_osedax.csv
29         done
30     else
31         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'
'""$'\t'"" >> orthogroups_annotations_originated_osedax.csv
32     fi
33 done < orthogroups_gene_IDs_originated_osedax_OK.txt

```

```

34
35 echo "Orthogroup"${'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_originated_osedax.csv
36 while read line; do
37     orthogroup_ID=$(cut -f 1 <<< "$line")
38     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_ori
ginated_osedax.csv | cut -f 7 | sort | uniq -c | sort -r | awk
'{$1=""; print $0}' | head -1)
39     echo $orthogroup_ID${'\t'}$annotation >> orthogroups_mostA
bundantAnnotation_originated_osedax.csv
40 done < gene_families_originated_osedax.csv

```

#### orthogroups\_annotations\_originated\_riftia.sh

```

1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
ul2021/gains/riftia
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
l2021/gains/riftia
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=72:00:0
8  #$ -l highmem
9
10 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
grep -w Rpac | cut -f 1 > gene_families_originated_riftia.txt #f
amilies originated in riftia
11 fgrep -f gene_families_originated_riftia.txt ../../Orthogroups.c
sv > gene_families_originated_riftia.csv
12
13 cut -f 1,24 gene_families_originated_riftia.csv > orthogroups_ge
ne_IDs_originated_riftia.txt
14 sed 's/Rpac|//g' orthogroups_gene_IDs_originated_riftia.txt > or

```

thogroups\_gene\_IDs\_originated\_riftia\_0K.txt

15

```
16 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t'
   t'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
   mber" > orthogroups_annotations_originated_riftia.csv
```

```
17 while read line; do
```

```
18 genes=$(cut -f 2 <<< "$line")
```

```
19 echo $genes
```

```
20 orthogroup_ID=$(cut -f 1 <<< "$line")
```

```
21 echo $orthogroup_ID
```

```
22 if [ "$genes" == RPAC* ]
```

23 then

```
24 IFS=' ' # space is set as delimiter
```

```
25 read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
```

```
26 for gene in "${ADDR[@]"}"; do
```

```
27 annotations=$(cut -f 13,14,15,18,20,24 ../../riftia_annotation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
```

```
28 echo $orthogroup_ID$'\t'"riftia"$'\t'$annotations >> ort
hogroups_annotations_originated_riftia.csv
```

29 done

```
30 else
```

```
31 echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'
    ""$'\t'"" >> orthogroups_annotations_originated_riftia.csv
```

32 | fi

```
33 done < orthogroups_gene_IDs_originated_riftia_OK.txt
```

34

```
35 echo "Orthogroup"`${t}"Panther_annotation" > orthogroups_mostAbu  
ndantAnnotation_originated_riftia.csv
```

```
36 while read line; do
```

```
37 orthogroup ID=$(cut -f 1 <<< "$line")
```

```
38 annotation=$(fgrep $orthogroup_ID orthogroups_annotations_ori
```

```

    originated_riftia.csv | cut -f 7 | sort | uniq -c | sort -r | awk
    '{$1=""; print $0}' | head -1)
39     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA
    bundantAnnotation_originated_riftia.csv
40 done < gene_families_originated_riftia.csv

```

#### orthogroups\_annotations\_originated\_lamellibrachia.sh

```

1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
    ul2021/gains/lamellibrachia
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
    l2021/gains/lamellibrachia
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=72:00:0
8  #$ -l highmem
9
10 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
    grep -w Lluy | cut -f 1 > gene_families_originated_lamellibrachi
    a.txt #families originated in riftia
11 fgrep -f gene_families_originated_lamellibrachia.txt ../../Ortho
    groups.csv > gene_families_originated_lamellibrachia.csv
12
13 cut -f 1,14 gene_families_originated_lamellibrachia.csv > orthog
    rroups_gene_IDs_originated_lamellibrachia.txt
14 sed 's/Lluy|//g' orthogroups_gene_IDs_originated_lamellibrachia.
    txt > orthogroups_gene_IDs_originated_lamellibrachia_OK.txt
15
16 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t'
    t'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
    mber" > orthogroups_annotations_originated_lamellibrachia.csv
17 while read line; do

```



```

18 genes=$(cut -f 2 <<< "$line")
19 echo $genes
20 orthogroup_ID=$(cut -f 1 <<< "$line")
21 echo $orthogroup_ID
22 if [[ "$genes" == FUN* ]]
23 then
24     IFS=', '      # space is set as delimiter
25     read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
26     for gene in "${ADDR[@]}"; do
27         annotations=$(cut -f 13,14,15,18,20,24 ../../lamellibrach
ia_annotation_Feb2021_TrinoPantherK0_OK.xls | fgrep $gene)
28         echo $orthogroup_ID$'\t'"lamellibrachia"$'\t'$annotations >> orthogroups_annotations_originated_lamellibrachia.csv
29     done
30 else
31     echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'
""$'\t'"" >> orthogroups_annotations_originated_lamellibrachia.
csv
32 fi
33 done < orthogroups_gene_IDs_originated_lamellibrachia_OK.txt
34
35 echo "Orthogroup"$'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_originated_lamellibrachia.csv
36 while read line; do
37     orthogroup_ID=$(cut -f 1 <<< "$line")
38     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_ori
ginated_lamellibrachia.csv | cut -f 7 | sort | uniq -c | sort -
r | awk '{ $1="" ; print $0 }' | head -1)
39     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA
bundantAnnotation_originated_lamellibrachia.csv
40 done < gene_families_originated_lamellibrachia.csv

```

## orthogroups\_annotations\_originated\_paraescarpia.sh

```
1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
   ul2021/gains/paraescarpia
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
   l2021/gains/paraescarpia
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=72:00:0
8  #$ -l highmem
9
10 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
   grep -w Pech | cut -f 1 > gene_families_originated_paraescarpia.
   txt #families originated in riftia
11 fgrep -f gene_families_originated_paraescarpia.txt ../../Orthogr
   oups.csv > gene_families_originated_paraescarpia.csv
12
13 cut -f 1,23 gene_families_originated_paraescarpia.csv > orthogro
   ups_gene_IDs_originated_paraescarpia.txt
14 sed 's/Pech|//g' orthogroups_gene_IDs_originated_paraescarpia.tx
   t > orthogroups_gene_IDs_originated_paraescarpia_OK.txt
15
16 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t'
   t'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
   mber" > orthogroups_annotations_originated_paraescarpia.csv
17 while read line; do
18     genes=$(cut -f 2 <<< "$line")
19     echo $genes
20     orthogroup_ID=$(cut -f 1 <<< "$line")
21     echo $orthogroup_ID
```

```

22     if [[ "$genes" == nbis* ]]
23     then
24         IFS=', '      # space is set as delimiter
25         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
26         for gene in "${ADDR[@]}; do
27             annotations=$(cut -f 13,14,15,18,20,24 ../../paraescarpia
_annotation_Jun2021_TrinoPantherK0.xls | fgrep $gene)
28             echo $orthogroup_ID$'\t'"paraescarpia"$'\t'$annotations
>> orthogroups_annotations_originated_paraescarpia.csv
29         done
30     else
31         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_originated_paraescarpia.cs
v
32     fi
33 done < orthogroups_gene_IDs_originated_paraescarpia_OK.txt
34
35 echo "Orthogroup"$'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_originated_paraescarpia.csv
36 while read line; do
37     orthogroup_ID=$(cut -f 1 <<< "$line")
38     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_ori
ginated_paraescarpia.csv | cut -f 7 | sort | uniq -c | sort -r
| awk '{ $1="" ; print $0 }' | head -1)
39     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA
bundantAnnotation_originated_paraescarpia.csv
40 done < gene_families_originated_paraescarpia.csv

```

orthogroups\_annotations\_originated\_siboglinidae.sh

```

1 #!/bin/bash

```

```
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
   ul2021/gains/siboglinidae
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
   l2021/gains/siboglinidae
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=140:00:0
8  #$ -l highmem
9
10 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
    grep -w Siboglinidae | cut -f 1 > gene_families_originated_sibog
    linidae.txt #families originated in oasisia
11 fgrep -f gene_families_originated_siboglinidae.txt ../../Orthogr
    oups.csv > gene_families_originated_siboglinidae.csv
12
13 cut -f 1,19 gene_families_originated_siboglinidae.csv > orthogro
    ups_gene_IDs_originated_siboglinidae_oasisia.txt
14 sed 's/0a1v|//g' orthogroups_gene_IDs_originated_siboglinidae_oa
    sisia.txt > orthogroups_gene_IDs_originated_siboglinidae_oasisia
    _OK.txt
15 cut -f 1,20 gene_families_originated_siboglinidae.csv > orthogro
    ups_gene_IDs_originated_siboglinidae_osedax.txt
16 sed 's/0fra|//g' orthogroups_gene_IDs_originated_siboglinidae_os
    edax.txt > orthogroups_gene_IDs_originated_siboglinidae_osedax_0
    K.txt
17 cut -f 1,24 gene_families_originated_siboglinidae.csv > orthogro
    ups_gene_IDs_originated_siboglinidae_riftia.txt
18 sed 's/Rpac|//g' orthogroups_gene_IDs_originated_siboglinidae_ri
    ftia.txt > orthogroups_gene_IDs_originated_siboglinidae_riftia_0
    K.txt
19 cut -f 1,14 gene_families_originated_siboglinidae.csv > orthogro
    ups_gene_IDs_originated_siboglinidae_lamellibrachia.txt
```

```

20 sed 's/Lluy|//g' orthogroups_gene_IDs_originated_siboglinidae_la
    mellibrachia.txt > orthogroups_gene_IDs_originated_siboglinidae_
    lamellibrachia_OK.txt
21 cut -f 1,23 gene_families_originated_siboglinidae.csv > orthogro
    ups_gene_IDs_originated_siboglinidae_paraescarpia.txt
22 sed 's/Pech|//g' orthogroups_gene_IDs_originated_siboglinidae_pa
    raescarpia.txt > orthogroups_gene_IDs_originated_siboglinidae_pa
    raescarpia_OK.txt
23
24
25 while read line; do
26     genes=$(cut -f 2 <<< "$line")
27     echo $genes
28     orthogroup_ID=$(cut -f 1 <<< "$line")
29     echo $orthogroup_ID
30     if [[ "$genes" == OALV* ]]
31     then
32         IFS=', '      # space is set as delimiter
33         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
34         for gene in "${ADDR[@]}; do
35             annotations=$(cut -f 13,14,15,18,20,24 ../../oasisia_anno
                tion_Jan2021_TrinoPantherK0.xls | fgrep $gene)
36             echo $orthogroup_ID$'\t'"oasisia"$'\t'$annotations >> or
                thogroups_annotations_originated_siboglinidae_oasisia.csv
37         done
38     else
39         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
                '""$'\t'"" >> orthogroups_annotations_originated_siboglinidae_oa
                sisia.csv
40     fi
41 done < orthogroups_gene_IDs_originated_siboglinidae_oasisia_OK.t

```

xt

```
42
43 while read line; do
44     genes=$(cut -f 2 <<< "$line")
45     echo $genes
46     orthogroup_ID=$(cut -f 1 <<< "$line")
47     echo $orthogroup_ID
48     if [[ "$genes" == OFRA* ]]
49     then
50         IFS=', '      # space is set as delimiter
51         read -ra ADDR <<< "$genes"  # str is read into an array a
s tokens separated by IFS
52         for gene in "${ADDR[@]}; do
53             annotations=$(cut -f 13,14,15,18,20,24 ../../osedax_annot
ation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
54             echo $orthogroup_ID$'\t'"osedax"$'\t'$annotations >> ort
hogroups_annotations_originated_siboglinidae_osedax.csv
55         done
56     else
57         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_originated_siboglinidae_os
edax.csv
58     fi
59 done < orthogroups_gene_IDs_originated_siboglinidae_osedax_OK.tx
t
60
61 while read line; do
62     genes=$(cut -f 2 <<< "$line")
63     echo $genes
64     orthogroup_ID=$(cut -f 1 <<< "$line")
65     echo $orthogroup_ID
```

```

66     if [[ "$genes" == RPAC* ]]
67     then
68         IFS=', '      # space is set as delimiter
69         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
70         for gene in "${ADDR[@]}"; do
71             annotations=$(cut -f 13,14,15,18,20,24 ../../riftia_annot
ation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
72             echo $orthogroup_ID$'\t'"riftia"$'\t'$annotations >> ort
hogroups_annotations_originated_siboglinidae_riftia.csv
73         done
74     else
75         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_originated_siboglinidae_rif
tia.csv
76     fi
77 done < orthogroups_gene_IDs_originated_siboglinidae_riftia_OK.tx
t
78
79 while read line; do
80     genes=$(cut -f 2 <<< "$line")
81     echo $genes
82     orthogroup_ID=$(cut -f 1 <<< "$line")
83     echo $orthogroup_ID
84     if [[ "$genes" == FUN* ]]
85     then
86         IFS=', '      # space is set as delimiter
87         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
88         for gene in "${ADDR[@]}"; do
89             annotations=$(cut -f 13,14,15,18,20,24 ../../lamellibrach
ia_annotation_Feb2021_TrinoPantherK0_OK.xls | fgrep $gene)

```

```

90     echo $orthogroup_ID$'\t'"lamellibrachia"$'\t'$annotation
s >> orthogroups_annotations_originated_siboglinidae_lamellibrac
hia.csv
91     done
92     else
93     echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_originated_siboglinidae_la
mellibrachia.csv
94     fi
95 done < orthogroups_gene_IDs_originated_siboglinidae_lamellibrach
ia_OK.txt
96
97 while read line; do
98     genes=$(cut -f 2 <<< "$line")
99     echo $genes
100    orthogroup_ID=$(cut -f 1 <<< "$line")
101    echo $orthogroup_ID
102    if [[ "$genes" == nbis* ]]
103    then
104        IFS=', '      # space is set as delimiter
105        read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
106        for gene in "${ADDR[@]}; do
107            annotations=$(cut -f 13,14,15,18,20,24 ../../paraescarpia
_annotation_Jun2021_TrinoPantherK0.xls | fgrep $gene)
108            echo $orthogroup_ID$'\t'"paraescarpia"$'\t'$annotations
>> orthogroups_annotations_originated_siboglinidae_paraescarpia.
csv
109        done
110    else
111    echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_originated_siboglinidae_pa

```



raescarpia.csv

```
112     fi
113 done < orthogroups_gene_IDs_originated_siboglinidae_paraescarpia
    _OK.txt
114
115 echo "Orthogroup"'\t'"Species"'\t'"GO_term1"'\t'"GO_term2"'\t'
    '\t'"GO_term3"'\t'"gene_ID"'\t'"Panther_annotation"'\t'"KEGG_nu
    mber" > orthogroups_annotations_originated_siboglinidae_ofra_0al
    v_Rpac_Lluy_Pech.csv
116 cat orthogroups_annotations_originated_siboglinidae_oasisia.csv
    orthogroups_annotations_originated_siboglinidae_osedax.csv ortho
    groups_annotations_originated_siboglinidae_riftia.csv orthogroup
    s_annotations_originated_siboglinidae_lamellibrachia.csv orthogr
    oups_annotations_originated_siboglinidae_paraescarpia.csv >> ort
    hogroups_annotations_originated_siboglinidae_ofra_0alv_Rpac_Lluy
    _Pech.csv
117 sort orthogroups_annotations_originated_siboglinidae_ofra_0alv_R
    pac_Lluy_Pech.csv > orthogroups_annotations_originated_siboglini
    dae_ofra_0alv_Rpac_Lluy_Pech_OK.csv
118
119 echo "Orthogroup"'\t'"Panther_annotation" > orthogroups_mostAbu
    ndantAnnotation_originated_siboglinidae.csv
120 while read line; do
121     orthogroup_ID=$(cut -f 1 <<< "$line")
122     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_ori
    genated_siboglinidae_ofra_0alv_Rpac_Lluy_Pech_OK.csv | cut -f 7
    | sed '/^$/d' | sort | uniq -c | sort -r | awk '{ $1="" }; print
    $0}' | head -1)
123     echo $orthogroup_ID'\t'$annotation >> orthogroups_mostA
    bundantAnnotation_originated_siboglinidae.csv
124 done < gene_families_originated_siboglinidae.csv
```

orthogroups\_annotations\_originated\_vestimentifera.sh

```
1 #!/bin/bash
```

```
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
   ul2021/gains/vestmentifera
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
   l2021/gains/vestmentifera
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=140:00:0
8  #$ -l highmem
9
10 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
    grep -w Vestimentifera | cut -f 1 > gene_families_originated_ves
    timentifera.txt #families originated in oasisia
11 fgrep -f gene_families_originated_vestimentifera.txt ../../Ortho
    groups.csv > gene_families_originated_vestimentifera.csv
12
13 cut -f 1,19 gene_families_originated_vestimentifera.csv > orthog
    rous_gene_IDs_originated_vestimentifera_oasisia.txt
14 sed 's/0alv|//g' orthogroups_gene_IDs_originated_vestimentifera_
    oasisia.txt > orthogroups_gene_IDs_originated_vestimentifera_oas
    isia_OK.txt
15 cut -f 1,24 gene_families_originated_vestimentifera.csv > orthog
    rous_gene_IDs_originated_vestimentifera_riftia.txt
16 sed 's/Rpac|//g' orthogroups_gene_IDs_originated_vestimentifera_
    riftia.txt > orthogroups_gene_IDs_originated_vestimentifera_rift
    ia_OK.txt
17 cut -f 1,14 gene_families_originated_vestimentifera.csv > orthog
    rous_gene_IDs_originated_vestimentifera_lamellibrachia.txt
18 sed 's/Lluy|//g' orthogroups_gene_IDs_originated_vestimentifera_
    lamellibrachia.txt > orthogroups_gene_IDs_originated_vestimentif
    era_lamellibrachia_OK.txt
19 cut -f 1,23 gene_families_originated_vestimentifera.csv > orthog
    rous_gene_IDs_originated_vestimentifera_paraescarpia.txt
```

```

20 sed 's/Pech|//g' orthogroups_gene_IDs_originated_vestimentifera_
    paraescarpia.txt > orthogroups_gene_IDs_originated_vestimentifer
    a_paraescarpia_0K.txt
21
22
23 while read line; do
24     genes=$(cut -f 2 <<< "$line")
25     echo $genes
26     orthogroup_ID=$(cut -f 1 <<< "$line")
27     echo $orthogroup_ID
28     if [[ "$genes" == OALV* ]]
29     then
30         IFS=', '      # space is set as delimiter
31         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
32         for gene in "${ADDR[@]}"; do
33             annotations=$(cut -f 13,14,15,18,20,24 ../../oasisia_anno
tation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
34             echo $orthogroup_ID$'\t'"oasisia"$'\t'$annotations >> or
thogroups_annotations_originated_vestimentifera_oasisia.csv
35         done
36     else
37         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_originated_vestimentifera_
oasisia.csv
38     fi
39 done < orthogroups_gene_IDs_originated_vestimentifera_oasisia_0
K.txt
40
41 while read line; do
42     genes=$(cut -f 2 <<< "$line")
43     echo $genes

```

```

44 orthogroup_ID=$(cut -f 1 <<< "$line")
45 echo $orthogroup_ID
46 if [[ "$genes" == RPAC* ]]
47 then
48     IFS=', '      # space is set as delimiter
49     read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
50     for gene in "${ADDR[@]}"; do
51         annotations=$(cut -f 13,14,15,18,20,24 ../../riftia_annot
ation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
52         echo $orthogroup_ID$'\t'"riftia"$'\t'$annotations >> ort
hogroups_annotations_originated_vestimentifera_riftia.csv
53     done
54 else
55     echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_originated_vestimentifera_
riftia.csv
56 fi
57 done < orthogroups_gene_IDs_originated_vestimentifera_riftia_OK.
txt
58
59 while read line; do
60     genes=$(cut -f 2 <<< "$line")
61     echo $genes
62     orthogroup_ID=$(cut -f 1 <<< "$line")
63     echo $orthogroup_ID
64     if [[ "$genes" == FUN* ]]
65     then
66         IFS=', '      # space is set as delimiter
67         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
68         for gene in "${ADDR[@]}"; do

```

```

69     annotations=$(cut -f 13,14,15,18,20,24 ../../lamellibrach
ia_annotation_Feb2021_TrinoPantherK0_OK.xls | fgrep $gene)
70     echo $orthogroup_ID$'\t'"lamellibrachia"$'\t'$annotations
s >> orthogroups_annotations_originated_vestimentifera_lamellibr
achia.csv
71     done
72     else
73     echo $orthogroup_ID$'\t'"$'\t'"$'\t'"$'\t'"$'\t'"$'\t
'"'"$'\t'" >> orthogroups_annotations_originated_vestimentifera_
lamellibrachia.csv
74     fi
75 done < orthogroups_gene_IDs_originated_vestimentifera_lamellibra
chia_OK.txt
76
77 while read line; do
78     genes=$(cut -f 2 <<< "$line")
79     echo $genes
80     orthogroup_ID=$(cut -f 1 <<< "$line")
81     echo $orthogroup_ID
82     if [[ "$genes" == nbis* ]]
83     then
84         IFS=', '      # space is set as delimiter
85         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
86         for gene in "${ADDR[@]}; do
87             annotations=$(cut -f 13,14,15,18,20,24 ../../paraescarpia
_annotation_Jun2021_TrinoPantherK0.xls | fgrep $gene)
88             echo $orthogroup_ID$'\t'"paraescarpia"$'\t'$annotations
>> orthogroups_annotations_originated_vestimentifera_paraescarpia
a.csv
89         done
90     else

```

```

91     echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'
    ""$'\t'"" >> orthogroups_annotations_originated_vestimentifera_
paraescarpia.csv
92     fi
93 done < orthogroups_gene_IDs_originated_vestimentifera_paraescarp
ia_OK.txt
94
95 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t'
t'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
mber" > orthogroups_annotations_originated_vestimentifera_Oalv_R
pac_Lluy_Pech.csv
96 cat orthogroups_annotations_originated_vestimentifera_oasisia.cs
v orthogroups_annotations_originated_vestimentifera_riftia.csv o
rthogroups_annotations_originated_vestimentifera_lamellibrachia.
csv orthogroups_annotations_originated_vestimentifera_paraescarp
ia.csv >> orthogroups_annotations_originated_vestimentifera_Oalv_R
pac_Lluy_Pech.csv
97 sort orthogroups_annotations_originated_vestimentifera_Oalv_Rpac
_Lluy_Pech.csv > orthogroups_annotations_originated_vestimentife
ra_Oalv_Rpac_Lluy_Pech_OK.csv
98
99 echo "Orthogroup"$'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_originated_vestimentifera.csv
100 while read line; do
101     orthogroup_ID=$(cut -f 1 <<< "$line")
102     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_ori
ginated_vestimentifera_Ofra_Oalv_Rpac_Lluy_Pech_OK.csv | cut -f
7 | sed '/^$/d' | sort | uniq -c | sort -r | awk '{ $1="" ; print
$0 }' | head -1)
103     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA
bundantAnnotation_originated_vestimentifera.csv
104 done < gene_families_originated_vestimentifera.csv

```

orthogroups\_annotations\_originated\_vestimentifera\_cl1.sh

```
1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
   ul2021/gains/vestmentifera_cl1
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
   l2021/gains/vestmentifera_cl1
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=140:00:0
8  #$ -l highmem
9
10 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
   grep -w Vestimentifera_cl1 | cut -f 1 > gene_families_originated
   _vestmentifera_cl1.txt #families originated in oasisia
11 fgrep -f gene_families_originated_vestimentifera_cl1.txt ../../O
   rthogroups.csv > gene_families_originated_vestimentifera_cl1.csv
12
13 cut -f 1,19 gene_families_originated_vestimentifera_cl1.csv > or
   thogroups_gene_IDs_originated_vestimentifera_cl1_oasisia.txt
14 sed 's/0a1v|//g' orthogroups_gene_IDs_originated_vestimentifera_
   cl1_oasisia.txt > orthogroups_gene_IDs_originated_vestimentifera
   _cl1_oasisia_OK.txt
15 cut -f 1,24 gene_families_originated_vestimentifera_cl1.csv > or
   thogroups_gene_IDs_originated_vestimentifera_cl1_riftia.txt
16 sed 's/Rpac|//g' orthogroups_gene_IDs_originated_vestimentifera_
   cl1_riftia.txt > orthogroups_gene_IDs_originated_vestimentifera_
   cl1_riftia_OK.txt
17 cut -f 1,23 gene_families_originated_vestimentifera_cl1.csv > or
   thogroups_gene_IDs_originated_vestimentifera_cl1_paraescarpia.tx
   t
18 sed 's/Pech|//g' orthogroups_gene_IDs_originated_vestimentifera_
   cl1_paraescarpia.txt > orthogroups_gene_IDs_originated_vestiment
   ifera_cl1_paraescarpia_OK.txt
```

```

19
20
21 while read line; do
22     genes=$(cut -f 2 <<< "$line")
23     echo $genes
24     orthogroup_ID=$(cut -f 1 <<< "$line")
25     echo $orthogroup_ID
26     if [[ "$genes" == OALV* ]]
27     then
28         IFS=', '      # space is set as delimiter
29         read -ra ADDR <<< "$genes"  # str is read into an array a
s tokens separated by IFS
30         for gene in "${ADDR[@]}; do
31             annotations=$(cut -f 13,14,15,18,20,24 ../../oasisia_anno
tation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
32             echo $orthogroup_ID$'\t'"oasisia"$'\t'$annotations >> or
thogroups_annotations_originated_vestimentifera_cl1_oasisia.csv
33         done
34     else
35         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_originated_vestimentifera_
cl1_oasisia.csv
36     fi
37 done < orthogroups_gene_IDs_originated_vestimentifera_cl1_oasisi
a_OK.txt
38
39 while read line; do
40     genes=$(cut -f 2 <<< "$line")
41     echo $genes
42     orthogroup_ID=$(cut -f 1 <<< "$line")
43     echo $orthogroup_ID

```



```

44     if [[ "$genes" == RPAC* ]]
45     then
46         IFS=', '      # space is set as delimiter
47         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
48         for gene in "${ADDR[@]}"; do
49             annotations=$(cut -f 13,14,15,18,20,24 ../../riftia_annot
ation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
50             echo $orthogroup_ID$'\t'"riftia"$'\t'$annotations >> ort
hogroups_annotations_originated_vestimentifera_cl1_riftia.csv
51         done
52     else
53         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_originated_vestimentifera_
cl1_riftia.csv
54     fi
55 done < orthogroups_gene_IDs_originated_vestimentifera_cl1_riftia
_OK.txt
56
57 while read line; do
58     genes=$(cut -f 2 <<< "$line")
59     echo $genes
60     orthogroup_ID=$(cut -f 1 <<< "$line")
61     echo $orthogroup_ID
62     if [[ "$genes" == nbis* ]]
63     then
64         IFS=', '      # space is set as delimiter
65         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
66         for gene in "${ADDR[@]}"; do
67             annotations=$(cut -f 13,14,15,18,20,24 ../../paraescarpia
_annotation_Jun2021_TrinoPantherK0.xls | fgrep $gene)

```

```

68     echo $orthogroup_ID$'\t'"paraescarpia"$'\t'$annotations
>> orthogroups_annotations_originated_vestimentifera_cl1_paraesc
arpia.csv
69     done
70     else
71     echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_originated_vestimentifera_
cl1_paraescarpia.csv
72     fi
73 done < orthogroups_gene_IDs_originated_vestimentifera_cl1_paraes
carpia_OK.txt
74
75 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t
't'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
mber" > orthogroups_annotations_originated_vestimentifera_cl1_Oa
lv_Rpac_Pech.csv
76 cat orthogroups_annotations_originated_vestimentifera_cl1_oasisi
a.csv orthogroups_annotations_originated_vestimentifera_cl1_rift
ia.csv orthogroups_annotations_originated_vestimentifera_cl1_par
aescarpia.csv >> orthogroups_annotations_originated_vestimentife
ra_cl1_Oalv_Rpac_Pech.csv
77 sort orthogroups_annotations_originated_vestimentifera_cl1_Oalv_
Rpac_Pech.csv > orthogroups_annotations_originated_vestimentifer
a_cl1_Oalv_Rpac_Pech_OK.csv
78
79 echo "Orthogroup"$'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_originated_vestimentifera_cl1.csv
80 while read line; do
81     orthogroup_ID=$(cut -f 1 <<< "$line")
82     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_ori
ginated_vestimentifera_cl1_Oalv_Rpac_Pech_OK.csv | cut
-f 7 | sed '/^$/d' | sort | uniq -c | sort -r | awk '{ $1="" ; pr
int $0 }' | head -1)
83     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA

```

```
bundantAnnotation_originated_vestimentifera_cl1.csv
```

```
84 done < gene_families_originated_vestimentifera_cl1.csv
```

#### orthogroups\_annotations\_originated\_vestimentifera\_cl2.sh

```
1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
   ul2021/gains/vestimentifera_cl2
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
   l2021/gains/vestimentifera_cl2
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=140:00:0
8  #$ -l highmem
9
10 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
   grep -w Vestimentifera_cl2 | cut -f 1 > gene_families_originated
   _vestimentifera_cl2.txt #families originated in oasisia
11 fgrep -f gene_families_originated_vestimentifera_cl2.txt ../../O
   rthogroups.csv > gene_families_originated_vestimentifera_cl2.csv
12
13 cut -f 1,19 gene_families_originated_vestimentifera_cl2.csv > or
   thogroups_gene_IDs_originated_vestimentifera_cl2_oasisia.txt
14 sed 's/0a1v|//g' orthogroups_gene_IDs_originated_vestimentifera_
   cl2_oasisia.txt > orthogroups_gene_IDs_originated_vestimentifera
   _cl2_oasisia_OK.txt
15 cut -f 1,24 gene_families_originated_vestimentifera_cl2.csv > or
   thogroups_gene_IDs_originated_vestimentifera_cl2_riftia.txt
16 sed 's/Rpac|//g' orthogroups_gene_IDs_originated_vestimentifera_
   cl2_riftia.txt > orthogroups_gene_IDs_originated_vestimentifera_
   cl2_riftia_OK.txt
17
18
```

```

19
20 while read line; do
21     genes=$(cut -f 2 <<< "$line")
22     echo $genes
23     orthogroup_ID=$(cut -f 1 <<< "$line")
24     echo $orthogroup_ID
25     if [[ "$genes" == OALV* ]]
26     then
27         IFS=', '      # space is set as delimiter
28         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
29         for gene in "${ADDR[@]}"; do
30             annotations=$(cut -f 13,14,15,18,20,24 ../../oasisia_anno
tation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
31             echo $orthogroup_ID$'\t'"oasisia"$'\t'$annotations >> or
thogroups_annotations_originated_vestimentifera_cl2_oasisia.csv
32         done
33     else
34         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_originated_vestimentifera_
cl2_oasisia.csv
35     fi
36 done < orthogroups_gene_IDs_originated_vestimentifera_cl2_oasisi
a_OK.txt
37
38 while read line; do
39     genes=$(cut -f 2 <<< "$line")
40     echo $genes
41     orthogroup_ID=$(cut -f 1 <<< "$line")
42     echo $orthogroup_ID
43     if [[ "$genes" == RPAC* ]]

```

```

44     then
45         IFS=', '      # space is set as delimiter
46         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
47         for gene in "${ADDR[@]}; do
48             annotations=$(cut -f 13,14,15,18,20,24 ../../riftia_annot
ation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
49             echo $orthogroup_ID$'\t'"riftia"$'\t'$annotations >> ort
hogroups_annotations_originated_vestimentifera_cl2_riftia.csv
50         done
51     else
52         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_originated_vestimentifera_
cl2_riftia.csv
53     fi
54 done < orthogroups_gene_IDs_originated_vestimentifera_cl2_riftia
_OK.txt
55
56 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t
't'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
mber" > orthogroups_annotations_originated_vestimentifera_cl2_Oa
lv_Rpac.csv
57 cat orthogroups_annotations_originated_vestimentifera_cl2_oasisi
a.csv orthogroups_annotations_originated_vestimentifera_cl2_rift
ia.csv >> orthogroups_annotations_originated_vestimentifera_cl2_
Oalv_Rpac.csv
58 sort orthogroups_annotations_originated_vestimentifera_cl2_Oalv_
Rpac.csv > orthogroups_annotations_originated_vestimentifera_cl2
_Oalv_Rpac_OK.csv
59
60 echo "Orthogroup"$'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_originated_vestimentifera_cl2.csv
61 while read line; do

```

```

62     orthogroup_ID=$(cut -f 1 <<< "$line")
63     annotation=$(fgrep $orthogroup_ID orthogroups_annotated_
        originated_vestimentifera_cl2_0alv_Rpac_OK.csv | cut -f 7 | sed '/
        ^$/d' | sort | uniq -c | sort -r | awk '{ $1=""; print $0}' | he
        ad -1)
64     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA
        bundantAnnotation_originated_vestimentifera_cl2.csv
65 done < gene_families_originated_vestimentifera_cl2.csv

```

## Losses

orthogroups\_annotated\_losses\_osedax.sh

```

1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
        ul2021/losses/osedax
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
        l2021/losses/osedax
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=140:00:0
8  #$ -l highmem
9
10 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
        grep -w Ofra | cut -f 1 > gene_families_losses_osedax.txt #famil
        ies losses in oasisia
11 fgrep -f gene_families_losses_osedax.txt ../../Orthogroups.csv >
        gene_families_losses_osedax.csv
12
13 cut -f 1,19 gene_families_losses_osedax.csv > orthogroups_gene_I
        Ds_losses_osedax_oasisia.txt
14 sed 's/0alv|//g' orthogroups_gene_IDs_losses_osedax_oasisia.txt

```

```

> orthogroups_gene_IDs_losses_osedax_oasisia_OK.txt
15 cut -f 1,24 gene_families_losses_osedax.csv > orthogroups_gene_I
   Ds_losses_osedax_riftia.txt
16 sed 's/Rpac|//g' orthogroups_gene_IDs_losses_osedax_riftia.txt >
   orthogroups_gene_IDs_losses_osedax_riftia_OK.txt
17 cut -f 1,14 gene_families_losses_osedax.csv > orthogroups_gene_I
   Ds_losses_osedax_lamellibrachia.txt
18 sed 's/Lluy|//g' orthogroups_gene_IDs_losses_osedax_lamellibrach
   ia.txt > orthogroups_gene_IDs_losses_osedax_lamellibrachia_OK.tx
   t
19 cut -f 1,23 gene_families_losses_osedax.csv > orthogroups_gene_I
   Ds_losses_osedax_paraescarpia.txt
20 sed 's/Pech|//g' orthogroups_gene_IDs_losses_osedax_paraescarpi
   a.txt > orthogroups_gene_IDs_losses_osedax_paraescarpia_OK.txt
21 cut -f 1,21 gene_families_losses_osedax.csv > orthogroups_gene_I
   Ds_losses_osedax_owenia.txt
22 sed 's/Ofus|//g' orthogroups_gene_IDs_losses_osedax_owenia.txt >
   orthogroups_gene_IDs_losses_osedax_owenia_OK.txt
23 cut -f 1,5 gene_families_losses_osedax.csv > orthogroups_gene_ID
   s_losses_osedax_capitella.txt
24 sed 's/Ctel|//g' orthogroups_gene_IDs_losses_osedax_capitella.tx
   t > orthogroups_gene_IDs_losses_osedax_capitella_OK.txt
25
26
27 while read line; do
28     genes=$(cut -f 2 <<< "$line")
29     echo $genes
30     orthogroup_ID=$(cut -f 1 <<< "$line")
31     echo $orthogroup_ID
32     if [[ "$genes" == OFUS* ]]
33     then
34         IFS=', '      # space is set as delimiter
35         read -ra ADDR <<< "$genes"    # str is read into an array a

```

s tokens separated by IFS

```
36     for gene in "${ADDR[@]"}"; do
37         cut -f 1,2,3,11,12,13 ../../Owenia_annotation_v250920.1_
TrinoPantherK0.xls | fgrep $gene > temp_file.txt
38         #K0_number=$(cut -f 1 temp_file.txt)
39         #gene_ID=$(cut -f 2 temp_file.txt)
40         #Panther_annotation=$(cut -f 3 temp_file.txt)
41         #GO_1=$(cut -f 4 temp_file.txt)
42         #GO_1=$(cut -f 5 temp_file.txt)
43         #GO_1=$(cut -f 6 temp_file.txt)
44         echo $orthogroup_ID$'\t'"owenia"$'\t'$(cut -f 4 temp_fil
e.txt)$'\t'$(cut -f 5 temp_file.txt)$'\t'$(cut -f 6 temp_file.tx
t)$'\t'$(cut -f 2 temp_file.txt)$'\t'$(cut -f 3 temp_file.tx
t)$'\t'$(cut -f 1 temp_file.txt) >> orthogroups_annotations_loss
es_osedax_owenia.csv
45     done
46     else
47         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_osedax_owenia.csv
48     fi
49 done < orthogroups_gene_IDs_losses_osedax_owenia_OK.txt
50
51 while read line; do
52     genes=$(cut -f 2 <<< "$line")
53     echo $genes
54     orthogroup_ID=$(cut -f 1 <<< "$line")
55     echo $orthogroup_ID
56     if [[ "$genes" == CapteT* ]]
57     then
58         IFS=', '      # space is set as delimiter
59         read -ra ADDR <<< "$genes"  # str is read into an array a
s tokens separated by IFS
```



```

60     for gene in "${ADDR[@]}"; do
61         cut -f 1,3,7,21,22 ../../Capitella_annotation_Feb2021_Tri
noPantherK0.xls | fgrep $gene > temp_file.txt
62         #K0_number=$(cut -f 7 temp_file.txt)
63         #gene_ID=$(cut -f 1 temp_file.txt)
64         #Panther_annotation=$(cut -f 3 temp_file.txt)
65         #GO_1=$(cut -f 21 temp_file.txt)
66         #GO_1=$(cut -f 22 temp_file.txt)
67         #GO_1=$(cut -f 6 temp_file.txt) NONE
68         echo $orthogroup_ID$'\t'"capitella"$'\t'$(cut -f 21 temp
_file.txt)$'\t'$(cut -f 22 temp_file.txt)$'\t'""$'\t'$(cut -f 1 t
emp_file.txt)$'\t'$(cut -f 3 temp_file.txt)$'\t'$(cut -f 7 temp_
file.txt) >> orthogroups_annotations_losses_osedax_capitella.csv
69     done
70     else
71         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_osedax_capitella.cs
v
72     fi
73 done < orthogroups_gene_IDs_losses_osedax_capitella_OK.txt
74
75 while read line; do
76     genes=$(cut -f 2 <<< "$line")
77     echo $genes
78     orthogroup_ID=$(cut -f 1 <<< "$line")
79     echo $orthogroup_ID
80     if [[ "$genes" == OALV* ]]
81     then
82         IFS=', '      # space is set as delimiter
83         read -ra ADDR <<< "$genes"  # str is read into an array a
s tokens separated by IFS
84         for gene in "${ADDR[@]}"; do

```

```

85     annotations=$(cut -f 13,14,15,18,20,24 ../../oasisia_anno
tation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
86     echo $orthogroup_ID$'\t'"oasisia"$'\t'$annotations >> or
thogroups_annotations_losses_osedax_oasisia.csv
87     done
88     else
89     echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_osedax_oasisia.csv
90     fi
91 done < orthogroups_gene_IDs_losses_osedax_oasisia_OK.txt
92
93 while read line; do
94     genes=$(cut -f 2 <<< "$line")
95     echo $genes
96     orthogroup_ID=$(cut -f 1 <<< "$line")
97     echo $orthogroup_ID
98     if [[ "$genes" == RPAC* ]]
99     then
100         IFS=', '      # space is set as delimiter
101         read -ra ADDR <<< "$genes"      # str is read into an array a
s tokens separated by IFS
102         for gene in "${ADDR[@]}"; do
103             annotations=$(cut -f 13,14,15,18,20,24 ../../riftia_annot
ation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
104             echo $orthogroup_ID$'\t'"riftia"$'\t'$annotations >> ort
hogroups_annotations_losses_osedax_riftia.csv
105             done
106             else
107             echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_osedax_riftia.csv
108             fi
109 done < orthogroups_gene_IDs_losses_osedax_riftia_OK.txt

```

```

110
111 while read line; do
112     genes=$(cut -f 2 <<< "$line")
113     echo $genes
114     orthogroup_ID=$(cut -f 1 <<< "$line")
115     echo $orthogroup_ID
116     if [[ "$genes" == FUN* ]]
117     then
118         IFS=', '      # space is set as delimiter
119         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
120         for gene in "${ADDR[@]}; do
121             annotations=$(cut -f 13,14,15,18,20,24 ../../lamellibrach
ia_annotation_Feb2021_TrinoPantherK0_OK.xls | fgrep $gene)
122             echo $orthogroup_ID$'\t'"lamellibrachia"$'\t'$annotation
s >> orthogroups_annotations_losses_osedax_lamellibrachia.csv
123         done
124     else
125         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_osedax_lamellibrach
ia.csv
126     fi
127 done < orthogroups_gene_IDs_losses_osedax_lamellibrachia_OK.txt
128
129 while read line; do
130     genes=$(cut -f 2 <<< "$line")
131     echo $genes
132     orthogroup_ID=$(cut -f 1 <<< "$line")
133     echo $orthogroup_ID
134     if [[ "$genes" == nbis* ]]
135     then

```

```

136     IFS=', '      # space is set as delimiter
137     read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
138     for gene in "${ADDR[@]}"; do
139         annotations=$(cut -f 13,14,15,18,20,24 ../../paraescarpia
_annotation_Jun2021_TrinoPantherK0.xls | fgrep $gene)
140         echo $orthogroup_ID$'\t'"paraescarpia"$'\t'$annotations
>> orthogroups_annotations_losses_osedax_paraescarpia.csv
141     done
142     else
143         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_osedax_paraescarpi
a.csv
144     fi
145 done < orthogroups_gene_IDs_losses_osedax_paraescarpia_OK.txt
146
147 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t
't'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
mber" > orthogroups_annotations_losses_osedax_Oalv_Rpac_Lluy_Pec
h_Ofus_Ctel.csv
148 cat orthogroups_annotations_losses_osedax_oasisia.csv orthogroup
s_annotations_losses_osedax_riftia.csv orthogroups_annotations_l
osses_osedax_lamellibrachia.csv orthogroups_annotations_losses_o
sedax_paraescarpia.csv orthogroups_annotations_losses_osedax_owe
nia.csv orthogroups_annotations_losses_osedax_capitella.csv >> o
rthogroups_annotations_losses_osedax_Oalv_Rpac_Lluy_Pech_Ofus_Ct
el.csv
149 sort orthogroups_annotations_losses_osedax_Oalv_Rpac_Lluy_Pech_O
fus_Ctel.csv > orthogroups_annotations_losses_osedax_Oalv_Rpac_L
luy_Pech_Ofus_Ctel_OK.csv
150
151 echo "Orthogroup"$'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_losses_osedax.csv

```

```

152 while read line; do
153     orthogroup_ID=$(cut -f 1 <<< "$line")
154     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_los
ses_osedax_0alv_Rpac_Lluy_Pech_0fus_Ctel_0K.csv | cut -f 7 | sed
'/^$/d' | sort | uniq -c | sort -r | awk '{$1=""; print $0}' |
head -1)
155     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA
bundantAnnotation_losses_osedax.csv
156 done < gene_families_losses_osedax.csv

```

#### orthogroups\_annotations\_losses\_oasisia.sh

```

1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
ul2021/losses/oasisia
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
l2021/losses/oasisia
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=140:00:0
8  #$ -l highmem
9
10 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
grep -w 0alv | cut -f 1 > gene_families_losses_oasisia.txt #fami
lies losses in oasisia
11 fgrep -f gene_families_losses_oasisia.txt ../../Orthogroups.csv
> gene_families_losses_oasisia.csv
12
13 cut -f 1,20 gene_families_losses_oasisia.csv > orthogroups_gene_
IDs_losses_oasisia_osedax.txt
14 sed 's/0fra|//g' orthogroups_gene_IDs_losses_oasisia_osedax.txt
> orthogroups_gene_IDs_losses_oasisia_osedax_0K.txt
15 cut -f 1,24 gene_families_losses_oasisia.csv > orthogroups_gene_

```

```

IDs_losses_oasisia_riftia.txt
16 sed 's/Rpac|//g' orthogroups_gene_IDs_losses_oasisia_riftia.txt
   > orthogroups_gene_IDs_losses_oasisia_riftia_OK.txt
17 cut -f 1,14 gene_families_losses_oasisia.csv > orthogroups_gene_
   IDs_losses_oasisia_lamellibrachia.txt
18 sed 's/Lluy|//g' orthogroups_gene_IDs_losses_oasisia_lamellibrac
   hia.txt > orthogroups_gene_IDs_losses_oasisia_lamellibrachia_OK.
   txt
19 cut -f 1,23 gene_families_losses_oasisia.csv > orthogroups_gene_
   IDs_losses_oasisia_paraescarpia.txt
20 sed 's/Pech|//g' orthogroups_gene_IDs_losses_oasisia_paraescarp
   a.txt > orthogroups_gene_IDs_losses_oasisia_paraescarpia_OK.txt
21 cut -f 1,21 gene_families_losses_oasisia.csv > orthogroups_gene_
   IDs_losses_oasisia_owenia.txt
22 sed 's/Ofus|//g' orthogroups_gene_IDs_losses_oasisia_owenia.txt
   > orthogroups_gene_IDs_losses_oasisia_owenia_OK.txt
23 cut -f 1,5 gene_families_losses_oasisia.csv > orthogroups_gene_I
   Ds_losses_oasisia_capitella.txt
24 sed 's/Ctel|//g' orthogroups_gene_IDs_losses_oasisia_capitella.t
   xt > orthogroups_gene_IDs_losses_oasisia_capitella_OK.txt
25
26
27 while read line; do
28     genes=$(cut -f 2 <<< "$line")
29     echo $genes
30     orthogroup_ID=$(cut -f 1 <<< "$line")
31     echo $orthogroup_ID
32     if [[ "$genes" == OFUS* ]]
33     then
34         IFS=', '      # space is set as delimiter
35         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS

```

```

36     for gene in "${ADDR[@]"}; do
37         cut -f 1,2,3,11,12,13 ../../Owenia_annotation_v250920.1_
TrinoPantherK0.xls | fgrep $gene > temp_file.txt
38         #KO_number=$(cut -f 1 temp_file.txt)
39         #gene_ID=$(cut -f 2 temp_file.txt)
40         #Panther_annotation=$(cut -f 3 temp_file.txt)
41         #GO_1=$(cut -f 4 temp_file.txt)
42         #GO_1=$(cut -f 5 temp_file.txt)
43         #GO_1=$(cut -f 6 temp_file.txt)
44         echo $orthogroup_ID$'\t'"owenia"$'\t'$(cut -f 4 temp_fil
e.txt)$'\t'$(cut -f 5 temp_file.txt)$'\t'$(cut -f 6 temp_file.tx
t)$'\t'$(cut -f 2 temp_file.txt)$'\t'$(cut -f 3 temp_file.tx
t)$'\t'$(cut -f 1 temp_file.txt) >> orthogroups_annotations_loss
es_oasisia_owenia.csv
45     done
46     else
47         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_oasisia_owenia.csv
48     fi
49 done < orthogroups_gene_IDs_losses_oasisia_owenia_OK.txt
50
51 while read line; do
52     genes=$(cut -f 2 <<< "$line")
53     echo $genes
54     orthogroup_ID=$(cut -f 1 <<< "$line")
55     echo $orthogroup_ID
56     if [[ "$genes" == CapteT* ]]
57     then
58         IFS=', '      # space is set as delimiter
59         read -ra ADDR <<< "$genes"  # str is read into an array a
s tokens separated by IFS
60         for gene in "${ADDR[@]"}; do

```

```

61         cut -f 1,3,7,21,22 ../../Capitella_annotation_Feb2021_Tri
noPantherK0.xls | fgrep $gene > temp_file.txt
62         #K0_number=$(cut -f 7 temp_file.txt)
63         #gene_ID=$(cut -f 1 temp_file.txt)
64         #Panther_annotation=$(cut -f 3 temp_file.txt)
65         #GO_1=$(cut -f 21 temp_file.txt)
66         #GO_1=$(cut -f 22 temp_file.txt)
67         #GO_1=$(cut -f 6 temp_file.txt) NONE
68         echo $orthogroup_ID$'\t'"capitella"$'\t'$(cut -f 21 temp
_file.txt)$'\t'$(cut -f 22 temp_file.txt)$'\t'""'\t'$(cut -f 1 t
emp_file.txt)$'\t'$(cut -f 3 temp_file.txt)$'\t'$(cut -f 7 temp_
file.txt) >> orthogroups_annotations_losses_oasisia_capitella.cs
v
69         done
70     else
71         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_oasisia_capitella.c
sv
72     fi
73 done < orthogroups_gene_IDs_losses_oasisia_capitella_OK.txt
74
75 while read line; do
76     genes=$(cut -f 2 <<< "$line")
77     echo $genes
78     orthogroup_ID=$(cut -f 1 <<< "$line")
79     echo $orthogroup_ID
80     if [[ "$genes" == OFRA* ]]
81     then
82         IFS=', '      # space is set as delimiter
83         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
84         for gene in "${ADDR[@]}; do

```



```

85     annotations=$(cut -f 13,14,15,18,20,24 ../../osedax_annotation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
86     echo $orthogroup_ID$'\t'"osedax"$'\t'$annotations >> orthogroups_annotations_losses_oasisia_osedax.csv
87     done
88     else
89     echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'"" >> orthogroups_annotations_losses_oasisia_osedax.csv
90     fi
91 done < orthogroups_gene_IDs_losses_oasisia_osedax_OK.txt
92
93 while read line; do
94     genes=$(cut -f 2 <<< "$line")
95     echo $genes
96     orthogroup_ID=$(cut -f 1 <<< "$line")
97     echo $orthogroup_ID
98     if [[ "$genes" == RPAC* ]]
99     then
100         IFS=', '      # space is set as delimiter
101         read -ra ADDR <<< "$genes"      # str is read into an array as tokens separated by IFS
102         for gene in "${ADDR[@]}; do
103             annotations=$(cut -f 13,14,15,18,20,24 ../../riftia_annotation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
104             echo $orthogroup_ID$'\t'"riftia"$'\t'$annotations >> orthogroups_annotations_losses_oasisia_riftia.csv
105         done
106     else
107         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'"" >> orthogroups_annotations_losses_oasisia_riftia.csv
108     fi
109 done < orthogroups_gene_IDs_losses_oasisia_riftia_OK.txt

```

```

110
111 while read line; do
112     genes=$(cut -f 2 <<< "$line")
113     echo $genes
114     orthogroup_ID=$(cut -f 1 <<< "$line")
115     echo $orthogroup_ID
116     if [[ "$genes" == FUN* ]]
117     then
118         IFS=', '      # space is set as delimiter
119         read -ra ADDR <<< "$genes"  # str is read into an array a
s tokens separated by IFS
120         for gene in "${ADDR[@]}; do
121             annotations=$(cut -f 13,14,15,18,20,24 ../../lamellibrach
ia_annotation_Feb2021_TrinoPantherK0_OK.xls | fgrep $gene)
122             echo $orthogroup_ID$'\t'"lamellibrachia"$'\t'$annotation
s >> orthogroups_annotations_losses_oasisia_lamellibrachia.csv
123         done
124     else
125         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_oasisia_lamellibrac
hia.csv
126     fi
127 done < orthogroups_gene_IDs_losses_oasisia_lamellibrachia_OK.txt
128
129 while read line; do
130     genes=$(cut -f 2 <<< "$line")
131     echo $genes
132     orthogroup_ID=$(cut -f 1 <<< "$line")
133     echo $orthogroup_ID
134     if [[ "$genes" == nbis* ]]
135     then

```

```

136     IFS=', '      # space is set as delimiter
137     read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
138     for gene in "${ADDR[@]}; do
139         annotations=$(cut -f 13,14,15,18,20,24 ../../paraescarpia
_annotation_Jun2021_TrinoPantherK0.xls | fgrep $gene)
140         echo $orthogroup_ID$'\t'"paraescarpia"$'\t'$annotations
>> orthogroups_annotations_losses_oasisia_paraescarpia.csv
141     done
142     else
143         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'
""$'\t'"" >> orthogroups_annotations_losses_oasisia_paraescarpia
a.csv
144     fi
145 done < orthogroups_gene_IDs_losses_oasisia_paraescarpia_0K.txt
146
147 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t'
t'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
mber" > orthogroups_annotations_losses_oasisia_Ofra_Rpac_Lluy_Pe
ch_Ofus_Ctel.csv
148 cat orthogroups_annotations_losses_oasisia_osedax.csv orthogroup
s_annotations_losses_oasisia_riftia.csv orthogroups_annotations_
losses_oasisia_lamellibrachia.csv orthogroups_annotations_losses
_oasisia_paraescarpia.csv orthogroups_annotations_losses_oasisia
_owenia.csv orthogroups_annotations_losses_oasisia_capitella.csv
>> orthogroups_annotations_losses_oasisia_Ofra_Rpac_Lluy_Pech_Of
us_Ctel.csv
149 sort orthogroups_annotations_losses_oasisia_Ofra_Rpac_Lluy_Pech_
Ofus_Ctel.csv > orthogroups_annotations_losses_oasisia_Ofra_Rpac
_Lluy_Pech_Ofus_Ctel_0K.csv
150
151 echo "Orthogroup"$'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_losses_oasisia.csv

```

```

152 while read line; do
153     orthogroup_ID=$(cut -f 1 <<< "$line")
154     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_los
ses_oasisia_Ofra_Rpac_Lluy_Pech_Ofus_Ctel_OK.csv | cut -f 7 | se
d '/^$/d' | sort | uniq -c | sort -r | awk '{s1=""; print $0}'
| head -1)
155     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA
bundantAnnotation_losses_oasisia.csv
156 done < gene_families_losses_oasisia.csv

```

#### orthogroups\_annotations\_losses\_riftia.sh

```

1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
ul2021/losses/riftia
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
l2021/losses/riftia
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=140:00:0
8  #$ -l highmem
9
10 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
grep -w Rpac | cut -f 1 > gene_families_losses_riftia.txt #famil
ies losses in oasisia
11 fgrep -f gene_families_losses_riftia.txt ../../Orthogroups.csv >
gene_families_losses_riftia.csv
12
13 cut -f 1,20 gene_families_losses_riftia.csv > orthogroups_gene_I
Ds_losses_riftia_osedax.txt
14 sed 's/Ofra|//g' orthogroups_gene_IDs_losses_riftia_osedax.txt >
orthogroups_gene_IDs_losses_riftia_osedax_OK.txt
15 cut -f 1,14 gene_families_losses_riftia.csv > orthogroups_gene_I

```

```

Ds_losses_riftia_lamellibrachia.txt
16 sed 's/Lluy|//g' orthogroups_gene_IDs_losses_riftia_lamellibrach
ia.txt > orthogroups_gene_IDs_losses_riftia_lamellibrachia_OK.tx
t
17 cut -f 1,23 gene_families_losses_riftia.csv > orthogroups_gene_I
Ds_losses_riftia_paraescarpia.txt
18 sed 's/Pech|//g' orthogroups_gene_IDs_losses_riftia_paraescarpi
a.txt > orthogroups_gene_IDs_losses_riftia_paraescarpia_OK.txt
19 cut -f 1,21 gene_families_losses_riftia.csv > orthogroups_gene_I
Ds_losses_riftia_owenia.txt
20 sed 's/Ofus|//g' orthogroups_gene_IDs_losses_riftia_owenia.txt >
orthogroups_gene_IDs_losses_riftia_owenia_OK.txt
21 cut -f 1,5 gene_families_losses_riftia.csv > orthogroups_gene_ID
s_losses_riftia_capitella.txt
22 sed 's/Ctel|//g' orthogroups_gene_IDs_losses_riftia_capitella.tx
t > orthogroups_gene_IDs_losses_riftia_capitella_OK.txt
23 cut -f 1,19 gene_families_losses_riftia.csv > orthogroups_gene_I
Ds_losses_riftia_oasisia.txt
24 sed 's/Oalv|//g' orthogroups_gene_IDs_losses_riftia_oasisia.txt
> orthogroups_gene_IDs_losses_riftia_oasisia_OK.txt
25
26 while read line; do
27     genes=$(cut -f 2 <<< "$line")
28     echo $genes
29     orthogroup_ID=$(cut -f 1 <<< "$line")
30     echo $orthogroup_ID
31     if [[ "$genes" == OALV* ]]
32     then
33         IFS=', '      # space is set as delimiter
34         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
35         for gene in "${ADDR[@]}"; do
36             annotations=$(cut -f 13,14,15,18,20,24 ../../oasisia_anno

```

```

tation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
37     echo $orthogroup_ID$'\t'"oasisia"$'\t'$annotations >> or
thogroups_annotations_losses_riftia_oasisia.csv
38     done
39     else
40     echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_riftia_oasisia.csv
41     fi
42 done < orthogroups_gene_IDs_losses_riftia_oasisia_OK.txt
43
44 while read line; do
45     genes=$(cut -f 2 <<< "$line")
46     echo $genes
47     orthogroup_ID=$(cut -f 1 <<< "$line")
48     echo $orthogroup_ID
49     if [[ "$genes" == OFUS* ]]
50     then
51         IFS=', '      # space is set as delimiter
52         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
53         for gene in "${ADDR[@]}; do
54             cut -f 1,2,3,11,12,13 ../../Owenia_annotation_v250920.1_
TrinoPantherK0.xls | fgrep $gene > temp_file.txt
55             #K0_number=$(cut -f 1 temp_file.txt)
56             #gene_ID=$(cut -f 2 temp_file.txt)
57             #Panther_annotation=$(cut -f 3 temp_file.txt)
58             #GO_1=$(cut -f 4 temp_file.txt)
59             #GO_1=$(cut -f 5 temp_file.txt)
60             #GO_1=$(cut -f 6 temp_file.txt)
61             echo $orthogroup_ID$'\t'"owenia"$'\t'$(cut -f 4 temp_fil
e.txt)$'\t'$(cut -f 5 temp_file.txt)$'\t'$(cut -f 6 temp_file.tx
t)$'\t'$(cut -f 2 temp_file.txt)$'\t'$(cut -f 3 temp_file.tx

```

```
t)${'\t'}$(cut -f 1 temp_file.txt) >> orthogroups_annotations_losses_riftia_owenia.csv
```

```
62     done
```

```
63     else
```

```
64     echo $orthogroup_ID${'\t'}${'\t'}${'\t'}${'\t'}${'\t'}${'\t'}${'\t'}${'\t'}${'\t'} >> orthogroups_annotations_losses_riftia_owenia.csv
```

```
65     fi
```

```
66 done < orthogroups_gene_IDs_losses_riftia_owenia_OK.txt
```

```
67
```

```
68 while read line; do
```

```
69     genes=$(cut -f 2 <<< "$line")
```

```
70     echo $genes
```

```
71     orthogroup_ID=$(cut -f 1 <<< "$line")
```

```
72     echo $orthogroup_ID
```

```
73     if [[ "$genes" == CapteT* ]]
```

```
74     then
```

```
75         IFS=', '      # space is set as delimiter
```

```
76         read -ra ADDR <<< "$genes"  # str is read into an array as tokens separated by IFS
```

```
77         for gene in "${ADDR[@]}; do
```

```
78             cut -f 1,3,7,21,22 ../../Capitella_annotation_Feb2021_Tri  
noPantherK0.xls | fgrep $gene > temp_file.txt
```

```
79             #K0_number=$(cut -f 7 temp_file.txt)
```

```
80             #gene_ID=$(cut -f 1 temp_file.txt)
```

```
81             #Panther_annotation=$(cut -f 3 temp_file.txt)
```

```
82             #GO_1=$(cut -f 21 temp_file.txt)
```

```
83             #GO_1=$(cut -f 22 temp_file.txt)
```

```
84             #GO_1=$(cut -f 6 temp_file.txt) NONE
```

```
85             echo $orthogroup_ID${'\t'}"capitella"${'\t'}$(cut -f 21 temp  
_file.txt)${'\t'}$(cut -f 22 temp_file.txt)${'\t'}${'\t'}$(cut -f 1 t  
emp_file.txt)${'\t'}$(cut -f 3 temp_file.txt)${'\t'}$(cut -f 7 temp_  
file.txt) >> orthogroups_annotations_losses_riftia_capitella.csv
```

```

86         done
87     else
88         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'
'""$'\t'"" >> orthogroups_annotations_losses_riftia_capitella.cs
v
89     fi
90 done < orthogroups_gene_IDs_losses_riftia_capitella_OK.txt
91
92 while read line; do
93     genes=$(cut -f 2 <<< "$line")
94     echo $genes
95     orthogroup_ID=$(cut -f 1 <<< "$line")
96     echo $orthogroup_ID
97     if [[ "$genes" == OFRA* ]]
98     then
99         IFS=', '      # space is set as delimiter
100         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
101         for gene in "${ADDR[@]}"; do
102             annotations=$(cut -f 13,14,15,18,20,24 ../../osedax_annot
ation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
103             echo $orthogroup_ID$'\t'"osedax"$'\t'$annotations >> ort
hogroups_annotations_losses_riftia_osedax.csv
104         done
105     else
106         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_riftia_osedax.csv
107     fi
108 done < orthogroups_gene_IDs_losses_riftia_osedax_OK.txt
109
110 while read line; do

```



```

111 genes=$(cut -f 2 <<< "$line")
112 echo $genes
113 orthogroup_ID=$(cut -f 1 <<< "$line")
114 echo $orthogroup_ID
115 if [[ "$genes" == FUN* ]]
116 then
117     IFS=', '      # space is set as delimiter
118     read -ra ADDR <<< "$genes"    # str is read into an array a
119     for gene in "${ADDR[@]}"; do
120         annotations=$(cut -f 13,14,15,18,20,24 ../../lamellibrach
121         ia_annotation_Feb2021_TrinoPantherKO_OK.xls | fgrep $gene)
122         echo $orthogroup_ID$'\t'"lamellibrachia"$'\t'$annotations
123     s >> orthogroups_annotations_losses_riftia_lamellibrachia.csv
124     done
125 else
126     echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
127     '""$'\t'"" >> orthogroups_annotations_losses_riftia_lamellibrach
128     ia.csv
129 fi
130 done < orthogroups_gene_IDs_losses_riftia_lamellibrachia_OK.txt
131
132 while read line; do
133     genes=$(cut -f 2 <<< "$line")
134     echo $genes
135     orthogroup_ID=$(cut -f 1 <<< "$line")
136     echo $orthogroup_ID
137     if [[ "$genes" == nbis* ]]
138     then
139         IFS=', '      # space is set as delimiter
140         read -ra ADDR <<< "$genes"    # str is read into an array a

```

s tokens separated by IFS

```
137     for gene in "${ADDR[@]"}"; do
138         annotations=$(cut -f 13,14,15,18,20,24 ../../paraescarpia
_annotation_Jun2021_TrinoPantherK0.xls | fgrep $gene)
139         echo $orthogroup_ID$'\t'"paraescarpia"$'\t'$annotations
>> orthogroups_annotations_losses_riftia_paraescarpia.csv
140     done
141     else
142         echo $orthogroup_ID$'\t'"'"$'\t'"'"$'\t'"'"$'\t'"'"$'\t'"'"$'\t
'"'"$'\t'"'" >> orthogroups_annotations_losses_riftia_paraescarpi
a.csv
143     fi
144 done < orthogroups_gene_IDs_losses_riftia_paraescarpia_OK.txt
145
146 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t
t'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
mber" > orthogroups_annotations_losses_riftia_0fra_0alv_Lluy_Pec
h_0fus_Ctel.csv
147 cat orthogroups_annotations_losses_riftia_oasisia.csv orthogroup
s_annotations_losses_riftia_osedax.csv orthogroups_annotations_l
osses_riftia_lamellibrachia.csv orthogroups_annotations_losses_r
iftia_paraescarpia.csv orthogroups_annotations_losses_riftia_owe
nia.csv orthogroups_annotations_losses_riftia_capitella.csv >> o
rthogroups_annotations_losses_riftia_0fra_0alv_Lluy_Pech_0fus_Ct
el.csv
148 sort orthogroups_annotations_losses_riftia_0fra_0alv_Lluy_Pech_0
fus_Ctel.csv > orthogroups_annotations_losses_riftia_0fra_0alv_L
luy_Pech_0fus_Ctel_OK.csv
149
150 echo "Orthogroup"$'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_losses_riftia.csv
151 while read line; do
152     orthogroup_ID=$(cut -f 1 <<< "$line")
```

```

153     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_los
ses_riftia_0fra_0alv_Lluy_Pech_0fus_Ctel_OK.csv | cut -f 7 | sed
'/^$/d' | sort | uniq -c | sort -r | awk '{ $1=""; print $0}' |
head -1)
154     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA
bundantAnnotation_losses_riftia.csv
155 done < gene_families_losses_riftia.csv

```

#### orthogroups\_annotations\_losses\_lamellibrachia.sh

```

1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
ul2021/losses/lamellibrachia
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
l2021/losses/lamellibrachia
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=140:00:0
8  #$ -l highmem
9
10 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
grep -w Lluy | cut -f 1 > gene_families_losses_lamellibrachia.tx
t #families losses in oasisia
11 fgrep -f gene_families_losses_lamellibrachia.txt ../../Orthogrou
ps.csv > gene_families_losses_lamellibrachia.csv
12
13 cut -f 1,20 gene_families_losses_lamellibrachia.csv > orthogroup
s_gene_IDs_losses_lamellibrachia_osedax.txt
14 sed 's/0fra|//g' orthogroups_gene_IDs_losses_lamellibrachia_osed
ax.txt > orthogroups_gene_IDs_losses_lamellibrachia_osedax_OK.tx
t
15 cut -f 1,23 gene_families_losses_lamellibrachia.csv > orthogroup
s_gene_IDs_losses_lamellibrachia_paraescarpia.txt

```

```

16 sed 's/Pech|//g' orthogroups_gene_IDs_losses_lamellibrachia_para
    escarpia.txt > orthogroups_gene_IDs_losses_lamellibrachia_paraes
    carpia_OK.txt
17 cut -f 1,21 gene_families_losses_lamellibrachia.csv > orthogroup
    s_gene_IDs_losses_lamellibrachia_owenia.txt
18 sed 's/Ofus|//g' orthogroups_gene_IDs_losses_lamellibrachia_owen
    ia.txt > orthogroups_gene_IDs_losses_lamellibrachia_owenia_OK.tx
    t
19 cut -f 1,5 gene_families_losses_lamellibrachia.csv > orthogroups
    _gene_IDs_losses_lamellibrachia_capitella.txt
20 sed 's/Ctel|//g' orthogroups_gene_IDs_losses_lamellibrachia_capi
    tella.txt > orthogroups_gene_IDs_losses_lamellibrachia_capitella
    _OK.txt
21 cut -f 1,19 gene_families_losses_lamellibrachia.csv > orthogroup
    s_gene_IDs_losses_lamellibrachia_oasisia.txt
22 sed 's/Oalv|//g' orthogroups_gene_IDs_losses_lamellibrachia_oasi
    sia.txt > orthogroups_gene_IDs_losses_lamellibrachia_oasisia_OK.
    txt
23 cut -f 1,24 gene_families_losses_lamellibrachia.csv > orthogroup
    s_gene_IDs_losses_lamellibrachia_riftia.txt
24 sed 's/Rpac|//g' orthogroups_gene_IDs_losses_lamellibrachia_rift
    ia.txt > orthogroups_gene_IDs_losses_lamellibrachia_riftia_OK.tx
    t
25
26 while read line; do
27     genes=$(cut -f 2 <<< "$line")
28     echo $genes
29     orthogroup_ID=$(cut -f 1 <<< "$line")
30     echo $orthogroup_ID
31     if [[ "$genes" == RPAC* ]]
32     then
33         IFS=', '      # space is set as delimiter
34         read -ra ADDR <<< "$genes"    # str is read into an array a

```

s tokens separated by IFS

```
35     for gene in "${ADDR[@]"}"; do
36         annotations=$(cut -f 13,14,15,18,20,24 ../../riftia_annotation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
37         echo $orthogroup_ID$'\t'"riftia"$'\t'$annotations >> orthogroups_annotations_losses_lamellibrachia_riftia.csv
38     done
39 else
40     echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'"" >> orthogroups_annotations_losses_lamellibrachia_riftia.csv
41 fi
42 done < orthogroups_gene_IDs_losses_lamellibrachia_riftia_OK.txt
43
44 while read line; do
45     genes=$(cut -f 2 <<< "$line")
46     echo $genes
47     orthogroup_ID=$(cut -f 1 <<< "$line")
48     echo $orthogroup_ID
49     if [[ "$genes" == OALV* ]]
50     then
51         IFS=', '      # space is set as delimiter
52         read -ra ADDR <<< "$genes"    # str is read into an array as
s tokens separated by IFS
53         for gene in "${ADDR[@]"}"; do
54             annotations=$(cut -f 13,14,15,18,20,24 ../../oasisia_annotation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
55             echo $orthogroup_ID$'\t'"oasisia"$'\t'$annotations >> orthogroups_annotations_losses_lamellibrachia_oasisia.csv
56         done
57     else
58         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""
```

```

"""$'\t'""" >> orthogroups_annotations_losses_lamellibrachia_oasi
sia.csv
59     fi
60 done < orthogroups_gene_IDs_losses_lamellibrachia_oasisia_OK.txt
61
62 while read line; do
63     genes=$(cut -f 2 <<< "$line")
64     echo $genes
65     orthogroup_ID=$(cut -f 1 <<< "$line")
66     echo $orthogroup_ID
67     if [[ "$genes" == OFUS* ]]
68     then
69         IFS=', '      # space is set as delimiter
70         read -ra ADDR <<< "$genes"  # str is read into an array a
s tokens separated by IFS
71         for gene in "${ADDR[@]}; do
72             cut -f 1,2,3,11,12,13 ../../Owenia_annotation_v250920.1_
TrinoPantherK0.xls | fgrep $gene > temp_file.txt
73             #K0_number=$(cut -f 1 temp_file.txt)
74             #gene_ID=$(cut -f 2 temp_file.txt)
75             #Panther_annotation=$(cut -f 3 temp_file.txt)
76             #GO_1=$(cut -f 4 temp_file.txt)
77             #GO_1=$(cut -f 5 temp_file.txt)
78             #GO_1=$(cut -f 6 temp_file.txt)
79             echo $orthogroup_ID$'\t'"owenia"$'\t'$(cut -f 4 temp_fil
e.txt)$'\t'$(cut -f 5 temp_file.txt)$'\t'$(cut -f 6 temp_file.tx
t)$'\t'$(cut -f 2 temp_file.txt)$'\t'$(cut -f 3 temp_file.tx
t)$'\t'$(cut -f 1 temp_file.txt) >> orthogroups_annotations_loss
es_lamellibrachia_owenia.csv
80         done
81     else
82         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'

```

```
""$'\t'"" >> orthogroups_annotations_losses_lamellibrachia_owen  
ia.csv
```

```
fi
```

```
done < orthogroups_gene_IDs_losses_lamellibrachia_owenia_OK.txt
```

```
while read line; do
```

```
genes=$(cut -f 2 <<< "$line")
```

```
echo $genes
```

```
orthogroup_ID=$(cut -f 1 <<< "$line")
```

```
echo $orthogroup_ID
```

```
if [[ "$genes" == CapteT* ]]
```

```
then
```

```
IFS=', ' # space is set as delimiter
```

```
read -ra ADDR <<< "$genes" # str is read into an array a  
s tokens separated by IFS
```

```
for gene in "${ADDR[@]}; do
```

```
cut -f 1,3,7,21,22 ../../Capitella_annotation_Feb2021_Tri  
noPantherK0.xls | fgrep $gene > temp_file.txt
```

```
#K0_number=$(cut -f 7 temp_file.txt)
```

```
#gene_ID=$(cut -f 1 temp_file.txt)
```

```
#Panther_annotation=$(cut -f 3 temp_file.txt)
```

```
#GO_1=$(cut -f 21 temp_file.txt)
```

```
#GO_1=$(cut -f 22 temp_file.txt)
```

```
#GO_1=$(cut -f 6 temp_file.txt) NONE
```

```
echo $orthogroup_ID$'\t'"capitella"$'\t'$(cut -f 21 temp  
_file.txt)$'\t'$(cut -f 22 temp_file.txt)$'\t'""$'\t'$(cut -f 1 t  
emp_file.txt)$'\t'$(cut -f 3 temp_file.txt)$'\t'$(cut -f 7 temp_  
file.txt) >> orthogroups_annotations_losses_lamellibrachia_capit  
ella.csv
```

```
done
```

```
else
```

```
echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'
```

```

107         fi
108 done < orthogroups_gene_IDs_losses_lamellibrachia_capitella_0K.txt
109
110 while read line; do
111     genes=$(cut -f 2 <<< "$line")
112     echo $genes
113     orthogroup_ID=$(cut -f 1 <<< "$line")
114     echo $orthogroup_ID
115     if [[ "$genes" == OFRA* ]]
116     then
117         IFS=', '      # space is set as delimiter
118         read -ra ADDR <<< "$genes"    # str is read into an array as
tokens separated by IFS
119         for gene in "${ADDR[@]}; do
120             annotations=$(cut -f 13,14,15,18,20,24 ../../osedax_annotation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
121             echo $orthogroup_ID$'\t'"osedax"$'\t'$annotations >> orthogroups_annotations_losses_lamellibrachia_osedax.csv
122         done
123     else
124         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'"" >> orthogroups_annotations_losses_lamellibrachia_osedax.csv
125     fi
126 done < orthogroups_gene_IDs_losses_lamellibrachia_osedax_0K.txt
127
128 while read line; do
129     genes=$(cut -f 2 <<< "$line")
130     echo $genes

```



```

131 orthogroup_ID=$(cut -f 1 <<< "$line")
132 echo $orthogroup_ID
133 if [[ "$genes" == nbis* ]]
134 then
135     IFS=', '      # space is set as delimiter
136     read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
137     for gene in "${ADDR[@]}; do
138         annotations=$(cut -f 13,14,15,18,20,24 ../../paraescarpia
_annotation_Jun2021_TrinoPantherK0.xls | fgrep $gene)
139         echo $orthogroup_ID$'\t'"paraescarpia"$'\t'$annotations
>> orthogroups_annotations_losses_lamellibrachia_paraescarpia.cs
v
140     done
141 else
142     echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_lamellibrachia_para
escarpia.csv
143 fi
144 done < orthogroups_gene_IDs_losses_lamellibrachia_paraescarpia_0
K.txt
145
146 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\
t'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
mber" > orthogroups_annotations_losses_lamellibrachia_0fra_0alv_
Rpac_Pech_0fus_Ctel.csv
147 cat orthogroups_annotations_losses_lamellibrachia_oasisia.csv or
thogroups_annotations_losses_lamellibrachia_osedax.csv orthogrou
ps_annotations_losses_lamellibrachia_riftia.csv orthogroups_anno
tations_losses_lamellibrachia_paraescarpia.csv orthogroups_annot
ations_losses_lamellibrachia_owenia.csv orthogroups_annotations_
losses_lamellibrachia_capitella.csv >> orthogroups_annotations_l
osses_lamellibrachia_0fra_0alv_Rpac_Pech_0fus_Ctel.csv

```

```

148 sort orthogroups_annotations_losses_lamellibrachia_Ofra_0alv_Rpac_Pech_0fus_Ctel.csv > orthogroups_annotations_losses_lamellibrachia_Ofra_0alv_Rpac_Pech_0fus_Ctel_OK.csv
149
150 echo "Orthogroup"$(\t)"Panther_annotation" > orthogroups_mostAbundantAnnotation_losses_lamellibrachia.csv
151 while read line; do
152     orthogroup_ID=$(cut -f 1 <<< "$line")
153     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_losses_lamellibrachia_Ofra_0alv_Rpac_Pech_0fus_Ctel_OK.csv | cut -f 7 | sed '/^$/d' | sort | uniq -c | sort -r | awk '{$1=""; print $0}' | head -1)
154     echo $orthogroup_ID$(\t)$annotation >> orthogroups_mostAbundantAnnotation_losses_lamellibrachia.csv
155 done < gene_families_losses_lamellibrachia.csv

```

#### orthogroups\_annotations\_losses\_paraescarpia.sh

```

1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/Jul2021/losses/paraescarpia
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Jul2021/losses/paraescarpia
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=140:00:0
8  #$ -l highmem
9
10 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
    grep -w Pech | cut -f 1 > gene_families_losses_paraescarpia.txt
    #families losses in oasisia
11 fgrep -f gene_families_losses_paraescarpia.txt ../../Orthogroups.csv > gene_families_losses_paraescarpia.csv

```

```
12
13 cut -f 1,20 gene_families_losses_paraescarpia.csv > orthogroups_
    gene_IDs_losses_paraescarpia_osedax.txt
14 sed 's/Ofra|//g' orthogroups_gene_IDs_losses_paraescarpia_oseda
    x.txt > orthogroups_gene_IDs_losses_paraescarpia_osedax_OK.txt
15 cut -f 1,21 gene_families_losses_paraescarpia.csv > orthogroups_
    gene_IDs_losses_paraescarpia_owenia.txt
16 sed 's/Ofus|//g' orthogroups_gene_IDs_losses_paraescarpia_oweni
    a.txt > orthogroups_gene_IDs_losses_paraescarpia_owenia_OK.txt
17 cut -f 1,5 gene_families_losses_paraescarpia.csv > orthogroups_g
    ene_IDs_losses_paraescarpia_capitella.txt
18 sed 's/Ctel|//g' orthogroups_gene_IDs_losses_paraescarpia_capite
    lla.txt > orthogroups_gene_IDs_losses_paraescarpia_capitella_OK.
    txt
19 cut -f 1,19 gene_families_losses_paraescarpia.csv > orthogroups_
    gene_IDs_losses_paraescarpia_oasisia.txt
20 sed 's/Oalv|//g' orthogroups_gene_IDs_losses_paraescarpia_oasisi
    a.txt > orthogroups_gene_IDs_losses_paraescarpia_oasisia_OK.txt
21 cut -f 1,24 gene_families_losses_paraescarpia.csv > orthogroups_
    gene_IDs_losses_paraescarpia_riftia.txt
22 sed 's/Rpac|//g' orthogroups_gene_IDs_losses_paraescarpia_rifti
    a.txt > orthogroups_gene_IDs_losses_paraescarpia_riftia_OK.txt
23 cut -f 1,14 gene_families_losses_paraescarpia.csv > orthogroups_
    gene_IDs_losses_paraescarpia_lamellibrachia.txt
24 sed 's/Lluy|//g' orthogroups_gene_IDs_losses_paraescarpia_lamell
    ibrachia.txt > orthogroups_gene_IDs_losses_paraescarpia_lamellib
    rachia_OK.txt
25
26 while read line; do
27     genes=$(cut -f 2 <<< "$line")
28     echo $genes
29     orthogroup_ID=$(cut -f 1 <<< "$line")
30     echo $orthogroup_ID
```

```

31     if [[ "$genes" == FUN* ]]
32     then
33         IFS=', '      # space is set as delimiter
34         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
35         for gene in "${ADDR[@]}"; do
36             annotations=$(cut -f 13,14,15,18,20,24 ../../lamellibrach
ia_annotation_Feb2021_TrinoPantherK0_OK.xls | fgrep $gene)
37             echo $orthogroup_ID$'\t'"lamellibrachia"$'\t'$annotation
s >> orthogroups_annotations_losses_paraescarpia_lamellibrachia.
csv
38         done
39     else
40         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_paraescarpia_lamell
ibrachia.csv
41     fi
42 done < orthogroups_gene_IDs_losses_paraescarpia_lamellibrachia_0
K.txt
43
44 while read line; do
45     genes=$(cut -f 2 <<< "$line")
46     echo $genes
47     orthogroup_ID=$(cut -f 1 <<< "$line")
48     echo $orthogroup_ID
49     if [[ "$genes" == RPAC* ]]
50     then
51         IFS=', '      # space is set as delimiter
52         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
53         for gene in "${ADDR[@]}"; do
54             annotations=$(cut -f 13,14,15,18,20,24 ../../riftia_annot

```

```

ation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
55     echo $orthogroup_ID$'\t'"riftia"$'\t'$annotations >> ortho
groups_annotations_losses_paraescarpia_riftia.csv
56     done
57     else
58     echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_paraescarpia_rifti
a.csv
59     fi
60 done < orthogroups_gene_IDs_losses_paraescarpia_riftia_OK.txt
61
62 while read line; do
63     genes=$(cut -f 2 <<< "$line")
64     echo $genes
65     orthogroup_ID=$(cut -f 1 <<< "$line")
66     echo $orthogroup_ID
67     if [[ "$genes" == OALV* ]]
68     then
69         IFS=', '      # space is set as delimiter
70         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
71         for gene in "${ADDR[@]}"; do
72             annotations=$(cut -f 13,14,15,18,20,24 ../../oasisia_anno
tation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
73             echo $orthogroup_ID$'\t'"oasisia"$'\t'$annotations >> or
thogroups_annotations_losses_paraescarpia_oasisia.csv
74             done
75             else
76             echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_paraescarpia_oasisi
a.csv
77             fi

```

```

78 done < orthogroups_gene_IDs_losses_paraescarpia_oasisia_0K.txt
79
80 while read line; do
81     genes=$(cut -f 2 <<< "$line")
82     echo $genes
83     orthogroup_ID=$(cut -f 1 <<< "$line")
84     echo $orthogroup_ID
85     if [[ "$genes" == OFUS* ]]
86     then
87         IFS=', '      # space is set as delimiter
88         read -ra ADDR <<< "$genes"  # str is read into an array a
s tokens separated by IFS
89         for gene in "${ADDR[@]}; do
90             cut -f 1,2,3,11,12,13 ../Owenia_annotation_v250920.1_
TrinoPantherK0.xls | fgrep $gene > temp_file.txt
91             #K0_number=$(cut -f 1 temp_file.txt)
92             #gene_ID=$(cut -f 2 temp_file.txt)
93             #Panther_annotation=$(cut -f 3 temp_file.txt)
94             #GO_1=$(cut -f 4 temp_file.txt)
95             #GO_1=$(cut -f 5 temp_file.txt)
96             #GO_1=$(cut -f 6 temp_file.txt)
97             echo $orthogroup_ID$'\t'"owenia"$'\t'$(cut -f 4 temp_fil
e.txt)$'\t'$(cut -f 5 temp_file.txt)$'\t'$(cut -f 6 temp_file.tx
t)$'\t'$(cut -f 2 temp_file.txt)$'\t'$(cut -f 3 temp_file.tx
t)$'\t'$(cut -f 1 temp_file.txt) >> orthogroups_annotations_loss
es_paraescarpia_owenia.csv
98         done
99     else
100         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_paraescarpia_oweni
a.csv
101     fi

```

```

102 done < orthogroups_gene_IDs_losses_paraescarpia_owenia_OK.txt
103
104 while read line; do
105     genes=$(cut -f 2 <<< "$line")
106     echo $genes
107     orthogroup_ID=$(cut -f 1 <<< "$line")
108     echo $orthogroup_ID
109     if [[ "$genes" == CapteT* ]]
110     then
111         IFS=', '      # space is set as delimiter
112         read -ra ADDR <<< "$genes"  # str is read into an array a
s tokens separated by IFS
113         for gene in "${ADDR[@]}; do
114             cut -f 1,3,7,21,22 ../../Capitella_annotation_Feb2021_Tri
noPantherK0.xls | fgrep $gene > temp_file.txt
115             #K0_number=$(cut -f 7 temp_file.txt)
116             #gene_ID=$(cut -f 1 temp_file.txt)
117             #Panther_annotation=$(cut -f 3 temp_file.txt)
118             #GO_1=$(cut -f 21 temp_file.txt)
119             #GO_1=$(cut -f 22 temp_file.txt)
120             #GO_1=$(cut -f 6 temp_file.txt) NONE
121             echo $orthogroup_ID$'\t'"capitella"$'\t'$(cut -f 21 temp
_file.txt)$'\t'$(cut -f 22 temp_file.txt)$'\t'""$'\t'$(cut -f 1 t
emp_file.txt)$'\t'$(cut -f 3 temp_file.txt)$'\t'$(cut -f 7 temp_
file.txt) >> orthogroups_annotations_losses_paraescarpia_capitel
la.csv
122         done
123     else
124         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_paraescarpia_capite
lla.csv
125     fi

```

```

126 done < orthogroups_gene_IDs_losses_paraescarpia_capitella_OK.txt
127
128 while read line; do
129     genes=$(cut -f 2 <<< "$line")
130     echo $genes
131     orthogroup_ID=$(cut -f 1 <<< "$line")
132     echo $orthogroup_ID
133     if [[ "$genes" == OFRA* ]]
134     then
135         IFS=', '      # space is set as delimiter
136         read -ra ADDR <<< "$genes"  # str is read into an array a
s tokens separated by IFS
137         for gene in "${ADDR[@]}; do
138             annotations=$(cut -f 13,14,15,18,20,24 ../../osedax_annot
ation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
139             echo $orthogroup_ID$'\t'"osedax"$'\t'$annotations >> ort
hogroups_annotations_losses_paraescarpia_osedax.csv
140         done
141     else
142         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_paraescarpia_oseda
x.csv
143     fi
144 done < orthogroups_gene_IDs_losses_paraescarpia_osedax_OK.txt
145
146
147 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t
't'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
mber" > orthogroups_annotations_losses_paraescarpia_ofra_0alv_Rp
ac_Lluy_0fus_Ctel.csv
148 cat orthogroups_annotations_losses_paraescarpia_oasisia.csv orth
ogroups_annotations_losses_paraescarpia_osedax.csv orthogroups_a

```



```

nnotations_losses_paraescarpia_riftia.csv orthogroups_annotation
s_losses_paraescarpia_lamellibrachia.csv orthogroups_annotations
_losses_paraescarpia_owenia.csv orthogroups_annotations_losses_p
araescarpia_capitella.csv >> orthogroups_annotations_losses_para
escarpia_Ofra_Oalv_Rpac_Lluy_Ofus_Ctel.csv
149 sort orthogroups_annotations_losses_paraescarpia_Ofra_Oalv_Rpac_
Lluy_Ofus_Ctel.csv > orthogroups_annotations_losses_paraescarpia
_Ofra_Oalv_Rpac_Lluy_Ofus_Ctel_OK.csv
150
151 echo "Orthogroup"$(\t)"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_losses_paraescarpia.csv
152 while read line; do
153     orthogroup_ID=$(cut -f 1 <<< "$line")
154     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_los
ses_paraescarpia_Ofra_Oalv_Rpac_Lluy_Ofus_Ctel_OK.csv | cut -f 7
| sed '/^$/d' | sort | uniq -c | sort -r | awk '{ $1=""'; print
$0}' | head -1)
155     echo $orthogroup_ID$(\t)$annotation >> orthogroups_mostA
bundantAnnotation_losses_paraescarpia.csv
156 done < gene_families_losses_paraescarpia.csv

```

#### orthogroups\_annotations\_losses\_vestimentifera\_cl2.sh

```

1 #!/bin/bash
2 #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
ul2021/losses/vestimentifera_cl2
3 #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
l2021/losses/vestimentifera_cl2
4 #$ -j y
5 #$ -pe smp 1
6 #$ -l h_vmem=100G
7 #$ -l h_rt=140:00:0
8 #$ -l highmem
9

```

```
10 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |  
    grep -w Vestimentifera_cl2 | cut -f 1 > gene_families_losses_ves  
    timentifera_cl2.txt #families losses in oasisia  
11 fgrep -f gene_families_losses_vestimentifera_cl2.txt ../../Ortho  
    groups.csv > gene_families_losses_vestimentifera_cl2.csv  
12  
13 cut -f 1,20 gene_families_losses_vestimentifera_cl2.csv > orthog  
    rroups_gene_IDs_losses_vestimentifera_cl2_osedax.txt  
14 sed 's/Ofra|//g' orthogroups_gene_IDs_losses_vestimentifera_cl2_  
    osedax.txt > orthogroups_gene_IDs_losses_vestimentifera_cl2_osed  
    ax_OK.txt  
15 cut -f 1,21 gene_families_losses_vestimentifera_cl2.csv > orthog  
    rroups_gene_IDs_losses_vestimentifera_cl2_owenia.txt  
16 sed 's/Ofus|//g' orthogroups_gene_IDs_losses_vestimentifera_cl2_  
    owenia.txt > orthogroups_gene_IDs_losses_vestimentifera_cl2_owen  
    ia_OK.txt  
17 cut -f 1,5 gene_families_losses_vestimentifera_cl2.csv > orthogr  
    oups_gene_IDs_losses_vestimentifera_cl2_capitella.txt  
18 sed 's/Ctel|//g' orthogroups_gene_IDs_losses_vestimentifera_cl2_  
    capitella.txt > orthogroups_gene_IDs_losses_vestimentifera_cl2_c  
    apitella_OK.txt  
19 cut -f 1,14 gene_families_losses_vestimentifera_cl2.csv > orthog  
    rroups_gene_IDs_losses_vestimentifera_cl2_lamellibrachia.txt  
20 sed 's/Lluy|//g' orthogroups_gene_IDs_losses_vestimentifera_cl2_  
    lamellibrachia.txt > orthogroups_gene_IDs_losses_vestimentifera_  
    cl2_lamellibrachia_OK.txt  
21 cut -f 1,23 gene_families_losses_vestimentifera_cl2.csv > orthog  
    rroups_gene_IDs_losses_vestimentifera_cl2_paraescarpia.txt  
22 sed 's/Pech|//g' orthogroups_gene_IDs_losses_vestimentifera_cl2_  
    paraescarpia.txt > orthogroups_gene_IDs_losses_vestimentifera_cl  
    2_paraescarpia_OK.txt  
23  
24 while read line; do  
25     genes=$(cut -f 2 <<< "$line")
```

```

26     echo $genes
27     orthogroup_ID=$(cut -f 1 <<< "$line")
28     echo $orthogroup_ID
29     if [[ "$genes" == nbis* ]]
30     then
31         IFS=', '      # space is set as delimiter
32         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
33         for gene in "${ADDR[@]}"; do
34             annotations=$(cut -f 13,14,15,18,20,24 ../../paraescarpia
_annotation_Jun2021_TrinoPantherK0.xls | fgrep $gene)
35             echo $orthogroup_ID$'\t'"paraescarpia"$'\t'$annotations
>> orthogroups_annotations_losses_vestimentifera_cl2_paraescarpia.csv
36         done
37     else
38         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'
'""$'\t'"" >> orthogroups_annotations_losses_vestimentifera_cl2_
paraescarpia.csv
39     fi
40 done < orthogroups_gene_IDs_losses_vestimentifera_cl2_paraescarpia_OK.txt
41
42 while read line; do
43     genes=$(cut -f 2 <<< "$line")
44     echo $genes
45     orthogroup_ID=$(cut -f 1 <<< "$line")
46     echo $orthogroup_ID
47     if [[ "$genes" == FUN* ]]
48     then
49         IFS=', '      # space is set as delimiter
50         read -ra ADDR <<< "$genes"    # str is read into an array a

```

s tokens separated by IFS

```
51     for gene in "${ADDR[@]}"; do
52         annotations=$(cut -f 13,14,15,18,20,24 ../../lamellibrachia_annotation_Feb2021_TrinoPantherK0_OK.xls | fgrep $gene)
53         echo $orthogroup_ID$'\t'"lamellibrachia"$'\t'$annotations >> orthogroups_annotations_losses_vestimentifera_cl2_lamellibrachia.csv
54     done
55 else
56     echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'"" >> orthogroups_annotations_losses_vestimentifera_cl2_lamellibrachia.csv
57 fi
58 done < orthogroups_gene_IDs_losses_vestimentifera_cl2_lamellibrachia_OK.txt
59
60 while read line; do
61     genes=$(cut -f 2 <<< "$line")
62     echo $genes
63     orthogroup_ID=$(cut -f 1 <<< "$line")
64     echo $orthogroup_ID
65     if [[ "$genes" == OFUS* ]]
66     then
67         IFS=', '      # space is set as delimiter
68         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
69         for gene in "${ADDR[@]}"; do
70             cut -f 1,2,3,11,12,13 ../../Owenia_annotation_v250920.1_TrinoPantherK0.xls | fgrep $gene > temp_file.txt
71             #K0_number=$(cut -f 1 temp_file.txt)
72             #gene_ID=$(cut -f 2 temp_file.txt)
73             #Panther_annotation=$(cut -f 3 temp_file.txt)
```

```

74         #GO_1=$(cut -f 4 temp_file.txt)
75         #GO_1=$(cut -f 5 temp_file.txt)
76         #GO_1=$(cut -f 6 temp_file.txt)
77         echo $orthogroup_ID$'\t'"owenia"$'\t'$(cut -f 4 temp_file.txt)$'\t'$(cut -f 5 temp_file.txt)$'\t'$(cut -f 6 temp_file.txt)$'\t'$(cut -f 2 temp_file.txt)$'\t'$(cut -f 3 temp_file.txt)$'\t'$(cut -f 1 temp_file.txt) >> orthogroups_annotations_losses_vestimentifera_cl2_owenia.csv
78     done
79     else
80         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t'"" >> orthogroups_annotations_losses_vestimentifera_cl2_owenia.csv
81     fi
82 done < orthogroups_gene_IDs_losses_vestimentifera_cl2_owenia_OK.txt
83
84 while read line; do
85     genes=$(cut -f 2 <<< "$line")
86     echo $genes
87     orthogroup_ID=$(cut -f 1 <<< "$line")
88     echo $orthogroup_ID
89     if [[ "$genes" == CapteT* ]]
90     then
91         IFS=', '      # space is set as delimiter
92         read -ra ADDR <<< "$genes"    # str is read into an array as tokens separated by IFS
93         for gene in "${ADDR[@]}; do
94             cut -f 1,3,7,21,22 ../../Capitella_annotation_Feb2021_Tri
noPantherK0.xls | fgrep $gene > temp_file.txt
95             #K0_number=$(cut -f 7 temp_file.txt)
96             #gene_ID=$(cut -f 1 temp_file.txt)

```

```

97         #Panther_annotation=$(cut -f 3 temp_file.txt)
98         #GO_1=$(cut -f 21 temp_file.txt)
99         #GO_1=$(cut -f 22 temp_file.txt)
100        #GO_1=$(cut -f 6 temp_file.txt) NONE
101        echo $orthogroup_ID$'\t'"capitella"$'\t'$(cut -f 21 temp
_file.txt)$'\t'$(cut -f 22 temp_file.txt)$'\t'""'\t'$(cut -f 1 t
emp_file.txt)$'\t'$(cut -f 3 temp_file.txt)$'\t'$(cut -f 7 temp_
file.txt) >> orthogroups_annotations_losses_vestimentifera_cl2_c
apitella.csv
102        done
103        else
104        echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_vestimentifera_cl2_
capitella.csv
105        fi
106    done < orthogroups_gene_IDs_losses_vestimentifera_cl2_capitella_
OK.txt
107
108    while read line; do
109        genes=$(cut -f 2 <<< "$line")
110        echo $genes
111        orthogroup_ID=$(cut -f 1 <<< "$line")
112        echo $orthogroup_ID
113        if [[ "$genes" == OFRA* ]]
114        then
115            IFS=', '      # space is set as delimiter
116            read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
117            for gene in "${ADDR[@]}; do
118                annotations=$(cut -f 13,14,15,18,20,24 ../../osedax_annot
ation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
119                echo $orthogroup_ID$'\t'"osedax"$'\t'$annotations >> ort

```

```

hogroups_annotations_losses_vestimentifera_cl2_osedax.csv
120     done
121     else
122         echo $orthogroup_ID$'\t'"'"'$'\t'"'"'$'\t'"'"'$'\t'"'"'$'\t'"'"'$'\t'
        '"'"'$'\t'"'"' >> orthogroups_annotations_losses_vestimentifera_cl2_
osedax.csv
123     fi
124 done < orthogroups_gene_IDs_losses_vestimentifera_cl2_osedax_OK.
txt
125
126
127 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t'
\t'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
mber" > orthogroups_annotations_losses_vestimentifera_cl2_ofra_P
ech_Lluy_Ofus_Ctel.csv
128 cat orthogroups_annotations_losses_vestimentifera_cl2_osedax.csv
orthogroups_annotations_losses_vestimentifera_cl2_paraescarpia.c
sv orthogroups_annotations_losses_vestimentifera_cl2_lamellibrac
hia.csv orthogroups_annotations_losses_vestimentifera_cl2_oweni
a.csv orthogroups_annotations_losses_vestimentifera_cl2_capitell
a.csv >> orthogroups_annotations_losses_vestimentifera_cl2_ofra_P
ech_Lluy_Ofus_Ctel.csv
129 sort orthogroups_annotations_losses_vestimentifera_cl2_ofra_Pech
_Lluy_Ofus_Ctel.csv > orthogroups_annotations_losses_vestimentif
era_cl2_ofra_Pech_Lluy_Ofus_Ctel_OK.csv
130
131 echo "Orthogroup"$'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_losses_vestimentifera_cl2.csv
132 while read line; do
133     orthogroup_ID=$(cut -f 1 <<< "$line")
134     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_los
ses_vestimentifera_cl2_ofra_Pech_Lluy_Ofus_Ctel_OK.csv | cut -f
7 | sed '/^$/d' | sort | uniq -c | sort -r | awk '{s1=""}; print
$s0}' | head -1)

```

```

135     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA
bundantAnnotation_losses_vestimentifera_cl2.csv
136 done < gene_families_losses_vestimentifera_cl2.csv

```

#### orthogroups\_annotations\_losses\_vestimentifera\_cl1.sh

```

1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
ul2021/losses/vestimentifera_cl1
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
l2021/losses/vestimentifera_cl1
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=140:00:0
8  #$ -l highmem
9
10 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
grep -w Vestimentifera_cl1 | cut -f 1 > gene_families_losses_ves
timentifera_cl1.txt #families losses in oasisia
11 fgrep -f gene_families_losses_vestimentifera_cl1.txt ../../Ortho
groups.csv > gene_families_losses_vestimentifera_cl1.csv
12
13 cut -f 1,20 gene_families_losses_vestimentifera_cl1.csv > orthog
roups_gene_IDs_losses_vestimentifera_cl1_osedax.txt
14 sed 's/Ofra|//g' orthogroups_gene_IDs_losses_vestimentifera_cl1_
osedax.txt > orthogroups_gene_IDs_losses_vestimentifera_cl1_osed
ax_OK.txt
15 cut -f 1,21 gene_families_losses_vestimentifera_cl1.csv > orthog
roups_gene_IDs_losses_vestimentifera_cl1_owenia.txt
16 sed 's/Ofus|//g' orthogroups_gene_IDs_losses_vestimentifera_cl1_
owenia.txt > orthogroups_gene_IDs_losses_vestimentifera_cl1_owen
ia_OK.txt
17 cut -f 1,5 gene_families_losses_vestimentifera_cl1.csv > orthogr

```



```

oups_gene_IDs_losses_vestimentifera_cl1_capitella.txt
18 sed 's/Ctel|//g' orthogroups_gene_IDs_losses_vestimentifera_cl1_
   capitella.txt > orthogroups_gene_IDs_losses_vestimentifera_cl1_c
   apitella_OK.txt
19 cut -f 1,14 gene_families_losses_vestimentifera_cl1.csv > orthog
   rroups_gene_IDs_losses_vestimentifera_cl1_lamellibrachia.txt
20 sed 's/Lluy|//g' orthogroups_gene_IDs_losses_vestimentifera_cl1_
   lamellibrachia.txt > orthogroups_gene_IDs_losses_vestimentifera_
   cl1_lamellibrachia_OK.txt
21
22 while read line; do
23     genes=$(cut -f 2 <<< "$line")
24     echo $genes
25     orthogroup_ID=$(cut -f 1 <<< "$line")
26     echo $orthogroup_ID
27     if [[ "$genes" == FUN* ]]
28     then
29         IFS=', '      # space is set as delimiter
30         read -ra ADDR <<< "$genes"  # str is read into an array a
   s tokens separated by IFS
31         for gene in "${ADDR[@]}; do
32             annotations=$(cut -f 13,14,15,18,20,24 ../../lamellibrach
   ia_annotation_Feb2021_TrinoPantherK0_OK.xls | fgrep $gene)
33             echo $orthogroup_ID$'\t'"lamellibrachia"$'\t'$annotations
   s >> orthogroups_annotations_losses_vestimentifera_cl1_lamellibr
   achia.csv
34         done
35     else
36         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
   '""$'\t'"" >> orthogroups_annotations_losses_vestimentifera_cl1_
   lamellibrachia.csv
37     fi
38 done < orthogroups_gene_IDs_losses_vestimentifera_cl1_lamellibra

```

chia\_OK.txt

```
39
40 while read line; do
41     genes=$(cut -f 2 <<< "$line")
42     echo $genes
43     orthogroup_ID=$(cut -f 1 <<< "$line")
44     echo $orthogroup_ID
45     if [[ "$genes" == OFUS* ]]
46     then
47         IFS=', '      # space is set as delimiter
48         read -ra ADDR <<< "$genes"  # str is read into an array a
s tokens separated by IFS
49         for gene in "${ADDR[@]}"; do
50             cut -f 1,2,3,11,12,13 ../Owenia_annotation_v250920.1_
TrinoPantherK0.xls | fgrep $gene > temp_file.txt
51             #K0_number=$(cut -f 1 temp_file.txt)
52             #gene_ID=$(cut -f 2 temp_file.txt)
53             #Panther_annotation=$(cut -f 3 temp_file.txt)
54             #GO_1=$(cut -f 4 temp_file.txt)
55             #GO_1=$(cut -f 5 temp_file.txt)
56             #GO_1=$(cut -f 6 temp_file.txt)
57             echo $orthogroup_ID$'\t'"owenia"$'\t'$(cut -f 4 temp_fil
e.txt)$'\t'$(cut -f 5 temp_file.txt)$'\t'$(cut -f 6 temp_file.tx
t)$'\t'$(cut -f 2 temp_file.txt)$'\t'$(cut -f 3 temp_file.tx
t)$'\t'$(cut -f 1 temp_file.txt) >> orthogroups_annotations_loss
es_vestimentifera_cl1_owenia.csv
58             done
59         else
60             echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_vestimentifera_cl1_
owenia.csv
61         fi
```

```

62 done < orthogroups_gene_IDs_losses_vestimentifera_cl1_owenia_OK.
    txt
63
64 while read line; do
65     genes=$(cut -f 2 <<< "$line")
66     echo $genes
67     orthogroup_ID=$(cut -f 1 <<< "$line")
68     echo $orthogroup_ID
69     if [[ "$genes" == CapteT* ]]
70     then
71         IFS=', '      # space is set as delimiter
72         read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
73         for gene in "${ADDR[@]}"; do
74             cut -f 1,3,7,21,22 ../../Capitella_annotation_Feb2021_Tri
noPantherK0.xls | fgrep $gene > temp_file.txt
75             #K0_number=$(cut -f 7 temp_file.txt)
76             #gene_ID=$(cut -f 1 temp_file.txt)
77             #Panther_annotation=$(cut -f 3 temp_file.txt)
78             #GO_1=$(cut -f 21 temp_file.txt)
79             #GO_1=$(cut -f 22 temp_file.txt)
80             #GO_1=$(cut -f 6 temp_file.txt) NONE
81             echo $orthogroup_ID$'\t'"capitella"$'\t'$(cut -f 21 temp
_file.txt)$'\t'$(cut -f 22 temp_file.txt)$'\t'""$'\t'$(cut -f 1 t
emp_file.txt)$'\t'$(cut -f 3 temp_file.txt)$'\t'$(cut -f 7 temp_
file.txt) >> orthogroups_annotations_losses_vestimentifera_cl1_c
apitella.csv
82         done
83     else
84         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_vestimentifera_cl1_
capitella.csv

```

```

85         fi
86     done < orthogroups_gene_IDs_losses_vestimentifera_cl1_capitella_
      OK.txt
87
88     while read line; do
89         genes=$(cut -f 2 <<< "$line")
90         echo $genes
91         orthogroup_ID=$(cut -f 1 <<< "$line")
92         echo $orthogroup_ID
93         if [[ "$genes" == OFRA* ]]
94         then
95             IFS=', '      # space is set as delimiter
96             read -ra ADDR <<< "$genes"    # str is read into an array a
      s tokens separated by IFS
97             for gene in "${ADDR[@]}"; do
98                 annotations=$(cut -f 13,14,15,18,20,24 ../../osedax_annot
      ation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
99                 echo $orthogroup_ID$'\t'"osedax"$'\t'$annotations >> ort
      hogroups_annotations_losses_vestimentifera_cl1_osedax.csv
100             done
101         else
102             echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
      '""$'\t'"" >> orthogroups_annotations_losses_vestimentifera_cl1_
      osedax.csv
103         fi
104     done < orthogroups_gene_IDs_losses_vestimentifera_cl1_osedax_OK.
      txt
105
106
107     echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t'
      t'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
      mber" > orthogroups_annotations_losses_vestimentifera_cl1_ofra_L

```

```

luy_Ofus_Ctel.csv
108 cat orthogroups_annotations_losses_vestimentifera_cl1_osedax.csv
orthogroups_annotations_losses_vestimentifera_cl1_lamellibrachi
a.csv orthogroups_annotations_losses_vestimentifera_cl1_owenia.c
sv orthogroups_annotations_losses_vestimentifera_cl1_capitella.c
sv >> orthogroups_annotations_losses_vestimentifera_cl1_Ofra_Lluy
y_Ofus_Ctel.csv
109 sort orthogroups_annotations_losses_vestimentifera_cl1_Ofra_Lluy
_Ofus_Ctel.csv > orthogroups_annotations_losses_vestimentifera_c
l1_Ofra_Lluy_Ofus_Ctel_OK.csv
110
111 echo "Orthogroup"$('\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_losses_vestimentifera_cl1.csv
112 while read line; do
113     orthogroup_ID=$(cut -f 1 <<< "$line")
114     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_lo
ses_vestimentifera_cl1_Ofra_Lluy_Ofus_Ctel_OK.csv | cut -f 7 | s
ed '/^$/d' | sort | uniq -c | sort -r | awk '{ $1="" ; print $0 }'
| head -1)
115     echo $orthogroup_ID$('\t'$annotation >> orthogroups_mostA
bundantAnnotation_losses_vestimentifera_cl1.csv
116 done < gene_families_losses_vestimentifera_cl1.csv

```

#### orthogroups\_annotations\_losses\_vestimentifera.sh

```

1  #!/bin/bash
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J
ul2021/losses/vestimentifera
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju
l2021/losses/vestimentifera
4  #$ -j y
5  #$ -pe smp 1
6  #$ -l h_vmem=100G
7  #$ -l h_rt=140:00:0

```

```

8  # $ -l highmem
9
10 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
    grep -w Vestimentifera | cut -f 1 > gene_families_losses_vestime
    ntifera.txt #families losses in oasisia
11 fgrep -f gene_families_losses_vestimentifera.txt ../../Orthogrou
    ps.csv > gene_families_losses_vestimentifera.csv
12
13 cut -f 1,20 gene_families_losses_vestimentifera.csv > orthogroup
    s_gene_IDs_losses_vestimentifera_osedax.txt
14 sed 's/Ofra|//g' orthogroups_gene_IDs_losses_vestimentifera_osed
    ax.txt > orthogroups_gene_IDs_losses_vestimentifera_osedax_OK.tx
    t
15 cut -f 1,21 gene_families_losses_vestimentifera.csv > orthogroup
    s_gene_IDs_losses_vestimentifera_owenia.txt
16 sed 's/Ofus|//g' orthogroups_gene_IDs_losses_vestimentifera_owen
    ia.txt > orthogroups_gene_IDs_losses_vestimentifera_owenia_OK.tx
    t
17 cut -f 1,5 gene_families_losses_vestimentifera.csv > orthogroups
    _gene_IDs_losses_vestimentifera_capitella.txt
18 sed 's/Ctel|//g' orthogroups_gene_IDs_losses_vestimentifera_capi
    tella.txt > orthogroups_gene_IDs_losses_vestimentifera_capitella
    _OK.txt
19
20 while read line; do
21     genes=$(cut -f 2 <<< "$line")
22     echo $genes
23     orthogroup_ID=$(cut -f 1 <<< "$line")
24     echo $orthogroup_ID
25     if [[ "$genes" == OFUS* ]]
26     then
27         IFS=', '      # space is set as delimiter
28         read -ra ADDR <<< "$genes"    # str is read into an array a

```

s tokens separated by IFS

```
29     for gene in "${ADDR[@]"}"; do
30         cut -f 1,2,3,11,12,13 ../../Owenia_annotation_v250920.1_
TrinoPantherK0.xls | fgrep $gene > temp_file.txt
31         #K0_number=$(cut -f 1 temp_file.txt)
32         #gene_ID=$(cut -f 2 temp_file.txt)
33         #Panther_annotation=$(cut -f 3 temp_file.txt)
34         #GO_1=$(cut -f 4 temp_file.txt)
35         #GO_1=$(cut -f 5 temp_file.txt)
36         #GO_1=$(cut -f 6 temp_file.txt)
37         echo $orthogroup_ID$'\t'"owenia"$'\t'$(cut -f 4 temp_fil
e.txt)$'\t'$(cut -f 5 temp_file.txt)$'\t'$(cut -f 6 temp_file.tx
t)$'\t'$(cut -f 2 temp_file.txt)$'\t'$(cut -f 3 temp_file.tx
t)$'\t'$(cut -f 1 temp_file.txt) >> orthogroups_annotations_loss
es_vestimentifera_owenia.csv
38     done
39     else
40         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_vestimentifera_owen
ia.csv
41     fi
42 done < orthogroups_gene_IDs_losses_vestimentifera_owenia_OK.txt
43
44 while read line; do
45     genes=$(cut -f 2 <<< "$line")
46     echo $genes
47     orthogroup_ID=$(cut -f 1 <<< "$line")
48     echo $orthogroup_ID
49     if [[ "$genes" == CapteT* ]]
50     then
51         IFS=', '      # space is set as delimiter
52         read -ra ADDR <<< "$genes"  # str is read into an array a
```

s tokens separated by IFS

```
53     for gene in "${ADDR[@]"}"; do
54         cut -f 1,3,7,21,22 ../../Capitella_annotation_Feb2021_Tri
noPantherK0.xls | fgrep $gene > temp_file.txt
55         #K0_number=$(cut -f 7 temp_file.txt)
56         #gene_ID=$(cut -f 1 temp_file.txt)
57         #Panther_annotation=$(cut -f 3 temp_file.txt)
58         #GO_1=$(cut -f 21 temp_file.txt)
59         #GO_1=$(cut -f 22 temp_file.txt)
60         #GO_1=$(cut -f 6 temp_file.txt) NONE
61         echo $orthogroup_ID$'\t'"capitella"$'\t'$(cut -f 21 temp
_file.txt)$'\t'$(cut -f 22 temp_file.txt)$'\t'""$'\t'$(cut -f 1 t
emp_file.txt)$'\t'$(cut -f 3 temp_file.txt)$'\t'$(cut -f 7 temp_
file.txt) >> orthogroups_annotations_losses_vestimentifera_capit
ella.csv
62         done
63     else
64         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_vestimentifera_capi
tella.csv
65     fi
66 done < orthogroups_gene_IDs_losses_vestimentifera_capitella_0K.t
xt
67
68 while read line; do
69     genes=$(cut -f 2 <<< "$line")
70     echo $genes
71     orthogroup_ID=$(cut -f 1 <<< "$line")
72     echo $orthogroup_ID
73     if [[ "$genes" == OFRA* ]]
74     then
75         IFS=', '      # space is set as delimiter
```



```

76     read -ra ADDR <<< "$genes"    # str is read into an array a
s tokens separated by IFS
77     for gene in "${ADDR[@]}"; do
78         annotations=$(cut -f 13,14,15,18,20,24 ../../osedax_annot
ation_Jan2021_TrinoPantherK0.xls | fgrep $gene)
79         echo $orthogroup_ID$'\t'"osedax"$'\t'$annotations >> ort
hogroups_annotations_losses_vestimentifera_osedax.csv
80     done
81     else
82         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_vestimentifera_osed
ax.csv
83     fi
84 done < orthogroups_gene_IDs_losses_vestimentifera_osedax_OK.txt
85
86
87 echo "Orthogroup"$'\t'"Species"$'\t'"GO_term1"$'\t'"GO_term2"$'\t'
t'"GO_term3"$'\t'"gene_ID"$'\t'"Panther_annotation"$'\t'"KEGG_nu
mber" > orthogroups_annotations_losses_vestimentifera_ofra_ofus_
Ctel.csv
88 cat orthogroups_annotations_losses_vestimentifera_osedax.csv ort
hogroups_annotations_losses_vestimentifera_owenia.csv orthogroup
s_annotations_losses_vestimentifera_capitella.csv >> orthogroups
_annotations_losses_vestimentifera_ofra_ofus_Ctel.csv
89 sort orthogroups_annotations_losses_vestimentifera_ofra_ofus_Cte
l.csv > orthogroups_annotations_losses_vestimentifera_ofra_ofus_
Ctel_OK.csv
90
91 echo "Orthogroup"$'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_losses_vestimentifera.csv
92 while read line; do
93     orthogroup_ID=$(cut -f 1 <<< "$line")
94     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_los

```

```
ses_vestimentifera_Ofra_Ofus_Ctel_OK.csv | cut -f 7 | sed '/^$/d  
' | sort | uniq -c | sort -r | awk '{ $1="" ; print $0 }' | head -  
1)
```

```
95     echo $orthogroup_ID$'\t'$annotation >> orthogroups_mostA  
    bundantAnnotation_losses_vestimentifera.csv  
96 done < gene_families_losses_vestimentifera.csv
```

#### orthogroups\_annotations\_losses\_siboglinidae.sh

```
1  #!/bin/bash  
2  #$ -wd /data/scratch/btx654/gene_family_evolution/ferdi_script/J  
    ul2021/losses/siboglinidae  
3  #$ -o /data/scratch/btx654/gene_family_evolution/ferdi_script/Ju  
    l2021/losses/siboglinidae  
4  #$ -j y  
5  #$ -pe smp 1  
6  #$ -l h_vmem=100G  
7  #$ -l h_rt=140:00:0  
8  #$ -l highmem  
9  
10 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |  
    grep -w Siboglinidae | cut -f 1 > gene_families_losses_siboglini  
    dae.txt #families losses in oasisia  
11 fgrep -f gene_families_losses_siboglinidae.txt ../../Orthogroup  
    s.csv > gene_families_losses_siboglinidae.csv  
12  
13 cut -f 1,21 gene_families_losses_siboglinidae.csv > orthogroups_  
    gene_IDs_losses_siboglinidae_owenia.txt  
14 sed 's/Ofus|//g' orthogroups_gene_IDs_losses_siboglinidae_oweni  
    a.txt > orthogroups_gene_IDs_losses_siboglinidae_owenia_OK.txt  
15 cut -f 1,5 gene_families_losses_siboglinidae.csv > orthogroups_g  
    ene_IDs_losses_siboglinidae_capitella.txt  
16 sed 's/Ctel|//g' orthogroups_gene_IDs_losses_siboglinidae_capite  
    lla.txt > orthogroups_gene_IDs_losses_siboglinidae_capitella_OK.
```

txt

```
17
18 while read line; do
19     genes=$(cut -f 2 <<< "$line")
20     echo $genes
21     orthogroup_ID=$(cut -f 1 <<< "$line")
22     echo $orthogroup_ID
23     if [[ "$genes" == OFUS* ]]
24     then
25         IFS=', '      # space is set as delimiter
26         read -ra ADDR <<< "$genes"  # str is read into an array a
s tokens separated by IFS
27         for gene in "${ADDR[@]}"; do
28             cut -f 1,2,3,11,12,13 ../Owenia_annotation_v250920.1_
TrinoPantherK0.xls | fgrep $gene > temp_file.txt
29             #K0_number=$(cut -f 1 temp_file.txt)
30             #gene_ID=$(cut -f 2 temp_file.txt)
31             #Panther_annotation=$(cut -f 3 temp_file.txt)
32             #GO_1=$(cut -f 4 temp_file.txt)
33             #GO_1=$(cut -f 5 temp_file.txt)
34             #GO_1=$(cut -f 6 temp_file.txt)
35             echo $orthogroup_ID$'\t'"owenia"$'\t'$(cut -f 4 temp_fil
e.txt)$'\t'$(cut -f 5 temp_file.txt)$'\t'$(cut -f 6 temp_file.tx
t)$'\t'$(cut -f 2 temp_file.txt)$'\t'$(cut -f 3 temp_file.tx
t)$'\t'$(cut -f 1 temp_file.txt) >> orthogroups_annotations_loss
es_siboglinidae_owenia.csv
36             done
37         else
38             echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_siboglinidae_oweni
a.csv
39         fi
```

```

40 done < orthogroups_gene_IDs_losses_siboglinidae_owenia_OK.txt
41
42 while read line; do
43     genes=$(cut -f 2 <<< "$line")
44     echo $genes
45     orthogroup_ID=$(cut -f 1 <<< "$line")
46     echo $orthogroup_ID
47     if [[ "$genes" == CapteT* ]]
48     then
49         IFS=', '      # space is set as delimiter
50         read -ra ADDR <<< "$genes"  # str is read into an array as
tokens separated by IFS
51         for gene in "${ADDR[@]}; do
52             cut -f 1,3,7,21,22 ../../Capitella_annotation_Feb2021_Tri
noPantherK0.xls | fgrep $gene > temp_file.txt
53             #K0_number=$(cut -f 7 temp_file.txt)
54             #gene_ID=$(cut -f 1 temp_file.txt)
55             #Panther_annotation=$(cut -f 3 temp_file.txt)
56             #GO_1=$(cut -f 21 temp_file.txt)
57             #GO_1=$(cut -f 22 temp_file.txt)
58             #GO_1=$(cut -f 6 temp_file.txt) NONE
59             echo $orthogroup_ID$'\t'"capitella"$'\t'$(cut -f 21 temp
_file.txt)$'\t'$(cut -f 22 temp_file.txt)$'\t'""$'\t'$(cut -f 1 t
emp_file.txt)$'\t'$(cut -f 3 temp_file.txt)$'\t'$(cut -f 7 temp_
file.txt) >> orthogroups_annotations_losses_siboglinidae_capitel
la.csv
60         done
61     else
62         echo $orthogroup_ID$'\t'""$'\t'""$'\t'""$'\t'""$'\t'""$'\t
'""$'\t'"" >> orthogroups_annotations_losses_siboglinidae_capite
lla.csv
63     fi

```

```

64 done < orthogroups_gene_IDs_losses_siboglinidae_capitella_OK.txt
65
66
67 echo "Orthogroup"'\t'"Species"'\t'"GO_term1"'\t'"GO_term2"'\t'
GO_term3"'\t'"gene_ID"'\t'"Panther_annotation"'\t'"KEGG_number" > orthogroups_
annotations_losses_siboglinidae_ofus_Ctel.csv
68 cat orthogroups_annotations_losses_siboglinidae_owenia.csv ortho
groups_annotations_losses_siboglinidae_capitella.csv >> orthogro
ups_annotations_losses_siboglinidae_ofus_Ctel.csv
69 sort orthogroups_annotations_losses_siboglinidae_ofus_Ctel.csv >
orthogroups_annotations_losses_siboglinidae_ofus_Ctel_OK.csv
70
71 echo "Orthogroup"'\t'"Panther_annotation" > orthogroups_mostAbu
ndantAnnotation_losses_siboglinidae.csv
72 while read line; do
73     orthogroup_ID=$(cut -f 1 <<< "$line")
74     annotation=$(fgrep $orthogroup_ID orthogroups_annotations_lo
ses_siboglinidae_ofus_Ctel_OK.csv | cut -f 7 | sed '/^$/d' | sor
t | uniq -c | sort -r | awk '{ $1=""; print $0 }' | head -1)
75     echo $orthogroup_ID'\t'$annotation >> orthogroups_mostA
bundantAnnotation_losses_siboglinidae.csv
76 done < gene_families_losses_siboglinidae.csv

```

## GO\_terms

### Base code:

```

1 install.packages("BiocManager")
2 BiocManager::install("topGO")
3 install.packages("ggpubr")
4

```

```

5 library(topGO)
6 library(ggplot2)
7 library(ggpubr)
8 library(cowplot)
9
10 # Import gene universe: whole (GO-annotated) genome
11 geneID2GO <- readMappings(file = "/Users/giacomo/Desktop/R/GO_enrichment/osedax/osedax_GO_universe.txt") ### 21108 transcripts have GO annotation
12 geneUniverse <- names(geneID2GO)
13
14 # Import and transform genes of interest: 8 clusters from step2a
15 cluster1 <- read.table("/Users/giacomo/Desktop/R/GO_enrichment/osedax/gene_IDs_originated_osedax_Siboglinidae.txt",header=FALSE)
16 cluster1 <- as.character(cluster1$V1)
17 cluster1genelist <- factor(as.integer(geneUniverse %in% cluster1))
18 names(cluster1genelist) <- geneUniverse
19
20 # fisher testing of GO term enrichment for Molecular Function (MF)
21 #cluster 1 - red genes
22 cluster1_G0data_MF <- new("topGOdata", description="Cluster1",
23                           ontology="MF", allGenes=cluster1genelist,
24                           annot = annFUN.gene2GO, gene2GO = geneID2GO)
25 cluster1_resultFisher_MF <- runTest(cluster1_G0data_MF,
26                                     algorithm="classic", statistic="fisher")
27 cluster1_MF <- GenTable(cluster1_G0data_MF, classicFisher = cluster1_resultFisher_MF,

```

```

28         orderBy = "resultFisher", ranksOf = "cla
    ssicFisher", topNodes = 15)
29
30 cluster1_MF[cluster1_MF == "< 1e-30"] <- "1e-30"
31 cluster1_MF[cluster1_MF == "<1e-30"] <- "1e-30"
32
33 goEnrichment <- cluster1_MF
34 goEnrichment$classicFisher <- as.numeric(goEnrichment$classicFis
    her)
35 goEnrichment <- goEnrichment[,c("GO.ID", "Term", "classicFisher")]
36 goEnrichment$Term <- gsub("[a-z]*\\.\\.\\.\\.$", "", goEnrichmen
    t$Term)
37 goEnrichment$Term <- gsub("\\\\.\\.\\.\\.$", "", goEnrichment$Term)
38 goEnrichment$Term <- paste(goEnrichment$GO.ID, goEnrichment$Ter
    m, sep=", ")
39 goEnrichment$Term <- factor(goEnrichment$Term, levels=rev(goEnri
    chment$Term))
40
41 #it could happen that the second line of the previous block will
    give the error " Warning message:NAs introduced by coercion "
42 # a fix for that is " goEnrichment$classicFisher <- c(30, 30, 3
    0, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30) "
43
44 cluster_1_plot <- ggplot(goEnrichment, aes(x=Term, y=-log10(clas
    sicFisher))) +
45   stat_summary(geom = "bar", fun = mean, position = "dodge") +
46   xlab("Molecular Function") +
47   ylab("-log10(p-value)") +
48   ggtitle("GF gains in Siboglinidae (Osedax)") +
49   scale_y_continuous(limits=c(0,30), breaks=round(seq(0,30, by =
    2), 1)) +
50   theme_classic() +

```

```

51   theme(
52     legend.position='none',
53     legend.background=element_rect(),
54     plot.title=element_text(angle=0, size=12, face="bold", vjust
=1),
55     axis.text.x=element_text(angle=0, size=10, hjust=1.10),
56     axis.text.y=element_text(angle=0, size=10, vjust=0.5),
57     axis.title=element_text(size=12),
58     legend.key=element_blank(),      #removes the border
59     legend.key.size=unit(1, "cm"),    #Sets overall area/size
of the legend
60     legend.text=element_text(size=12), #Text size
61     title=element_text(size=12)) +
62     guides(colour=guide_legend(override.aes=list(size=2.5))) +
63     coord_flip()
64
65 cluster_1_plot
66
67 cluster1_G0data_BP <- new("topG0data", description="Cluster1",
68                           ontology="BP", allGenes=cluster1geneli
st,
69                           annot = annFUN.gene2GO, gene2GO = gene
ID2GO)
70 cluster1_resultFisher_BP <- runTest(cluster1_G0data_BP,
71                                     algorithm="classic", statist
ic="fisher")
72 cluster1_BP <- GenTable(cluster1_G0data_BP, classicFisher = clus
ter1_resultFisher_BP,
73                           orderBy = "resultFisher", ranksOf = "cla
ssicFisher", topNodes = 15)
74
75 cluster1_BP[cluster1_BP == "< 1e-30"] <- "1e-30"

```



```

76 cluster1_BP[cluster1_BP == "<1e-30"] <- "1e-30"
77
78 goEnrichment <- cluster1_BP
79 goEnrichment$classicFisher <- as.numeric(goEnrichment$classicFisher)
80 goEnrichment <- goEnrichment[,c("GO.ID", "Term", "classicFisher")]
81 goEnrichment$Term <- gsub("[a-z]*\\.\\.\\.\\$", "", goEnrichment$Term)
82 goEnrichment$Term <- gsub("\\\\.\\.\\.\\$", "", goEnrichment$Term)
83 goEnrichment$Term <- paste(goEnrichment$GO.ID, goEnrichment$Term, sep=", ")
84 goEnrichment$Term <- factor(goEnrichment$Term, levels=rev(goEnrichment$Term))
85
86 #it could happen that the second line of the previous block will
87 # give the error " Warning message:NAs introduced by coercion "
88 # a fix for that is " goEnrichment$classicFisher <- c(30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30, 30) "
89
90 cluster_1BP_plot <- ggplot(goEnrichment, aes(x=Term, y=-log10(classicFisher))) +
91   stat_summary(geom = "bar", fun = mean, position = "dodge") +
92   xlab("Biological Process") +
93   ylab("-log10(p-value)") +
94   scale_y_continuous(limits=c(0,30), breaks=round(seq(0,30, by = 2), 1)) +
95   theme_classic() +
96   theme(
97     legend.position='none',
98     legend.background=element_rect(),
99     plot.title=element_text(angle=0, size=12, face="bold", vjust=1),

```

```

99     axis.text.x=element_text(angle=0, size=10, hjust=1.10),
100     axis.text.y=element_text(angle=0, size=10, vjust=0.5),
101     axis.title=element_text(size=12),
102     legend.key=element_blank(),      #removes the border
103     legend.key.size=unit(1, "cm"),    #Sets overall area/size
of the legend
104     legend.text=element_text(size=12), #Text size
105     title=element_text(size=12)) +
106     guides(colour=guide_legend(override.aes=list(size=2.5))) +
107     coord_flip()
108
109 cluster_1BP_plot
110
111 plot_grid(cluster_1_plot + rremove("x.title"),
112           cluster_1BP_plot,
113           ncol = 1, align="v")

```

- exported in pdf 7x8 inches

## Riftia

```

1 cd /data/scratch/btx654/gene_family_evolution/ferdi_script/Jul20
21/GO_terms/riftia
2 # I will use BlastX GO terms
3 cut -f2,13 ../../riftia_annotation_Jan2021_TrinoPantherK0.xls |
tail -n +2 > riftia_GO_raw.txt # 38179 genes in riftia_GO_raw.tx
t
4 grep 'GO' riftia_GO_raw.txt > riftia_GO_only_raw.txt # 20737 rif
tia_GO_only_raw.txt
5

```

## python.py

```

1 if __name__ == "__main__":
2

```

```

3 import re
4
5 i = open("riftia_GO_only_raw.txt", "r")
6 o = open("riftia_GO_universe.txt", "w")
7
8 regex = re.compile(r'GO:\d+')
9
10 for line in i:
11     GOMatches = regex.findall(line)
12     Gene_ID = line.split("\t",1)[0]
13     if not GOMatches == []:
14         o.write(Gene_ID+"\t")
15         for i, match in enumerate(GOMatches):
16             if i+1 == len(GOMatches):
17                 o.write(match.strip("'")+"\n")
18             else:
19                 o.write(match.strip("'")+", ")

```

```

1 module load python
2 python python.py

```

Now we have the file riftia\_GO\_universe.txt and we need to select a subgroup of genes:

## Expansions

```

1 cut -f 1,7 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
  grep -w Rpac | cut -f 1 > gene_families_expanded_riftia.txt #fam
  ilies expanded in riftia
2 fgrep -f gene_families_expanded_riftia.txt ../../Orthogroups.csv
  > gene_families_expanded_riftia.csv
3 cut -f 1,24 gene_families_expanded_riftia.csv | sed 's/Rpac//g'
  | cut -f 2 | sed 's/, /\n/g' > gene_IDs_expanded_riftia.txt

```

R script:

```
install.packages("BiocManager")  
BiocManager::install("topGO")  
install.packages("ggpubr")  
  
library(topGO)  
library(ggplot2)  
library(ggpubr)  
library(cowplot)  
  
# Import gene universe: whole (GO-annotated) genome  
geneID2GO <- readMappings(file = "/Users/giacomo/Desktop/R/GO_enrichment/riftia_GO_universe.txt") ### 21108 transcripts have GO annotation  
geneUniverse <- names(geneID2GO)  
  
# Import and transform genes of interest: 8 clusters from step2a  
cluster1 <- read.table("/Users/giacomo/Desktop/R/GO_enrichment/gene_IDs_expanded_riftia.txt", header=FALSE)  
cluster1 <- as.character(cluster1$V1)  
cluster1genelist <- factor(as.integer(geneUniverse %in% cluster1))  
names(cluster1genelist) <- geneUniverse  
  
# fisher testing of GO term enrichment for Molecular Function (MF)  
#cluster 1 - red genes  
cluster1_G0data_MF <- new("topGOdata", description="Cluster1",  
                           ontology="MF", allGenes=cluster1genelist,  
                           annot = annFUN.gene2GO, gene2GO = gene
```

```

ID2G0)
25 cluster1_resultFisher_MF <- runTest(cluster1_G0data_MF,
26                                     algorithm="classic", statist
ic="fisher")
27 cluster1_MF <- GenTable(cluster1_G0data_MF, classicFisher = clus
ter1_resultFisher_MF,
28                             orderBy = "resultFisher", ranksOf = "cla
ssicFisher", topNodes = 15)
29
30 cluster1_MF[cluster1_MF == "< 1e-30"] <- "1e-30"
31
32 goEnrichment <- cluster1_MF
33 goEnrichment$classicFisher <- as.numeric(goEnrichment$classicFis
her)
34 goEnrichment <- goEnrichment[,c("GO.ID", "Term", "classicFisher")]
35 goEnrichment$Term <- gsub(" [a-z]*\\.\\.\\.\\.\\$", "", goEnrichmen
t$Term)
36 goEnrichment$Term <- gsub("\\\\.\\.\\.\\.\\$", "", goEnrichment$Term)
37 goEnrichment$Term <- paste(goEnrichment$GO.ID, goEnrichment$Ter
m, sep=", ")
38 goEnrichment$Term <- factor(goEnrichment$Term, levels=rev(goEnri
chment$Term))
39
40
41 cluster_1_plot <- ggplot(goEnrichment, aes(x=Term, y=-log10(clas
sicFisher))) +
42   stat_summary(geom = "bar", fun = mean, position = "dodge") +
43   xlab("Biological process") +
44   ylab("-log10(p-value)") +
45   ggtitle("GF expansions in Riftia") +
46   scale_y_continuous(limits=c(0,30), breaks=round(seq(0,30, by =
2), 1)) +

```

```

47 theme_classic() +
48 theme(
49   legend.position='none',
50   legend.background=element_rect(),
51   plot.title=element_text(angle=0, size=12, face="bold", vjust
=1),
52   axis.text.x=element_text(angle=0, size=10, hjust=1.10),
53   axis.text.y=element_text(angle=0, size=10, vjust=0.5),
54   axis.title=element_text(size=12),
55   legend.key=element_blank(),      #removes the border
56   legend.key.size=unit(1, "cm"),    #Sets overall area/size
of the legend
57   legend.text=element_text(size=12), #Text size
58   title=element_text(size=12)) +
59   guides(colour=guide_legend(override.aes=list(size=2.5))) +
60   coord_flip()
61
62 cluster_1_plot

```

- exported in pdf 4x8 inches

## Gains

```

1 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
grep -w Rpac | cut -f 1 > gene_families_originated_riftia.txt #f
amilies expanded in riftia
2 fgrep -f gene_families_originated_riftia.txt ../../Orthogroups.c
sv > gene_families_originated_riftia.csv
3 cut -f 1,24 gene_families_originated_riftia.csv | sed 's/Rpac|//
g' | cut -f 2 | sed 's/, /\n/g' > gene_IDs_originated_riftia.txt
4
5 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
grep -w Siboglinidae | cut -f 1 > gene_families_originated_rifti
a_siboglinidae.txt #families expanded in riftia

```

```

6 fgrep -f gene_families_originated_riftia_siboglinidae.txt ../../Orthogroups.csv > gene_families_originated_riftia_siboglinidae.csv
7 cut -f 1,24 gene_families_originated_riftia_siboglinidae.csv | sed 's/Rpac|//g' | cut -f 2 | sed 's/, /\n/g' | sed '/^$/d' > gene_IDs_originated_riftia_siboglinidae.txt
8
9 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv | grep -w Vestimentifera | cut -f 1 > gene_families_originated_riftia_Vestimentifera.txt #families expanded in riftia
10 fgrep -f gene_families_originated_riftia_Vestimentifera.txt ../../Orthogroups.csv > gene_families_originated_riftia_Vestimentifera.csv
11 cut -f 1,24 gene_families_originated_riftia_Vestimentifera.csv | sed 's/Rpac|//g' | cut -f 2 | sed 's/, /\n/g' | sed '/^$/d' > gene_IDs_originated_riftia_Vestimentifera.txt
12
13 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv | grep -w Vestimentifera_cl1 | cut -f 1 > gene_families_originated_riftia_Vestimentifera_cl1.txt #families expanded in riftia
14 fgrep -f gene_families_originated_riftia_Vestimentifera_cl1.txt ../../Orthogroups.csv > gene_families_originated_riftia_Vestimentifera_cl1.csv
15 cut -f 1,24 gene_families_originated_riftia_Vestimentifera_cl1.csv | sed 's/Rpac|//g' | cut -f 2 | sed 's/, /\n/g' | sed '/^$/d' > gene_IDs_originated_riftia_Vestimentifera_cl1.txt
16
17 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv | grep -w Vestimentifera_cl2 | cut -f 1 > gene_families_originated_riftia_Vestimentifera_cl2.txt #families expanded in riftia
18 fgrep -f gene_families_originated_riftia_Vestimentifera_cl2.txt ../../Orthogroups.csv > gene_families_originated_riftia_Vestimentifera_cl2.csv
19 cut -f 1,24 gene_families_originated_riftia_Vestimentifera_cl2.c

```

```
sv | sed 's/Rpac|//g' | cut -f 2 | sed 's/, /\n/g' | sed '/^$/d'  
> gene_IDs_originated_riftia_Vestimentifera_cl2.txt
```

## Oasisia

```
1 cd /data/scratch/btx654/gene_family_evolution/ferdi_script/Jul20  
21/GO_terms/oasisia  
2 # I will use BlastX GO terms  
3 cut -f2,13 ../../oasisia_annotation_Jan2021_TrinoPantherK0.xls |  
tail -n +2 > oasisia_GO_raw.txt # 38179 genes in riftia_GO_raw.t  
xt  
4 grep 'GO' oasisia_GO_raw.txt > oasisia_GO_only_raw.txt # 20737 r  
iftia_GO_only_raw.txt  
5
```

### python.py

```
1 if __name__ == "__main__":  
2  
3     import re  
4  
5     i = open("oasisia_GO_only_raw.txt", "r")  
6     o = open("oasisia_GO_universe.txt", "w")  
7  
8     regex = re.compile(r'GO:\d+')  
9  
10    for line in i:  
11        GOmatches = regex.findall(line)  
12        Gene_ID = line.split("\t",1)[0]  
13        if not GOmatches == []:  
14            o.write(Gene_ID+"\t")  
15            for i, match in enumerate(GOmatches):  
16                if i+1 == len(GOmatches):
```



```

17         o.write(match.strip('"')+"\n")
18     else:
19         o.write(match.strip('"')+", ")

```

```

1 module load python
2 python python.py

```

## Expansions

exp.sh

```

1 cut -f 1,7 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
  grep -w Oalv | cut -f 1 > gene_families_expanded_oasisia.txt #fa
  milies expanded in riftia
2 fgrep -f gene_families_expanded_oasisia.txt ../../Orthogroups.cs
  v > gene_families_expanded_oasisia.csv
3 cut -f 1,19 gene_families_expanded_oasisia.csv | sed 's/Oalv|//g
  ' | cut -f 2 | sed 's/, /\n/g' > gene_IDs_expanded_oasisia.txt

```

## Gains

gains.sh

```

1 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
  grep -w Oalv | cut -f 1 > gene_families_originated_oasisia.txt #
  families expanded in riftia
2 fgrep -f gene_families_originated_oasisia.txt ../../Orthogroups.
  csv > gene_families_originated_oasisia.csv
3 cut -f 1,19 gene_families_originated_oasisia.csv | sed 's/Oal
  v|//g' | cut -f 2 | sed 's/, /\n/g' > gene_IDs_originated_oasisi
  a.txt
4
5 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
  grep -w Siboglinidae | cut -f 1 > gene_families_originated_oasisi
  a_siboglinidae.txt #families expanded in riftia
6 fgrep -f gene_families_originated_oasisia_siboglinidae.txt
  ../../Orthogroups.csv > gene_families_originated_oasisia_sibogli
  nidae.csv

```

```

7 cut -f 1,19 gene_families_originated_oasisia_siboglinidae.csv |
  sed 's/0alv|//g' | cut -f 2 | sed 's/, /\n/g' | sed '/^$/d' > ge
  ne_IDs_originated_oasisia_siboglinidae.txt
8
9 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
  grep -w Vestimentifera | cut -f 1 > gene_families_originated_oas
  isia_Vestimentifera.txt #families expanded in riftia
10 fgrep -f gene_families_originated_oasisia_Vestimentifera.txt
  ../../Orthogroups.csv > gene_families_originated_oasisia_Vestime
  ntifera.csv
11 cut -f 1,19 gene_families_originated_oasisia_Vestimentifera.csv
  | sed 's/0alv|//g' | cut -f 2 | sed 's/, /\n/g' | sed '/^$/d' >
  gene_IDs_originated_oasisia_Vestimentifera.txt
12
13 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
  grep -w Vestimentifera_cl1 | cut -f 1 > gene_families_originated
  _oasisia_Vestimentifera_cl1.txt #families expanded in riftia
14 fgrep -f gene_families_originated_oasisia_Vestimentifera_cl1.txt
  ../../Orthogroups.csv > gene_families_originated_oasisia_Vestime
  ntifera_cl1.csv
15 cut -f 1,19 gene_families_originated_oasisia_Vestimentifera_cl1.
  csv | sed 's/0alv|//g' | cut -f 2 | sed 's/, /\n/g' | sed '/^$/d
  ' > gene_IDs_originated_oasisia_Vestimentifera_cl1.txt
16
17 cut -f 1,4 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
  grep -w Vestimentifera_cl2 | cut -f 1 > gene_families_originated
  _oasisia_Vestimentifera_cl2.txt #families expanded in riftia
18 fgrep -f gene_families_originated_oasisia_Vestimentifera_cl2.txt
  ../../Orthogroups.csv > gene_families_originated_oasisia_Vestime
  ntifera_cl2.csv
19 cut -f 1,19 gene_families_originated_oasisia_Vestimentifera_cl2.
  csv | sed 's/0alv|//g' | cut -f 2 | sed 's/, /\n/g' | sed '/^$/d
  ' > gene_IDs_originated_oasisia_Vestimentifera_cl2.txt

```

## Paraescarpia

```
1 cd /data/scratch/btx654/gene_family_evolution/ferdi_script/Jul20
2 21/GO_terms/paraescarpia
3 # I will use BlastX GO terms
4 cut -f2,13 ../../paraescarpia_annotation_Jun2021_TrinoPantherK0.
5 xls | tail -n +2 > paraescarpia_GO_raw.txt # 38179 genes in rift
   ia_GO_raw.txt
6
7 grep 'GO' paraescarpia_GO_raw.txt > paraescarpia_GO_only_raw.txt
8 # 20737 riftia_GO_only_raw.txt
9
```

### python.py

```
1 if __name__ == "__main__":
2
3     import re
4
5     i = open("paraescarpia_GO_only_raw.txt", "r")
6     o = open("paraescarpia_GO_universe.txt", "w")
7
8     regex = re.compile(r'GO:\d+')
9
10    for line in i:
11        GOmatches = regex.findall(line)
12        Gene_ID = line.split("\t",1)[0]
13        if not GOmatches == []:
14            o.write(Gene_ID+"\t")
15            for i, match in enumerate(GOmatches):
16                if i+1 == len(GOmatches):
17                    o.write(match.strip('"')+"\n")
18                else:
19                    o.write(match.strip('"')+", ")
```

```
1 module load python
2 python python.py
```

## Expansions

```
1 cp ../oasisia/*.sh ./
2 sed -i 's/oasisia/paraescarpia/g' *.sh
3 sed -i 's/Oa1v/Pech/g' *.sh
4 sed -i 's/19/23/g' *.sh
5
6 bash exp.sh
```

## Gains

```
bash gains.sh
```

## Lamellibrachia

```
1 cd /data/scratch/btx654/gene_family_evolution/ferdi_script/Jul20
  21/GO_terms/lamellibrachia
2 # I will use BlastX GO terms
3 cut -f2,13 ../../lamellibrachia_annotation_Feb2021_TrinoPantherK
  0_OK.xls | tail -n +2 > lamellibrachia_GO_raw.txt # 38179 genes
  in riftia_GO_raw.txt
4 grep 'GO' lamellibrachia_GO_raw.txt > lamellibrachia_GO_only_ra
  w.txt # 20737 riftia_GO_only_raw.txt
5
```

## python.py

```
1 if __name__ == "__main__":
2
3     import re
4
5     i = open("lamellibrachia_GO_only_raw.txt", "r")
```

```

6     o = open("lamellibrachia_G0_universe.txt", "w")
7
8     regex = re.compile(r'G0:\d+')
9
10    for line in i:
11        G0matches = regex.findall(line)
12        Gene_ID = line.split("\t",1)[0]
13        if not G0matches == []:
14            o.write(Gene_ID+"\t")
15            for i, match in enumerate(G0matches):
16                if i+1 == len(G0matches):
17                    o.write(match.strip("'")+"\n")
18                else:
19                    o.write(match.strip("'")+", ")

```

```

1 module load python
2 python python.py

```

## Expansions

```

1 cp ../oasisia/*.sh ./
2 sed -i 's/oasisia/lamellibrachia/g' *.sh
3 sed -i 's/Oalv/Lluy/g' *.sh
4 sed -i 's/19/14/g' *.sh
5
6 bash exp.sh

```

## Gains

```
bash gains.sh
```

## Osedax

```
1 cd /data/scratch/btx654/gene_family_evolution/ferdi_script/Jul20
2 21/GO_terms/0sedax
3 # I will use BlastX GO terms
4 cut -f2,13 ../../osedax_annotation_Jan2021_TrinoPantherK0.xls |
5 tail -n +2 > osedax_GO_raw.txt #
6 grep 'GO' osedax_GO_raw.txt > osedax_GO_only_raw.txt # 10674 ose
7 dax_GO_only_raw.txt
```

python.py

```
1 if __name__ == "__main__":
2
3     import re
4
5     i = open("osedax_GO_only_raw.txt", "r")
6     o = open("osedax_GO_universe.txt", "w")
7
8     regex = re.compile(r'GO:\d+')
9
10    for line in i:
11        GOMatches = regex.findall(line)
12        Gene_ID = line.split("\t",1)[0]
13        if not GOMatches == []:
14            o.write(Gene_ID+"\t")
15            for i, match in enumerate(GOMatches):
16                if i+1 == len(GOMatches):
17                    o.write(match.strip('"')+"\n")
18                else:
19                    o.write(match.strip('"')+", ")
```

```
1 module load python
```

```
2 python python.py
```

## Expansions

```
1 cp ../oasisia/*.sh ./
2 sed -i 's/oasisia/osedax/g' *.sh
3 sed -i 's/Oalv/Ofra/g' *.sh
4 sed -i 's/19/20/g' *.sh
5
6 bash exp.sh
```

## Gains

```
bash gains.sh
```

## Losses

using Owenia to annotate the losses

```
1 cd /data/scratch/btx654/gene_family_evolution/ferdi_script/Jul20
  21/GO_terms/losses
2 cut -f2,11 ../../Owenia_annotation_v250920.1_TrinoPantherK0.xls
  | tail -n +2 | sed 's/"//g' > owenia_GO_raw.txt #
3 grep 'GO' owenia_GO_raw.txt > owenia_GO_only_raw.txt # 21108 ose
  dax_GO_only_raw.txt
```

## python.py

```
1 if __name__ == "__main__":
2
3     import re
4
5     i = open("owenia_GO_only_raw.txt", "r")
6     o = open("owenia_GO_universe.txt", "w")
7
8     regex = re.compile(r'GO:\d+')

```

```

9
10     for line in i:
11         G0matches = regex.findall(line)
12         Gene_ID = line.split("\t",1)[0]
13         if not G0matches == []:
14             o.write(Gene_ID+"\t")
15             for i, match in enumerate(G0matches):
16                 if i+1 == len(G0matches):
17                     o.write(match.strip("'")+"\n")
18                 else:
19                     o.write(match.strip("'")+", ")

```

```

1 module load python
2 python python.py

```

#### losses.sh

```

1 cut -f 1,21 ../../Orthogroups.csv | tail -n +2 | grep -w Ofus |
  cut -f 1 > owenia_all_GF
2
3 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
  grep -w Ofra | cut -f 1 > gene_families_lost_osedax.txt
4 grep -f gene_families_lost_osedax.txt owenia_all_GF > losses_ose
  dax_owenia
5 fgrep -f losses_osedax_owenia ../../Orthogroups.csv > gene_famil
  ies_lost_osedax_owenia.csv
6 cut -f 1,21 gene_families_lost_osedax_owenia.csv | sed 's/Ofu
  s|//g' | cut -f 2 | sed 's/, /\n/g' > gene_IDs_lost_osedax_oweni
  a.txt
7
8 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
  grep -w Oalv | cut -f 1 > gene_families_lost_oasisia.txt

```



```
9  grep -f gene_families_lost_oasisia.txt owenia_all_GF > losses_oa
sisia_owenia
10 fgrep -f losses_oasisia_owenia ../../Orthogroups.csv > gene_fami
lies_lost_oasisia_owenia.csv
11 cut -f 1,21 gene_families_lost_oasisia_owenia.csv | sed 's/Ofu
s|//g' | cut -f 2 | sed 's/, /\n/g' > gene_IDs_lost_oasisia_owen
ia.txt
12
13 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
grep -w Rpac | cut -f 1 > gene_families_lost_riftia.txt
14 grep -f gene_families_lost_riftia.txt owenia_all_GF > losses_rif
tia_owenia
15 fgrep -f losses_riftia_owenia ../../Orthogroups.csv > gene_famil
ies_lost_riftia_owenia.csv
16 cut -f 1,21 gene_families_lost_riftia_owenia.csv | sed 's/Ofu
s|//g' | cut -f 2 | sed 's/, /\n/g' > gene_IDs_lost_riftia_oweni
a.txt
17
18 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
grep -w Pech | cut -f 1 > gene_families_lost_paraescarpia.txt
19 grep -f gene_families_lost_paraescarpia.txt owenia_all_GF > loss
es_paraescarpia_owenia
20 fgrep -f losses_paraescarpia_owenia ../../Orthogroups.csv > gene
_families_lost_paraescarpia_owenia.csv
21 cut -f 1,21 gene_families_lost_paraescarpia_owenia.csv | sed 's/
Ofus|//g' | cut -f 2 | sed 's/, /\n/g' > gene_IDs_lost_paraescar
pia_owenia.txt
22
23 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
grep -w Lluy | cut -f 1 > gene_families_lost_lamellibrachia.txt
24 grep -f gene_families_lost_lamellibrachia.txt owenia_all_GF > lo
sses_lamellibrachia_owenia
```

```
25 fgrep -f losses_lamellibrachia_owenia ../../Orthogroups.csv > ge
ne_families_lost_lamellibrachia_owenia.csv
26 cut -f 1,21 gene_families_lost_lamellibrachia_owenia.csv | sed '
s/Ofus|//g' | cut -f 2 | sed 's/, /\n/g' > gene_IDs_lost_lamelli
brachia_owenia.txt
27
28 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
grep -w Siboglinidae | cut -f 1 > gene_families_lost_Siboglinida
e.txt
29 grep -f gene_families_lost_Siboglinidae.txt owenia_all_GF > loss
es_Siboglinidae_owenia
30 fgrep -f losses_Siboglinidae_owenia ../../Orthogroups.csv > gene
_families_lost_Siboglinidae_owenia.csv
31 cut -f 1,21 gene_families_lost_Siboglinidae_owenia.csv | sed 's/
Ofus|//g' | cut -f 2 | sed 's/, /\n/g' > gene_IDs_lost_Siboglini
dae_owenia.txt
32
33 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
grep -w Vestimentifera | cut -f 1 > gene_families_lost_Vestiment
ifera.txt
34 grep -f gene_families_lost_Vestimentifera.txt owenia_all_GF > lo
sses_Vestimentifera_owenia
35 fgrep -f losses_Vestimentifera_owenia ../../Orthogroups.csv > ge
ne_families_lost_Vestimentifera_owenia.csv
36 cut -f 1,21 gene_families_lost_Vestimentifera_owenia.csv | sed '
s/Ofus|//g' | cut -f 2 | sed 's/, /\n/g' > gene_IDs_lost_Vestime
ntifera_owenia.txt
37
38 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
grep -w Vestimentifera_cl1 | cut -f 1 > gene_families_lost_Vesti
mentifera_cl1.txt
39 grep -f gene_families_lost_Vestimentifera_cl1.txt owenia_all_GF
> losses_Vestimentifera_cl1_owenia
40 fgrep -f losses_Vestimentifera_cl1_owenia ../../Orthogroups.csv
```

```

> gene_families_lost_Vestimentifera_cl1_owenia.csv
41 cut -f 1,21 gene_families_lost_Vestimentifera_cl1_owenia.csv | s
   ed 's/0fus|//g' | cut -f 2 | sed 's/, /\n/g' > gene_IDs_lost_Ves
   timentifera_cl1_owenia.txt
42
43 cut -f 1,6 ../../orthofinder_ultrasensitive_stats_Jun2021.tsv |
   grep -w Vestimentifera_cl2 | cut -f 1 > gene_families_lost_Vesti
   mentifera_cl2.txt
44 grep -f gene_families_lost_Vestimentifera_cl2.txt owenia_all_GF
   > losses_Vestimentifera_cl2_owenia
45 fgrep -f losses_Vestimentifera_cl2_owenia ../../Orthogroups.csv
   > gene_families_lost_Vestimentifera_cl2_owenia.csv
46 cut -f 1,21 gene_families_lost_Vestimentifera_cl2_owenia.csv | s
   ed 's/0fus|//g' | cut -f 2 | sed 's/, /\n/g' > gene_IDs_lost_Ves
   timentifera_cl2_owenia.txt

```

## Results

```

1 scp -i ~/.ssh/id_rsa_apocrita -r btx654@login.hpc.qmul.ac.uk:/da
   ta/scratch/btx654/gene_family_evolution/ferdi_script/Jul2021/GO_
   terms/*/*_GO_universe.txt /Users/giacomo/Desktop/R/GO_enrichment
   /New_aug21/
2 scp -i ~/.ssh/id_rsa_apocrita -r btx654@login.hpc.qmul.ac.uk:/da
   ta/scratch/btx654/gene_family_evolution/ferdi_script/Jul2021/GO_
   terms/*/*_gene_IDs_* /Users/giacomo/Desktop/R/GO_enrichment/New_au
   g21/

```

### Riftia

Expansions

MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher

<b>1</b>	GO:0005509	calcium ion binding	1315	44	4.36	1e-30
<b>2</b>	GO:0046872	metal ion binding	5198	47	17.25	5.7e-17
<b>3</b>	GO:0043169	cation binding	5286	47	17.54	1.2e-16
<b>4</b>	GO:0009374	biotin binding	99	9	0.33	3.3e-11
<b>5</b>	GO:0033293	monocarboxylic acid binding	117	9	0.39	1.5e-10
<b>6</b>	GO:0043167	ion binding	7525	47	24.97	4.4e-10
<b>7</b>	GO:0019842	vitamin binding	248	9	0.82	1.1e-07
<b>8</b>	GO:0031406	carboxylic acid binding	328	9	1.09	1.2e-06
<b>9</b>	GO:0043177	organic acid binding	328	9	1.09	1.2e-06

<b>10</b>	GO:1901681	sulfur compound binding	333	9	1.10	1.3e-06
<b>11</b>	GO:0033218	amide binding	367	9	1.22	3.0e-06
<b>12</b>	GO:0005488	binding	12748	52	42.30	0.00023
<b>13</b>	GO:0004063	aryldialkylphosphatase activity	3	1	0.01	0.00992
<b>14</b>	GO:0043878	glyceraldehyde-3-phosphate dehydrogenase...	6	1	0.02	0.01975
<b>15</b>	GO:0097603	temperature-gated ion channel activity	7	1	0.02	0.02300

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0007160	cell-matrix adhesion	218	32	0.62	1e-30
<b>2</b>	GO:0031589	cell-substrate adhesion	356	32	1.01	1e-30
<b>3</b>	GO:0007155	cell adhesion	1518	34	4.31	2.5e-25
<b>4</b>	GO:0022610	biological adhesion	1571	34	4.46	7.7e-25
<b>5</b>	GO:0022001	negative regulation of anterior neural c...	1	1	0.00	0.0028
<b>6</b>	GO:0022002	negative regulation of anterior neural c...	1	1	0.00	0.0028
<b>7</b>	GO:0060898	eye field cell fate commitment involved ...	1	1	0.00	0.0028
<b>8</b>	GO:0046619	optic placode formation	2	1	0.01	0.0057

		involved in came...				
<b>9</b>	GO:0070 654	sensory epitheliu m regenerat ion	2	1	0.01	0.0057
<b>10</b>	GO:0001 743	optic placode formation	3	1	0.01	0.0085
<b>11</b>	GO:0002 765	immune response - inhibiting signal transd...	3	1	0.01	0.0085
<b>12</b>	GO:0002 767	immune response - inhibiting cell surface ...	3	1	0.01	0.0085
<b>13</b>	GO:0002 769	natural killer cell inhibitory signaling. ..	3	1	0.01	0.0085
<b>14</b>	GO:0043 049	otic placode formation	3	1	0.01	0.0085
<b>15</b>	GO:1990 399	epitheliu m regenerat	3	1	0.01	0.0085

		ion				
--	--	-----	--	--	--	--

Gains  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

1	GO:0090729	toxin activity	185	64	6.17	1e-30
2	GO:0008191	metalloendopeptidase inhibitor activity	128	47	4.27	1e-30
3	GO:0004866	endopeptidase inhibitor activity	288	51	9.61	5.4e-23
4	GO:0061135	endopeptidase regulator activity	296	51	9.88	2.0e-22
5	GO:0030414	peptidase inhibitor activity	297	51	9.91	2.3e-22



<b>6</b>	GO:0061134	peptidase regulator activity	317	51	10.58	4.7e-21
<b>7</b>	GO:0004857	enzyme inhibitor activity	400	51	13.35	1.3e-16
<b>8</b>	GO:0005507	copper ion binding	301	40	10.04	7.4e-14
<b>9</b>	GO:0003711	transcript ion elongation regulator activ...	74	20	2.47	2.3e-13
<b>10</b>	GO:0050733	RS domain binding	75	19	2.50	3.2e-12
<b>11</b>	GO:0098772	molecular function regulator	2429	130	81.05	1.3e-08
<b>12</b>	GO:0005179	hormone activity	65	14	2.17	2.3e-08
<b>13</b>	GO:1901681	sulfur compound binding	333	32	11.11	8.3e-08
<b>14</b>	GO:0045159	myosin II binding	34	10	1.13	1.0e-07

<b>15</b>	GO:0005509	calcium ion binding	1315	75	43.88	2.9e-06
-----------	------------	---------------------	------	----	-------	---------

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0042311	vasodilation	145	47	4.88	1e-30
<b>2</b>	GO:0097755	positive regulation of blood vessel diam...	157	48	5.28	1e-30
<b>3</b>	GO:0008015	blood circulation	674	91	22.67	1e-30
<b>4</b>	GO:0003013	circulatory system process	780	97	26.24	5.9e-30
<b>5</b>	GO:0097746	regulation of blood vessel diameter	221	50	7.43	1.8e-27

<b>6</b>	GO:0008217	regulation of blood pressure	203	48	6.83	2.5e-27
<b>7</b>	GO:0035296	regulation of tube diameter	223	50	7.50	2.8e-27
<b>8</b>	GO:0035150	regulation of tube size	249	51	8.38	7.6e-26
<b>9</b>	GO:0003018	vascular process in circulatory system	290	53	9.75	2.5e-24
<b>10</b>	GO:0090066	regulation of anatomical structure size	533	63	17.93	2.1e-18
<b>11</b>	GO:0003008	system process	1872	132	62.97	3.8e-17
<b>12</b>	GO:0043484	regulation of RNA splicing	245	36	8.24	8.2e-14
<b>13</b>	GO:0050	regulation	512	54	17.22	8.2e-14

	878	n of body fluid levels				
<b>14</b>	GO:0050817	coagulation	418	48	14.06	9.1e-14
<b>15</b>	GO:0007596	blood coagulation	396	46	13.32	2.0e-13

Gains siboglinidae  
MF

<b>GO.ID</b>	<b>Term</b>	<b>Annotated</b>	<b>Significant</b>	<b>Expected</b>	<b>classicFisher</b>
--------------	-------------	------------------	--------------------	-----------------	----------------------

<b>1</b>	GO:0038023	signaling receptor activity	606	23	5.42	1.5e-09
<b>2</b>	GO:0060089	molecular transducer activity	606	23	5.42	1.5e-09
<b>3</b>	GO:0004888	transmembrane signaling receptor activit...	492	19	4.40	4.2e-08

<b>4</b>	GO:0005344	oxygen carrier activity	35	6	0.31	5.6e-07
<b>5</b>	GO:0005215	transporter activity	978	25	8.75	6.8e-07
<b>6</b>	GO:0005509	calcium ion binding	533	18	4.77	7.6e-07
<b>7</b>	GO:0005216	ion channel activity	331	14	2.96	1.1e-06
<b>8</b>	GO:0022836	gated channel activity	263	12	2.35	3.3e-06
<b>9</b>	GO:0015267	channel activity	367	14	3.28	3.8e-06
<b>10</b>	GO:0030594	neurotransmitter receptor activity	105	8	0.94	4.0e-06
<b>11</b>	GO:0022803	passive transmembrane transporter activity	369	14	3.30	4.1e-06
<b>12</b>	GO:0015464	acetylcholine receptor activity	17	4	0.15	1.3e-05

<b>13</b>	GO:0008227	G protein-coupled amine receptor activit...	35	5	0.31	1.3e-05
<b>14</b>	GO:0004937	alpha1-adrenergic receptor activity	6	3	0.05	1.4e-05
<b>15</b>	GO:0004252	serine-type endopeptidase activity	61	6	0.55	1.6e-05

BP

<b>GO.ID</b>	<b>Term</b>	<b>Annotated</b>	<b>Significant</b>	<b>Expected</b>	<b>classicFisher</b>
--------------	-------------	------------------	--------------------	-----------------	----------------------

<b>1</b>	GO:0051716	cellular response to stimulus	5550	51	25.26	1.9e-09
<b>2</b>	GO:0007165	signal transduction	4236	44	19.28	1.9e-09
<b>3</b>	GO:0023	signaling	4621	46	21.03	2.3e-09

	052					
4	GO:0007154	cell communication	4673	46	21.27	3.5e-09
5	GO:0050896	response to stimulus	7282	56	33.15	1.2e-07
6	GO:0002224	toll-like receptor signaling pathway	102	7	0.46	4.0e-07
7	GO:0009605	response to external stimulus	2081	25	9.47	2.9e-06
8	GO:0019932	second-messenger-mediated signaling	437	11	1.99	4.3e-06
9	GO:0002221	pattern recognition receptor signaling p...	146	7	0.66	4.5e-06
10	GO:0003146	heart jogging	10	3	0.05	1.1e-05

<b>11</b>	GO:0044057	regulation of system process	359	9	1.63	3.6e-05
<b>12</b>	GO:0009887	animal organ morphogenesis	1029	15	4.68	5.2e-05
<b>13</b>	GO:0050794	regulation of cellular process	8584	56	39.07	6.7e-05
<b>14</b>	GO:0050771	negative regulation of axonogenesis	53	4	0.24	9.8e-05
<b>15</b>	GO:0019935	cyclic-nucleotide-mediated signaling	236	7	1.07	9.9e-05

Gains vestimentifera  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

--	--	--	--	--	--	--



1	GO:0005509	calcium ion binding	1315	108	50.18	1.2e-14
2	GO:0038023	signaling receptor activity	1177	88	44.91	6.8e-10
3	GO:0060089	molecular transducer activity	1177	88	44.91	6.8e-10
4	GO:0046872	metal ion binding	5198	264	198.35	1.1e-08
5	GO:0043169	cation binding	5286	267	201.71	1.5e-08
6	GO:0043167	ion binding	7525	354	287.15	2.7e-08
7	GO:0022848	acetylcholine-gated cation-selective cha...	89	17	3.40	3.6e-08
8	GO:0099529	neurotransmitter receptor activity invol...	126	20	4.81	6.1e-08

<b>9</b>	GO:0005261	cation channel activity	515	46	19.65	7.7e-08
<b>10</b>	GO:1904315	transmitter-gated ion channel activity i...	118	19	4.50	1.0e-07
<b>11</b>	GO:0005216	ion channel activity	651	53	24.84	1.6e-07
<b>12</b>	GO:0030594	neurotransmitter receptor activity	199	25	7.59	1.7e-07
<b>13</b>	GO:0003700	DNA-binding transcription factor activit...	883	65	33.69	2.6e-07
<b>14</b>	GO:0004888	transmembrane signaling receptor activit...	926	67	35.34	3.1e-07
<b>15</b>	GO:0005230	extracellular ligand-gated ion	165	22	6.30	3.2e-07

		channel a...				
--	--	-----------------	--	--	--	--

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

1	GO:0007165	signal transduction	4236	218	144.99	5.8e-12
2	GO:0001867	complement activation, lectin pathway	83	19	2.84	3.5e-11
3	GO:0006956	complement activation	93	20	3.18	3.6e-11
4	GO:0050896	response to stimulus	7282	326	249.25	4.8e-11
5	GO:0006959	humoral immune response	163	26	5.58	5.5e-11
6	GO:0023052	signaling	4621	226	158.17	3.3e-10

<b>7</b>	GO:0007154	cell communication	4673	227	159.95	5.6e-10
<b>8</b>	GO:0098542	defense response to other organism	805	63	27.55	6.9e-10
<b>9</b>	GO:0045087	innate immune response	628	53	21.50	1.3e-09
<b>10</b>	GO:0009605	response to external stimulus	2081	121	71.23	2.0e-09
<b>11</b>	GO:0007155	cell adhesion	1518	95	51.96	4.5e-09
<b>12</b>	GO:0002252	immune effector process	525	46	17.97	5.2e-09
<b>13</b>	GO:0022610	biological adhesion	1571	97	53.77	6.0e-09
<b>14</b>	GO:0043207	response to external biotic stimulus	1022	70	34.98	2.0e-08
<b>15</b>	GO:0051707	response to other organism	1022	70	34.98	2.0e-08

Gains vestimentifera\_cl1

MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

1	GO:0008467	[heparan sulfate]-glucosamine 3-sulfotransferase activity	22	10	0.12	9.8e-18
2	GO:0034483	heparan sulfate sulfotransferase activity	30	10	0.17	4.4e-16
3	GO:0050656	3'-phosphoadenosine 5'-phosphosulfate binding	16	8	0.09	7.9e-15
4	GO:1901681	sulfur compound binding	333	17	1.84	3.1e-12
5	GO:0008146	sulfotransferase activity	91	10	0.50	7.1e-11
6	GO:0016016	transferase activity	126	10	0.70	1.8e-09

	782	se activity, transferri ng sulfu...				
<b>7</b>	GO:0008 810	cellulase activity	6	3	0.03	3.2e-06
<b>8</b>	GO:0008 028	monocar boxylic acid transme mbrane transp...	97	6	0.54	1.6e-05
<b>9</b>	GO:0015 129	lactate transme mbrane transport er activi...	16	3	0.09	8.7e-05
<b>10</b>	GO:0009 374	biotin binding	99	5	0.55	0.00022
<b>11</b>	GO:0036 094	small molecule binding	3178	32	17.58	0.00027
<b>12</b>	GO:0005 342	organic acid transme mbrane transport er a...	228	7	1.26	0.00027
<b>13</b>	GO:0046 943	carboxyli c acid	228	7	1.26	0.00027

		transmembrane transport e...				
<b>14</b>	GO:0070403	NAD+ binding	59	4	0.33	0.00032
<b>15</b>	GO:1990404	protein ADP-ribosylase activity	60	4	0.33	0.00034

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0015015	heparan sulfate proteoglycan biosynthetic...	21	10	0.09	2.6e-19
<b>2</b>	GO:0006477	protein sulfation	27	10	0.11	6.0e-18
<b>3</b>	GO:0051923	sulfation	29	10	0.12	1.4e-17
<b>4</b>	GO:0046	regulation	37	10	0.15	2.4e-16

	596	n of viral entry into host cell				
5	GO:0052372	modulation by symbiont of entry into hos...	41	10	0.17	7.6e-16
6	GO:0015012	heparan sulfate proteoglycan biosynthe ti...	43	10	0.18	1.3e-15
7	GO:0050818	regulation of coagulation	69	11	0.29	4.0e-15
8	GO:0030201	heparan sulfate proteoglycan metabolic p...	48	10	0.20	4.4e-15
9	GO:0050819	negative regulation of coagulation	57	10	0.24	2.8e-14
10	GO:0046718	viral entry into host cell	58	10	0.24	3.4e-14



<b>11</b>	GO:0044409	entry into host	63	10	0.26	8.1e-14
<b>12</b>	GO:0052126	movement in host environment	72	10	0.30	3.3e-13
<b>13</b>	GO:0030166	proteoglycan biosynthetic process	92	10	0.38	4.2e-12
<b>14</b>	GO:0051701	interaction with host	146	11	0.60	1.9e-11
<b>15</b>	GO:1903900	regulation of viral life cycle	115	10	0.48	4.0e-11

Gains vestimentifera\_cl2  
MF

<b>GO.ID</b>	<b>Term</b>	<b>Annotated</b>	<b>Significant</b>	<b>Expected</b>	<b>classicFisher</b>
--------------	-------------	------------------	--------------------	-----------------	----------------------

<b>1</b>	GO:0005	calcium ion	1315	19	5.17	3.9e-07
----------	---------	-------------	------	----	------	---------

	509	binding				
<b>2</b>	GO:0004252	serine-type endopeptidase activity	119	7	0.47	4.1e-07
<b>3</b>	GO:0008236	serine-type peptidase activity	152	7	0.60	2.1e-06
<b>4</b>	GO:0034185	apolipoprotein binding	56	5	0.22	2.6e-06
<b>5</b>	GO:0042806	fucose binding	56	5	0.22	2.6e-06
<b>6</b>	GO:0017171	serine hydrolase activity	159	7	0.63	2.9e-06
<b>7</b>	GO:0048029	monosaccharide binding	118	6	0.46	6.8e-06
<b>8</b>	GO:0004175	endopeptidase activity	453	10	1.78	1.0e-05
<b>9</b>	GO:1990405	protein antigen binding	15	3	0.06	2.6e-05
<b>10</b>	GO:0003	antigen binding	49	4	0.19	4.0e-05

	823					
<b>11</b>	GO:0008233	peptidase activity	721	11	2.84	0.00010
<b>12</b>	GO:0038023	signaling receptor activity	1177	14	4.63	0.00015
<b>13</b>	GO:0060089	molecular transducer activity	1177	14	4.63	0.00015
<b>14</b>	GO:0004952	dopamine neurotransmitter receptor activ...	10	2	0.04	0.00067
<b>15</b>	GO:0043169	cation binding	5286	32	20.79	0.00267

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

1	GO:0006956	complement	93	7	0.30	1.7e-08

		activation				
2	GO:0030449	regulation of complement activation	54	6	0.17	1.8e-08
3	GO:0006959	humoral immune response	163	8	0.52	4.5e-08
4	GO:0002920	regulation of humoral immune response	70	6	0.22	8.9e-08
5	GO:0051918	negative regulation of fibrinolysis	14	4	0.04	9.1e-08
6	GO:0051917	regulation of fibrinolysis	17	4	0.05	2.1e-07
7	GO:0001867	complement activation, lectin pathway	83	6	0.26	2.5e-07
8	GO:0001	regulation	46	5	0.15	3.4e-07

	868	n of complem ent activation , lec...				
<b>9</b>	GO:0010 185	regulatio n of cellular defense response	46	5	0.15	3.4e-07
<b>10</b>	GO:0030 194	positive regulatio n of blood coagulati on	20	4	0.06	4.3e-07
<b>11</b>	GO:1900 048	positive regulatio n of hemostas is	20	4	0.06	4.3e-07
<b>12</b>	GO:0050 820	positive regulatio n of coagulati on	21	4	0.07	5.3e-07
<b>13</b>	GO:0002 252	immune effector process	525	11	1.68	6.6e-07
<b>14</b>	GO:1903 036	positive regulatio n of response	56	5	0.18	9.3e-07

		to wound...				
<b>15</b>	GO:0006968	cellular defense response	58	5	0.19	1.1e-06

**Oasisia**  
Expansions  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0046982	protein heterodimerization activity	337	103	10.48	1e-30
<b>2</b>	GO:0008233	peptidase activity	951	139	29.57	1e-30
<b>3</b>	GO:0046872	metal ion binding	5800	339	180.34	1e-30
<b>4</b>	GO:0043169	cation binding	5894	339	183.26	1e-30
<b>5</b>	GO:0005509	calcium ion binding	1702	167	52.92	1e-30

<b>6</b>	GO:0005229	intracellular calcium activated chloride...	62	39	1.93	1e-30
<b>7</b>	GO:0061778	intracellular chloride channel activity	62	39	1.93	1e-30
<b>8</b>	GO:0046983	protein dimerization activity	922	118	28.67	1e-30
<b>9</b>	GO:0061578	Lys63-specific deubiquitinase activity	36	30	1.12	1e-30
<b>10</b>	GO:0022839	ion gated channel activity	80	39	2.49	1e-30
<b>11</b>	GO:0008237	metalloproteinase activity	420	75	13.06	1e-30
<b>12</b>	GO:1990380	Lys48-specific deubiquitinase activity	45	30	1.40	1e-30

<b>13</b>	GO:0070064	proline-rich region binding	47	30	1.46	1e-30
<b>14</b>	GO:0005488	binding	13397	510	416.56	1e-30
<b>15</b>	GO:0005254	chloride channel activity	112	39	3.48	1.1e-30

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0007155	cell adhesion	1817	170	41.12	1e-30
<b>2</b>	GO:0022610	biological adhesion	1876	170	42.46	1e-30
<b>3</b>	GO:1903753	negative regulation of p38MAPK cascade	32	30	0.72	1e-30
<b>4</b>	GO:0050857	positive regulation of	33	30	0.75	1e-30



		antigen receptor- ..				
<b>5</b>	GO:0050862	positive regulation of T cell receptor s...	33	30	0.75	1e-30
<b>6</b>	GO:1990108	protein linear deubiquit ination	34	30	0.77	1e-30
<b>7</b>	GO:0097343	rioptosome assembly	35	30	0.79	1e-30
<b>8</b>	GO:1901026	rioptosome assembly involved in necropt...	35	30	0.79	1e-30
<b>9</b>	GO:0045577	regulation of B cell differenti ation	37	30	0.84	1e-30
<b>10</b>	GO:0002360	T cell lineage commitment	38	30	0.86	1e-30

<b>11</b>	GO:0043369	CD4-positive or CD8-positive, alpha-beta...	38	30	0.86	1e-30
<b>12</b>	GO:2001238	positive regulation of extrinsic apoptot...	61	35	1.38	1e-30
<b>13</b>	GO:0045581	negative regulation of T cell differenti.. ..	39	29	0.88	1e-30
<b>14</b>	GO:1901223	negative regulation of NIK/NF-kappaB sig...	85	38	1.92	1e-30
<b>15</b>	GO:0050856	regulation of T cell receptor signaling ...	45	30	1.02	1e-30

Gains  
MF

GO.ID	Term	Annotated	Significant	Expected	
-------	------	-----------	-------------	----------	--

						<b>classicFisher</b>
--	--	--	--	--	--	----------------------

<b>1</b>	GO:0004745	retinol dehydrogenase activity	39	34	2.12	1e-30
<b>2</b>	GO:0005509	calcium ion binding	1702	199	92.44	8.6e-27
<b>3</b>	GO:0098631	cell adhesion mediator activity	124	41	6.73	8.7e-22
<b>4</b>	GO:0009374	biotin binding	205	44	11.13	2.5e-15
<b>5</b>	GO:0033293	monocarboxylic acid binding	230	44	12.49	1.9e-13
<b>6</b>	GO:0008061	chitin binding	109	28	5.92	3.0e-12
<b>7</b>	GO:0004866	endopeptidase inhibitor activity	247	42	13.41	3.7e-11
<b>8</b>	GO:0061	endopept	251	42	13.63	6.3e-11

	135	idase regulator activity				
<b>9</b>	GO:0030 414	peptidase inhibitor activity	263	43	14.28	8.0e-11
<b>10</b>	GO:0008 146	sulfotran sferase activity	108	26	5.87	8.8e-11
<b>11</b>	GO:1901 681	sulfur compoun d binding	419	57	22.76	1.3e-10
<b>12</b>	GO:0030 020	extracellu lar matrix structural constitu...	22	12	1.19	2.4e-10
<b>13</b>	GO:0016 616	oxidored uctase activity, acting on the C...	184	34	9.99	2.8e-10
<b>14</b>	GO:0042 806	fucose binding	123	27	6.68	3.6e-10
<b>15</b>	GO:0004 252	serine- type endopept idase activity	205	36	11.13	3.7e-10

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

1	GO:0007094	mitotic spindle assembly checkpoint	165	76	8.22	1e-30
2	GO:0031577	spindle checkpoint	165	76	8.22	1e-30
3	GO:0045841	negative regulation of mitotic metaphase...	165	76	8.22	1e-30
4	GO:0071173	spindle assembly checkpoint	165	76	8.22	1e-30
5	GO:0071174	mitotic spindle checkpoint	165	76	8.22	1e-30
6	GO:1902100	negative regulation of metaphase	165	76	8.22	1e-30

		se/anaph as...				
<b>7</b>	GO:1905 819	negative regulatio n of chromos ome separa...	165	76	8.22	1e-30
<b>8</b>	GO:2000 816	negative regulatio n of mitotic sister ch...	165	76	8.22	1e-30
<b>9</b>	GO:0033 048	negative regulatio n of mitotic sister ch...	170	76	8.47	1e-30
<b>10</b>	GO:0033 046	negative regulatio n of sister chromati d ...	172	76	8.57	1e-30
<b>11</b>	GO:0051 985	negative regulatio n of chromos ome segreg...	172	76	8.57	1e-30
<b>12</b>	GO:0045	negative regulatio	178	76	8.87	1e-30

	839	n of mitotic nuclear d...				
<b>13</b>	GO:0007091	metaphase/anaphase transition of mitotic...	220	83	10.96	1e-30
<b>14</b>	GO:0010965	regulation of mitotic sister chromatids...	220	83	10.96	1e-30
<b>15</b>	GO:0044784	metaphase/anaphase transition of cell cycle...	220	83	10.96	1e-30

Gains siboglinidae  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0003953	NAD+ nucleosidase activity	79	17	0.55	3.5e-21
<b>2</b>	GO:0050135	NAD(P)+ nucleosidase activity	79	17	0.55	3.5e-21
<b>3</b>	GO:0061809	NAD+ nucleotidase, cyclic ADP-ribose gen...	79	17	0.55	3.5e-21
<b>4</b>	GO:0016799	hydrolase activity, hydrolyzing N-glycos. ..	106	17	0.74	7.2e-19
<b>5</b>	GO:0038023	signaling receptor activity	1391	43	9.66	7.6e-18
<b>6</b>	GO:0060089	molecular transducer activity	1391	43	9.66	7.6e-18
<b>7</b>	GO:0004888	transmembrane	1113	37	7.73	3.1e-16



		signaling receptor activit...				
<b>8</b>	GO:0016 798	hydrolase activity, acting on glycosyl b...	437	19	3.03	1.7e-10
<b>9</b>	GO:0005 262	calcium channel activity	221	13	1.53	4.5e-09
<b>10</b>	GO:0015 085	calcium ion transme mbrane transport er ac...	242	13	1.68	1.3e-08
<b>11</b>	GO:0005 245	voltage- gated calcium channel activity	66	8	0.46	1.7e-08
<b>12</b>	GO:0001 609	G protein- coupled adenosin e receptor act...	13	5	0.09	1.8e-08
<b>13</b>	GO:0001 594	trace- amine receptor activity	6	4	0.04	3.3e-08

<b>14</b>	GO:0072509	divalent inorganic cation transmembrane ...	268	13	1.86	4.5e-08
<b>15</b>	GO:0008227	G protein-coupled amine receptor activit...	88	8	0.61	1.7e-07

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0002224	toll-like receptor signaling pathway	146	17	0.84	6.9e-18
<b>2</b>	GO:0002221	pattern recognition receptor signaling p...	202	17	1.16	1.8e-15
<b>3</b>	GO:0007154	cell communication	5266	59	30.17	2.3e-09

<b>4</b>	GO:0023052	signaling	5200	58	29.79	4.5e-09
<b>5</b>	GO:0007165	signal transduction	4753	55	27.23	4.5e-09
<b>6</b>	GO:0010951	negative regulation of endopeptidase act...	154	9	0.88	2.5e-07
<b>7</b>	GO:0051716	cellular response to stimulus	6077	60	34.82	2.5e-07
<b>8</b>	GO:0010466	negative regulation of peptidase activit...	159	9	0.91	3.2e-07
<b>9</b>	GO:0050896	response to stimulus	7920	69	45.38	1.7e-06
<b>10</b>	GO:0045861	negative regulation of proteolysis	336	11	1.93	3.5e-06
<b>11</b>	GO:0006952	defense response	1531	24	8.77	4.2e-06

<b>12</b>	GO:0061515	myeloid cell development	76	6	0.44	4.8e-06
<b>13</b>	GO:2000697	negative regulation of epithelial cell d...	7	3	0.04	6.3e-06
<b>14</b>	GO:0010460	positive regulation of heart rate	23	4	0.13	8.3e-06
<b>15</b>	GO:0045087	innate immune response	881	17	5.05	9.9e-06

Gains vestimentifera  
MF

<b>GO.ID</b>	<b>Term</b>	<b>Annotated</b>	<b>Significant</b>	<b>Expected</b>	<b>classicFisher</b>
--------------	-------------	------------------	--------------------	-----------------	----------------------

<b>1</b>	GO:0005509	calcium ion binding	1702	193	76.65	1e-30
<b>2</b>	GO:0034185	apolipoprotein binding	85	30	3.83	1.9e-19

<b>3</b>	GO:0030971	receptor tyrosine kinase binding	104	28	4.68	8.5e-15
<b>4</b>	GO:1990782	protein tyrosine kinase binding	123	29	5.54	1.2e-13
<b>5</b>	GO:0043167	ion binding	8110	458	365.24	4.9e-12
<b>6</b>	GO:0043169	cation binding	5894	351	265.44	4.8e-11
<b>7</b>	GO:0046872	metal ion binding	5800	345	261.21	1.0e-10
<b>8</b>	GO:0038023	signaling receptor activity	1391	113	62.64	3.8e-10
<b>9</b>	GO:0060089	molecular transducer activity	1391	113	62.64	3.8e-10
<b>10</b>	GO:0015026	coreceptor activity	39	14	1.76	6.6e-10
<b>11</b>	GO:0005102	signaling receptor binding	1191	100	53.64	7.9e-10
<b>12</b>	GO:0030	lipoprotei	31	12	1.40	4.1e-09

	228	n particle receptor activity				
<b>13</b>	GO:0042806	fucose binding	123	23	5.54	5.5e-09
<b>14</b>	GO:0004096	catalase activity	32	12	1.44	6.3e-09
<b>15</b>	GO:0038024	cargo receptor activity	109	21	4.91	1.5e-08

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0022610	biological adhesion	1876	179	82.37	1.0e-24
<b>2</b>	GO:0007155	cell adhesion	1817	175	79.78	1.4e-24
<b>3</b>	GO:0150094	amyloid- beta clearance by cellular catab...	54	25	2.37	4.0e-20

4	GO:0097242	amyloid-beta clearance	66	27	2.90	6.9e-20
5	GO:0007160	cell-matrix adhesion	306	55	13.43	2.5e-19
6	GO:0030279	negative regulation of ossification	113	33	4.96	1.0e-18
7	GO:1901629	regulation of presynaptic membrane organization	45	21	1.98	3.3e-17
8	GO:1901631	positive regulation of presynaptic membrane organization	45	21	1.98	3.3e-17
9	GO:1904395	positive regulation of skeletal muscle attachment	46	21	2.02	5.9e-17
10	GO:1904393	regulation of skeletal muscle attachment	47	21	2.06	1.0e-16

		muscle acetylcho l...				
<b>11</b>	GO:1901 628	positive regulatio n of postsyna ptic memb...	49	21	2.15	2.9e-16
<b>12</b>	GO:1903 911	positive regulatio n of receptor clusteri...	50	21	2.20	4.8e-16
<b>13</b>	GO:0031 589	cell- substrate adhesion	470	64	20.64	5.5e-16
<b>14</b>	GO:1903 909	regulatio n of receptor clusterin g	52	21	2.28	1.3e-15
<b>15</b>	GO:0032 612	interleuki n-1 productio n	108	29	4.74	1.6e-15

Gains vestimentifera\_cl1  
MF

GO.ID	Term	Annotated	Significant	Expected	
-------	------	-----------	-------------	----------	--



						classicFisher
--	--	--	--	--	--	---------------

1	GO:0008028	monocarboxylic acid transmembrane transp...	117	15	0.82	3.4e-15
2	GO:0046943	carboxylic acid transmembrane transport e...	251	16	1.76	2.2e-11
3	GO:0005342	organic acid transmembrane transporter a...	252	16	1.76	2.4e-11
4	GO:0015293	symporter activity	273	15	1.91	7.8e-10
5	GO:0008514	organic anion transmembrane transporter ...	326	16	2.28	1.1e-09

<b>6</b>	GO:0004842	ubiquitin-protein transferase activity	432	17	3.02	8.6e-09
<b>7</b>	GO:0015291	secondary active transmembrane transport.	381	16	2.67	9.8e-09
<b>8</b>	GO:0019787	ubiquitin-like protein transferase activ...	444	17	3.11	1.3e-08
<b>9</b>	GO:0008509	anion transmembrane transporter activity	485	16	3.40	2.7e-07
<b>10</b>	GO:0022804	active transmembrane transporter activity...	553	16	3.87	1.5e-06
<b>11</b>	GO:0004100	chitin synthase activity	15	4	0.11	2.9e-06
<b>12</b>	GO:0008467	[heparan sulfate]-glucosam	15	4	0.11	2.9e-06

		ine 3-sulfotra ...				
<b>13</b>	GO:0005308	creatine transme mbrane transport er activ...	33	5	0.23	3.1e-06
<b>14</b>	GO:0034483	heparan sulfate sulfotran sferase activit...	18	4	0.13	6.5e-06
<b>15</b>	GO:0005509	calcium ion binding	1702	27	11.91	3.8e-05

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0015718	monocar boxylic acid transport	180	15	1.15	5.8e-13
<b>2</b>	GO:0070534	protein K63- linked ubiquitin	69	10	0.44	1.9e-11

		ation				
<b>3</b>	GO:0032608	interferon-beta production	59	9	0.38	1.3e-10
<b>4</b>	GO:0051607	defense response to virus	208	13	1.33	8.0e-10
<b>5</b>	GO:0046942	carboxylic acid transport	351	16	2.25	8.4e-10
<b>6</b>	GO:0015849	organic acid transport	352	16	2.26	8.7e-10
<b>7</b>	GO:0001816	cytokine production	526	19	3.37	9.5e-10
<b>8</b>	GO:0034340	response to type I interferon	81	9	0.52	2.3e-09
<b>9</b>	GO:0009615	response to virus	240	13	1.54	4.6e-09
<b>10</b>	GO:0032606	type I interferon production	103	9	0.66	2.0e-08
<b>11</b>	GO:0032621	interleukin-18 production	33	6	0.21	5.8e-08

		n				
<b>12</b>	GO:0006952	defense response	1531	29	9.81	7.4e-08
<b>13</b>	GO:0000209	protein polyubiquitination	310	13	1.99	9.5e-08
<b>14</b>	GO:0015711	organic anion transport	500	16	3.20	1.2e-07
<b>15</b>	GO:0098542	defense response to other organism	1126	24	7.22	1.5e-07

Gains vestimentifera\_cl2  
MF

<b>GO.ID</b>	<b>Term</b>	<b>Annotated</b>	<b>Significant</b>	<b>Expected</b>	<b>classicFisher</b>
--------------	-------------	------------------	--------------------	-----------------	----------------------

<b>1</b>	GO:0005509	calcium ion binding	1702	56	13.01	8.1e-23
<b>2</b>	GO:0004939	beta-adrenergic receptor activity	30	9	0.23	8.4e-13
<b>3</b>	GO:0004	adrenergi	46	9	0.35	5.8e-11

	935	c receptor activity				
<b>4</b>	GO:0008 227	G protein- coupled amine receptor activit...	88	11	0.67	6.6e-11
<b>5</b>	GO:0004 940	beta1- adrenergi c receptor activity	11	6	0.08	7.9e-11
<b>6</b>	GO:0009 374	biotin binding	205	14	1.57	5.6e-10
<b>7</b>	GO:0038 023	signaling receptor activity	1391	34	10.63	7.1e-10
<b>8</b>	GO:0060 089	molecula r transduc er activity	1391	34	10.63	7.1e-10
<b>9</b>	GO:0033 293	monocar boxylic acid binding	230	14	1.76	2.5e-09
<b>10</b>	GO:0042 806	fucose binding	123	11	0.94	2.5e-09

<b>11</b>	GO:0031406	carboxylic acid binding	413	17	3.16	1.7e-08
<b>12</b>	GO:0043177	organic acid binding	413	17	3.16	1.7e-08
<b>13</b>	GO:0005112	Notch binding	159	11	1.22	3.8e-08
<b>14</b>	GO:0043169	cation binding	5894	75	45.04	6.6e-08
<b>15</b>	GO:0046872	metal ion binding	5800	74	44.32	8.1e-08

BP

<b>GO.ID</b>	<b>Term</b>	<b>Annotated</b>	<b>Significant</b>	<b>Expected</b>	<b>classicFisher</b>
--------------	-------------	------------------	--------------------	-----------------	----------------------

<b>1</b>	GO:0006959	humoral immune response	263	18	1.72	9.8e-14
<b>2</b>	GO:0035239	tube morphogenesis	958	30	6.25	3.4e-13
<b>3</b>	GO:1903524	positive regulation of	49	10	0.32	6.2e-13

		blood circulation				
4	GO:0045823	positive regulation of heart contraction	36	9	0.23	1.3e-12
5	GO:0031960	response to corticosteroid	140	13	0.91	7.0e-12
6	GO:0001525	angiogenesis	509	21	3.32	1.3e-11
7	GO:0001867	complement activation, lectin pathway	154	13	1.00	2.3e-11
8	GO:0071880	adenylate cyclase-activating adrenergic ...	50	9	0.33	3.1e-11
9	GO:0006956	complement activation	159	13	1.04	3.5e-11
10	GO:0071875	adrenergic	52	9	0.34	4.5e-11



		receptor signaling pathway				
<b>11</b>	GO:0007189	adenylate cyclase- activatin g G protein- c...	165	13	1.08	5.6e-11
<b>12</b>	GO:0035295	tube developm ent	1182	30	7.71	6.6e-11
<b>13</b>	GO:0051384	response to glucocort icoid	107	11	0.70	1.0e-10
<b>14</b>	GO:0002682	regulatio n of immune system process	1144	29	7.46	1.5e-10
<b>15</b>	GO:0048514	blood vessel morphog enesis	589	21	3.84	2.0e-10

## Paraescarpia

Expansions

MF

GO.ID	Term	Annotated	Significant	Expected	
-------	------	-----------	-------------	----------	--

						<b>classicFisher</b>
--	--	--	--	--	--	----------------------

<b>1</b>	GO:0004867	serine-type endopeptidase inhibitor activity	166	82	5.19	1e-30
<b>2</b>	GO:0004866	endopeptidase inhibitor activity	222	82	6.95	1e-30
<b>3</b>	GO:0061135	endopeptidase regulator activity	226	82	7.07	1e-30
<b>4</b>	GO:0030414	peptidase inhibitor activity	236	82	7.39	1e-30
<b>5</b>	GO:0061134	peptidase regulator activity	254	82	7.95	1e-30
<b>6</b>	GO:0071558	histone demethylase activity (H3-K27 spe...	94	53	2.94	1e-30

<b>7</b>	GO:0004857	enzyme inhibitor activity	312	82	9.76	1e-30
<b>8</b>	GO:0032452	histone demethylase activity	108	53	3.38	1e-30
<b>9</b>	GO:0140457	protein demethylase activity	108	53	3.38	1e-30
<b>10</b>	GO:0032451	demethylase activity	117	53	3.66	1e-30
<b>11</b>	GO:0031490	chromatin DNA binding	132	53	4.13	1e-30
<b>12</b>	GO:0015280	ligand-gated sodium channel activity	96	47	3.00	1e-30
<b>13</b>	GO:0051213	dioxygenase activity	153	53	4.79	1e-30
<b>14</b>	GO:0005272	sodium channel activity	118	47	3.69	1e-30
<b>15</b>	GO:0030	enzyme regulator	657	82	20.56	1.9e-28

	234	activity				
--	-----	----------	--	--	--	--

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

1	GO:0071557	histone H3-K27 demethylation	94	53	2.68	1e-30
2	GO:0070076	histone lysine demethylation	107	53	3.05	1e-30
3	GO:0016577	histone demethylation	108	53	3.08	1e-30
4	GO:0006482	protein demethylation	110	53	3.14	1e-30
5	GO:0008214	protein dealkylation	110	53	3.14	1e-30
6	GO:0070988	demethylation	134	53	3.82	1e-30
7	GO:0051568	histone H3-K4 methylation	111	47	3.17	1e-30

<b>8</b>	GO:0034968	histone lysine methylation	142	47	4.05	1e-30
<b>9</b>	GO:0018022	peptidyl-lysine methylation	152	47	4.34	1e-30
<b>10</b>	GO:0035725	sodium ion transmembrane transport	161	47	4.59	1e-30
<b>11</b>	GO:0016571	histone methylation	179	47	5.11	1e-30
<b>12</b>	GO:0006338	chromatin remodeling	194	48	5.53	1e-30
<b>13</b>	GO:0006479	protein methylation	196	47	5.59	1.2e-30
<b>14</b>	GO:0008213	protein alkylation	196	47	5.59	1.2e-30
<b>15</b>	GO:0018205	peptidyl-lysine modification	305	55	8.70	5.4e-29

Gains  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

1	GO:0005540	hyaluronic acid binding	67	40	2.66	1e-30
2	GO:0000978	RNA polymerase II cis-regulatory region ...	653	93	25.89	1.6e-28
3	GO:0000987	cis-regulatory region sequence-specific ...	669	94	26.53	2.3e-28
4	GO:0005201	extracellular matrix structural constitu...	154	46	6.11	3.1e-28
5	GO:0000977	RNA polymerase II transcription regulation...	758	98	30.06	1.1e-26

<b>6</b>	GO:0003724	RNA helicase activity	143	42	5.67	1.6e-25
<b>7</b>	GO:0070883	pre-miRNA binding	52	27	2.06	1.3e-24
<b>8</b>	GO:0046872	metal ion binding	4005	264	158.82	1.4e-24
<b>9</b>	GO:0070034	telomerase RNA binding	63	29	2.50	2.1e-24
<b>10</b>	GO:0031406	carboxylic acid binding	284	56	11.26	3.6e-24
<b>11</b>	GO:0043177	organic acid binding	284	56	11.26	3.6e-24
<b>12</b>	GO:0002151	G-quadruplex RNA binding	54	27	2.14	4.9e-24
<b>13</b>	GO:0000976	transcription regulatory region sequence ...	835	99	33.11	5.0e-24
<b>14</b>	GO:0001	regulator	835	99	33.11	5.0e-24

	067	y region nucleic acid binding				
<b>15</b>	GO:0043 169	cation binding	4066	264	161.24	1.7e-23

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0016 182	synaptic vesicle budding from endosome	68	52	2.71	1e-30
<b>2</b>	GO:0098 943	neurotransmitter receptor transport, pos...	68	52	2.71	1e-30
<b>3</b>	GO:0036 466	synaptic vesicle recycling via endosome	73	52	2.91	1e-30



<b>4</b>	GO:0048499	synaptic vesicle membrane organization	74	52	2.95	1e-30
<b>5</b>	GO:0070142	synaptic vesicle budding	75	52	2.99	1e-30
<b>6</b>	GO:0099532	synaptic vesicle endosomal processing	76	52	3.03	1e-30
<b>7</b>	GO:0048490	anterograde synaptic vesicle transport	78	52	3.11	1e-30
<b>8</b>	GO:0099514	synaptic vesicle cytoskeletal transport	78	52	3.11	1e-30
<b>9</b>	GO:0099517	synaptic vesicle transport along microtu...	78	52	3.11	1e-30
<b>10</b>	GO:0006896	Golgi to vacuole transport	82	52	3.27	1e-30

<b>11</b>	GO:0008089	anterograde axonal transport	87	52	3.47	1e-30
<b>12</b>	GO:0099637	neurotransmitter receptor transport	93	53	3.71	1e-30
<b>13</b>	GO:0018146	keratan sulfate biosynthetic process	47	40	1.88	1e-30
<b>14</b>	GO:0042340	keratan sulfate catabolic process	48	40	1.92	1e-30
<b>15</b>	GO:0047496	vesicle transport along microtubule	94	52	3.75	1e-30

Gains siboglinidae  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

--	--	--	--	--	--	--

1	GO:0005344	oxygen carrier activity	57	8	0.25	1.1e-10
2	GO:0005215	transporter activity	1207	21	5.31	1.0e-08
3	GO:0019825	oxygen binding	50	6	0.22	7.3e-08
4	GO:0042166	acetylcholine binding	13	4	0.06	2.3e-07
5	GO:0042165	neurotransmitter binding	18	4	0.08	9.7e-07
6	GO:0022848	acetylcholine-gated cation-selective cha...	55	5	0.24	4.0e-06
7	GO:0008559	ATPase-coupled xenobiotic transmembrane ...	63	5	0.28	7.8e-06
8	GO:1904315	transmitter-gated ion	79	5	0.35	2.4e-05

		channel activity i...				
<b>9</b>	GO:0099 529	neurotran smitter receptor activity invol...	80	5	0.35	2.5e-05
<b>10</b>	GO:0005 231	excitatory extracellu lar ligand- gated io...	86	5	0.38	3.6e-05
<b>11</b>	GO:0098 960	postsyna ptic neurotran smitter receptor a...	87	5	0.38	3.8e-05
<b>12</b>	GO:0042 910	xenobioti c transme mbrane transport er act...	88	5	0.39	4.0e-05
<b>13</b>	GO:0020 037	heme binding	163	6	0.72	7.5e-05
<b>14</b>	GO:0022 824	transmitt er-gated ion	102	5	0.45	8.1e-05

		channel activity				
15	GO:0022835	transmitter-gated channel activity	102	5	0.45	8.1e-05

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

1	GO:0007271	synaptic transmission, cholinergic	41	5	0.14	2.4e-07
2	GO:0035094	response to nicotine	34	4	0.12	4.9e-06
3	GO:2000300	regulation of synaptic vesicle exocytosis.	43	4	0.15	1.3e-05
4	GO:0098693	regulation of	55	4	0.19	3.4e-05

		synaptic vesicle cycle				
5	GO:0016079	synaptic vesicle exocytosis	58	4	0.20	4.2e-05
6	GO:0046928	regulation of neurotransmitter secretion	64	4	0.22	6.3e-05
7	GO:0051588	regulation of neurotransmitter transport	64	4	0.22	6.3e-05
8	GO:1903305	regulation of regulated secretory pathway	72	4	0.24	9.9e-05
9	GO:0043362	nucleate erythrocyte maturation	6	2	0.02	0.00017
10	GO:0043363	nucleate erythrocyte differentiation	6	2	0.02	0.00017

		ation				
<b>11</b>	GO:0048823	nucleate erythrocyte development	6	2	0.02	0.00017
<b>12</b>	GO:0007269	neurotransmitter secretion	91	4	0.31	0.00025
<b>13</b>	GO:0099643	signal release from synapse	91	4	0.31	0.00025
<b>14</b>	GO:0017157	regulation of exocytosis	100	4	0.34	0.00035
<b>15</b>	GO:0061515	myeloid cell development	42	3	0.14	0.00038

Gains vestimentifera  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

--	--	--	--	--	--	--

<b>1</b>	GO:0042301	phosphate ion binding	28	16	1.11	5.9e-16
<b>2</b>	GO:0052689	carboxylic ester hydrolase activity	115	28	4.57	5.0e-15
<b>3</b>	GO:0004096	catalase activity	30	15	1.19	7.0e-14
<b>4</b>	GO:0003700	DNA-binding transcription factor activit...	632	66	25.12	2.9e-13
<b>5</b>	GO:0000981	DNA-binding transcription factor activit...	595	63	23.65	5.9e-13
<b>6</b>	GO:0017171	serine hydrolase activity	141	27	5.60	7.3e-12
<b>7</b>	GO:0005509	calcium ion binding	945	79	37.56	1.1e-10
<b>8</b>	GO:0004601	peroxidase activity	57	16	2.27	3.9e-10
<b>9</b>	GO:0046	tetrapyrr	181	28	7.19	5.7e-10



	906	ole binding				
<b>10</b>	GO:0016684	oxidoreductase activity, acting on perox...	59	16	2.34	6.8e-10
<b>11</b>	GO:0020037	heme binding	163	26	6.48	1.2e-09
<b>12</b>	GO:0038024	cargo receptor activity	77	17	3.06	6.2e-09
<b>13</b>	GO:0016209	antioxidant activity	77	16	3.06	4.4e-08
<b>14</b>	GO:0017070	U6 snRNA binding	16	8	0.64	5.7e-08
<b>15</b>	GO:0000977	RNA polymerase II transcription regulation	758	61	30.12	8.2e-08

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0046485	ether lipid metabolic process	48	26	1.81	6.3e-25
<b>2</b>	GO:0006662	glycerol ether metabolic process	50	26	1.89	2.6e-24
<b>3</b>	GO:0018904	ether metabolic process	50	26	1.89	2.6e-24
<b>4</b>	GO:0060395	SMAD protein signal transduction	43	16	1.62	1.4e-12
<b>5</b>	GO:0006805	xenobiotic metabolic process	49	16	1.85	1.4e-11
<b>6</b>	GO:0071466	cellular response to xenobiotic stimulus	50	16	1.89	2.0e-11
<b>7</b>	GO:1903584	regulation of histone	9	8	0.34	3.4e-11

		deubiquitination				
<b>8</b>	GO:1903586	positive regulation of histone deubiquitination	9	8	0.34	3.4e-11
<b>9</b>	GO:0009410	response to xenobiotic stimulus	53	16	2.00	5.4e-11
<b>10</b>	GO:0090085	regulation of protein deubiquitination	14	9	0.53	2.4e-10
<b>11</b>	GO:1903003	positive regulation of protein deubiquitination	14	9	0.53	2.4e-10
<b>12</b>	GO:0016042	lipid catabolic process	224	31	8.46	3.3e-10
<b>13</b>	GO:0034383	low-density lipoprotein particle clearance	22	10	0.83	2.3e-09

<b>14</b>	GO:0099560	synaptic membrane adhesion	19	9	0.72	9.5e-09
<b>15</b>	GO:0007156	homophilic cell adhesion via plasma memb...	141	22	5.33	1.4e-08

Gains vestimentifera\_cl1  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0140110	transcription regulator activity	936	28	8.96	3.6e-08
<b>2</b>	GO:0000981	DNA-binding transcription factor activit...	595	22	5.69	3.7e-08
<b>3</b>	GO:0003700	DNA-binding transcript	632	22	6.05	1.1e-07

		ion factor activit...				
4	GO:0000 977	RNA polymera se II transcript ion regulato...	758	23	7.25	6.1e-07
5	GO:0000 978	RNA polymera se II cis- regulator y region ...	653	21	6.25	8.0e-07
6	GO:0000 987	cis- regulator y region sequence -specific ...	669	21	6.40	1.2e-06
7	GO:0001 228	DNA- binding transcript ion activator acti...	181	11	1.73	1.2e-06
8	GO:0043 565	sequence -specific DNA binding	981	26	9.39	1.3e-06
9	GO:0001 216	DNA- binding transcript	184	11	1.76	1.4e-06

		ion activator acti...				
<b>10</b>	GO:0000976	transcript ion regulator y region sequence ...	835	23	7.99	3.2e-06
<b>11</b>	GO:0001067	regulator y region nucleic acid binding	835	23	7.99	3.2e-06
<b>12</b>	GO:1990837	sequence -specific double- stranded DNA bi...	862	23	8.25	5.5e-06
<b>13</b>	GO:0003690	double- stranded DNA binding	930	24	8.90	5.8e-06
<b>14</b>	GO:0070403	NAD+ binding	52	5	0.50	0.00013
<b>15</b>	GO:0003677	DNA binding	1673	30	16.01	0.00037

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

1	GO:0035641	locomotory exploration behavior	15	6	0.15	4.6e-09
2	GO:0035640	exploration behavior	17	6	0.17	1.1e-08
3	GO:0007219	Notch signaling pathway	177	12	1.80	2.4e-07
4	GO:0007495	visceral mesoderm-endoderm interaction i...	7	4	0.07	3.5e-07
5	GO:0006357	regulation of transcription by RNA	1306	33	13.31	4.9e-07

		polym...				
<b>6</b>	GO:0006355	regulation of transcription, DNA-templat...	1755	39	17.88	9.0e-07
<b>7</b>	GO:1903506	regulation of nucleic acid-templated tra...	1757	39	17.90	9.2e-07
<b>8</b>	GO:2001141	regulation of RNA biosynthetic process	1764	39	17.97	1.0e-06
<b>9</b>	GO:0048537	mucosa-associated lymphoid tissue develo...	9	4	0.09	1.2e-06
<b>10</b>	GO:0048541	Peyer's patch development	9	4	0.09	1.2e-06
<b>11</b>	GO:0061146	Peyer's patch morphogenesis	9	4	0.09	1.2e-06



<b>12</b>	GO:0006366	transcription by RNA polymerase II	1377	33	14.03	1.6e-06
<b>13</b>	GO:0007155	cell adhesion	1007	27	10.26	2.3e-06
<b>14</b>	GO:0022610	biological adhesion	1009	27	10.28	2.4e-06
<b>15</b>	GO:0006351	transcription, DNA-templated	1873	39	19.08	4.6e-06

## Lamellibrachia

Expansions

MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

1	GO:0005509	calcium ion binding	1351	73	13.31	1e-30

2	GO:0004930	G protein-coupled receptor activity	658	44	6.48	7.2e-25
3	GO:0008188	neuropeptide receptor activity	187	25	1.84	1.9e-21
4	GO:0008528	G protein-coupled peptide receptor activ...	268	26	2.64	1.0e-18
5	GO:0042806	fucose binding	67	16	0.66	2.6e-18
6	GO:0001653	peptide receptor activity	341	26	3.36	4.0e-16
7	GO:0004888	transmembrane signaling receptor activit...	1175	45	11.58	7.3e-16
8	GO:0008142	oxysterol binding	10	8	0.10	3.3e-15

<b>9</b>	GO:0048029	monosaccharide binding	112	16	1.10	1.5e-14
<b>10</b>	GO:0038023	signaling receptor activity	1467	45	14.45	2.1e-12
<b>11</b>	GO:0060089	molecular transducer activity	1467	45	14.45	2.1e-12
<b>12</b>	GO:0036312	phosphatidylinositol 3-kinase regulatory...	19	8	0.19	5.1e-12
<b>13</b>	GO:0043548	phosphatidylinositol 3-kinase binding	30	8	0.30	3.6e-10
<b>14</b>	GO:0009374	biotin binding	108	12	1.06	6.7e-10
<b>15</b>	GO:0004995	tachykinin receptor activity	16	6	0.16	6.1e-09

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

1	GO:0007160	cell-matrix adhesion	232	33	2.06	1.2e-30
2	GO:0006955	immune response	1122	53	9.97	2.2e-25
3	GO:0031589	cell-substrate adhesion	366	34	3.25	2.9e-25
4	GO:0032101	regulation of response to external stimu...	844	46	7.50	2.3e-24
5	GO:0007155	cell adhesion	1488	57	13.23	6.3e-23
6	GO:0050776	regulation of immune response	542	37	4.82	8.8e-23

<b>7</b>	GO:0045087	innate immune response	701	41	6.23	9.4e-23
<b>8</b>	GO:1901223	negative regulation of NIK/NF-kappaB sig...	27	14	0.24	1.8e-22
<b>9</b>	GO:0045088	regulation of innate immune response	320	30	2.84	2.0e-22
<b>10</b>	GO:0022610	biological adhesion	1529	57	13.59	2.4e-22
<b>11</b>	GO:0002287	alpha-beta T cell activation involved in...	30	14	0.27	1.3e-21
<b>12</b>	GO:0002293	alpha-beta T cell differentiation involv...	30	14	0.27	1.3e-21
<b>13</b>	GO:0002294	CD4-positive, alpha-beta T	30	14	0.27	1.3e-21

		cell differen...				
<b>14</b>	GO:0042093	T-helper cell differenti ation	30	14	0.27	1.3e-21
<b>15</b>	GO:0002682	regulatio n of immune system process	909	43	8.08	2.5e-20

Gains  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0004521	endoribo nuclease activity	91	21	2.36	1.2e-14
<b>2</b>	GO:0004519	endonucl ease activity	194	26	5.03	6.3e-12
<b>3</b>	GO:0051082	unfolded protein binding	119	20	3.09	2.8e-11

<b>4</b>	GO:0004540	ribonucle ase activity	140	21	3.63	8.2e-11
<b>5</b>	GO:0015081	sodium ion transme mbrane transport er act...	330	31	8.56	6.2e-10
<b>6</b>	GO:0140098	catalytic activity, acting on RNA	493	37	12.79	6.9e-09
<b>7</b>	GO:0015280	ligand- gated sodium channel activity	76	14	1.97	7.9e-09
<b>8</b>	GO:0004674	protein serine/th reonine kinase activity	506	37	13.12	1.4e-08
<b>9</b>	GO:0005272	sodium channel activity	128	17	3.32	3.4e-08
<b>10</b>	GO:0004518	nuclease activity	294	26	7.63	5.3e-08
<b>11</b>	GO:0004672	protein kinase activity	686	41	17.79	5.7e-07

12	GO:0016772	transferase activity, transferring phosph...	1094	55	28.37	1.6e-06
13	GO:0097367	carbohydrate derivative binding	2949	114	76.49	2.9e-06
14	GO:0046873	metal ion transmembrane transporter acti...	733	41	19.01	3.1e-06
15	GO:0015077	monovalent inorganic cation transmembran...	496	31	12.86	6.0e-06

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

--	--	--	--	--	--	--



1	GO:0036498	IRE1-mediated unfolded protein response	60	19	1.54	3.0e-16
2	GO:0030968	endoplasmic reticulum unfolded protein r...	95	22	2.43	2.0e-15
3	GO:0034620	cellular response to unfolded protein	112	22	2.87	7.9e-14
4	GO:0035967	cellular response to topologically incor...	126	22	3.23	9.8e-13
5	GO:0006986	response to unfolded protein	140	23	3.59	1.1e-12
6	GO:0035966	response to topologically incorrect prot...	155	23	3.97	1.0e-11
7	GO:0034	response to	239	26	6.12	5.3e-10

	976	endoplas mic reticulum stress				
<b>8</b>	GO:0070 059	intrinsic apoptotic signaling pathway in...	69	14	1.77	1.8e-09
<b>9</b>	GO:0071 852	funga- type cell wall organizat ion or bi...	28	9	0.72	1.9e-08
<b>10</b>	GO:0036 290	protein trans- autophos phorylati on	14	7	0.36	2.0e-08
<b>11</b>	GO:0031 505	funga- type cell wall organizat ion	27	8	0.69	2.5e-07
<b>12</b>	GO:0043 507	positive regulatio n of JUN kinase activi...	50	9	1.28	4.3e-06
<b>13</b>	GO:0016 071	mRNA metaboli c process	679	37	17.39	1.3e-05
<b>14</b>	GO:0043	regulatio	57	9	1.46	1.3e-05

	506	n of JUN kinase activity				
<b>15</b>	GO:0006033	chitin localization	32	7	0.82	1.3e-05

Gains siboglinidae  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0003953	NAD+ nucleosidase activity	73	20	0.55	2.2e-26
<b>2</b>	GO:0050135	NAD(P)+ nucleosidase activity	73	20	0.55	2.2e-26
<b>3</b>	GO:0061809	NAD+ nucleotidase, cyclic ADP- ribose gen...	73	20	0.55	2.2e-26
<b>4</b>	GO:0016799	hydrolase activity, hydrolyzi	97	20	0.73	1.2e-23

		ng N-glycos.. .				
5	GO:0004888	transmembrane signaling receptor activit...	1175	43	8.89	3.3e-19
6	GO:0038023	signaling receptor activity	1467	47	11.10	8.4e-19
7	GO:0060089	molecular transducer activity	1467	47	11.10	8.4e-19
8	GO:0016798	hydrolase activity, acting on glycosyl b...	351	20	2.66	1.8e-12
9	GO:0008227	G protein-coupled amine receptor activit...	78	8	0.59	1.3e-07
10	GO:0090554	phosphatidylcholine floppase activity	39	6	0.30	4.4e-07

<b>11</b>	GO:0008525	phosphatidylcholine transporter activity	42	6	0.32	6.9e-07
<b>12</b>	GO:0005344	oxygen carrier activity	70	7	0.53	9.6e-07
<b>13</b>	GO:0140328	floppase activity	46	6	0.35	1.2e-06
<b>14</b>	GO:0001594	trace-amine receptor activity	4	3	0.03	1.7e-06
<b>15</b>	GO:0030273	melanin-concentrating hormone receptor a...	4	3	0.03	1.7e-06

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher	
-------	------	-----------	-------------	----------	---------------	--

<b>1</b>	GO:0002224	toll-like receptor signaling	157	20	1.02	1.4e-20
----------	------------	------------------------------	-----	----	------	---------

		pathway				
<b>2</b>	GO:0002221	pattern recognition receptor signaling p...	235	20	1.53	4.6e-17
<b>3</b>	GO:0045087	innate immune response	701	25	4.58	2.3e-12
<b>4</b>	GO:0098542	defense response to other organism	864	25	5.64	2.0e-10
<b>5</b>	GO:0006952	defense response	1201	29	7.84	3.8e-10
<b>6</b>	GO:0007165	signal transduction	4536	60	29.60	4.5e-10
<b>7</b>	GO:0023052	signaling	4935	63	32.21	4.6e-10
<b>8</b>	GO:0007154	cell communication	5001	62	32.64	2.8e-09
<b>9</b>	GO:0009605	response to external stimulus	2350	39	15.34	9.2e-09
<b>10</b>	GO:0006955	immune response	1122	26	7.32	9.2e-09

<b>11</b>	GO:0043207	response to external biotic stimulus	1100	25	7.18	2.8e-08
<b>12</b>	GO:0051707	response to other organism	1100	25	7.18	2.8e-08
<b>13</b>	GO:0009607	response to biotic stimulus	1123	25	7.33	4.2e-08
<b>14</b>	GO:0050896	response to stimulus	7358	75	48.02	8.0e-08
<b>15</b>	GO:0002376	immune system process	1815	32	11.85	8.5e-08

Gains vestimentifera  
MF

<b>GO.ID</b>	<b>Term</b>	<b>Annotated</b>	<b>Significant</b>	<b>Expected</b>	<b>classicFisher</b>
--------------	-------------	------------------	--------------------	-----------------	----------------------

<b>1</b>	GO:0005509	calcium ion binding	1351	156	76.43	1.2e-18
<b>2</b>	GO:0001537	N-acetylglactosamine	19	12	1.07	3.5e-11

		4-O-sulfotransfera...				
3	GO:0038023	signaling receptor activity	1467	141	83.00	1.3e-10
4	GO:0060089	molecular transducer activity	1467	141	83.00	1.3e-10
5	GO:0005262	calcium channel activity	304	45	17.20	2.8e-09
6	GO:0008146	sulfotransferase activity	97	23	5.49	3.2e-09
7	GO:0008107	galactoside 2-alpha-L-fucosyl transferase...	23	11	1.30	1.3e-08
8	GO:0031127	alpha-(1,2)-fucosyltransferase activity	23	11	1.30	1.3e-08
9	GO:0015085	calcium ion transmembrane	321	45	18.16	1.6e-08



		transport er ac...				
<b>10</b>	GO:0004888	transme mbrane signaling receptor activit...	1175	112	66.48	2.2e-08
<b>11</b>	GO:0005261	cation channel activity	609	69	34.45	2.4e-08
<b>12</b>	GO:0009374	biotin binding	108	23	6.11	2.8e-08
<b>13</b>	GO:0072509	divalent inorganic cation transme mbrane ...	358	47	20.25	6.0e-08
<b>14</b>	GO:0005215	transport er activity	1905	160	107.78	1.1e-07
<b>15</b>	GO:0098960	postsyna ptic neurotran smmitter receptor a...	153	27	8.66	1.2e-07

BP

	Term	Annotated	Significant	Expected	
--	------	-----------	-------------	----------	--

GO.ID						classicFisher
-------	--	--	--	--	--	---------------

1	GO:0022610	biological adhesion	1529	145	79.64	3.4e-13
2	GO:0007155	cell adhesion	1488	141	77.51	8.3e-13
3	GO:0001867	complement activation, lectin pathway	128	28	6.67	7.0e-11
4	GO:0006956	complement activation	138	28	7.19	4.5e-10
5	GO:0050655	dermatan sulfate proteoglycan metabolic ...	24	11	1.25	9.6e-09
6	GO:0006968	cellular defense response	65	17	3.39	2.3e-08
7	GO:0009605	response to external stimulus	2350	180	122.41	2.4e-08

<b>8</b>	GO:0010185	regulation of cellular defense response	58	16	3.02	2.6e-08
<b>9</b>	GO:0050982	detection of mechanical stimulus	148	26	7.71	4.4e-08
<b>10</b>	GO:0001868	regulation of complement activation, lec...	62	16	3.23	7.3e-08
<b>11</b>	GO:0036065	fucosylation	88	19	4.58	1.0e-07
<b>12</b>	GO:0030449	regulation of complement activation	67	16	3.49	2.3e-07
<b>13</b>	GO:0002920	regulation of humoral immune response	84	18	4.38	2.4e-07

<b>14</b>	GO:0035641	locomotory exploration behavior	20	9	1.04	2.7e-07
<b>15</b>	GO:0035640	exploration behavior	26	10	1.35	3.5e-07

**Osedax**  
Expansions  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0004222	metalloendopeptidase activity	199	34	1.34	1e-30
<b>2</b>	GO:0008237	metallopeptidase activity	279	34	1.87	1e-30
<b>3</b>	GO:0004175	endopeptidase	331	34	2.22	1e-30

		activity				
<b>4</b>	GO:0008233	peptidase activity	488	34	3.28	7.0e-28
<b>5</b>	GO:0008270	zinc ion binding	561	34	3.77	7.6e-26
<b>6</b>	GO:0046914	transition metal ion binding	724	34	4.86	3.4e-22
<b>7</b>	GO:0008201	heparin binding	73	12	0.49	3.5e-14
<b>8</b>	GO:0004252	serine-type endopeptidase activity	61	11	0.41	1.5e-13
<b>9</b>	GO:0005539	glycosaminoglycan binding	97	12	0.65	1.2e-12
<b>10</b>	GO:0140096	catalytic activity, acting on a protein	1493	34	10.02	2.1e-12
<b>11</b>	GO:0008236	serine-type peptidase activity	90	11	0.60	1.3e-11
<b>12</b>	GO:0017171	serine hydrolase activity	94	11	0.63	2.1e-11

<b>13</b>	GO:1901681	sulfur compound binding	147	12	0.99	1.8e-10
<b>14</b>	GO:0016787	hydrolase activity	1828	34	12.27	7.0e-10
<b>15</b>	GO:0004930	G protein-coupled receptor activity	311	15	2.09	1.2e-09

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0030574	collagen catabolic process	72	31	0.55	1e-30
<b>2</b>	GO:0032963	collagen metabolic process	100	31	0.77	1e-30
<b>3</b>	GO:0030198	extracellular matrix organization	195	35	1.50	1e-30
<b>4</b>	GO:0043062	extracellular	205	35	1.57	1e-30

		structure organizat ion				
<b>5</b>	GO:0022 617	extracellu lar matrix disassem bly	37	18	0.28	1.3e-29
<b>6</b>	GO:0042 698	ovulation cycle	43	17	0.33	5.2e-26
<b>7</b>	GO:0035 987	endoder mal cell differenti ation	31	15	0.24	1.1e-24
<b>8</b>	GO:0001 706	endoder m formation	32	15	0.25	2.0e-24
<b>9</b>	GO:0007 492	endoder m developm ent	57	15	0.44	6.8e-20
<b>10</b>	GO:0001 704	formation of primary germ layer	60	15	0.46	1.6e-19
<b>11</b>	GO:0001 554	luteolysis	15	9	0.12	2.7e-16

<b>12</b>	GO:0001542	ovulation from ovarian follicle	16	9	0.12	6.0e-16
<b>13</b>	GO:0007155	cell adhesion	609	28	4.68	1.5e-15
<b>14</b>	GO:0022610	biological adhesion	619	28	4.76	2.2e-15
<b>15</b>	GO:0007369	gastrulation	112	15	0.86	3.6e-15

Gains  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0098631	cell adhesion mediator activity	76	43	2.72	1e-30
<b>2</b>	GO:0004222	metalloendopeptidase activity	199	59	7.12	1e-30
<b>3</b>	GO:0008	metalloprotease activity	279	59	9.99	5.3e-30



	237	peptidase activity				
4	GO:0004175	endopeptidase activity	331	63	11.85	2.8e-29
5	GO:0050839	cell adhesion molecule binding	207	50	7.41	2.7e-28
6	GO:0008233	peptidase activity	488	63	17.47	8.7e-20
7	GO:0042806	fucose binding	13	10	0.47	7.8e-13
8	GO:0005201	extracellular matrix structural constituent	85	21	3.04	1.2e-12
9	GO:0008270	zinc ion binding	561	55	20.08	3.9e-12
10	GO:0043169	cation binding	2687	149	96.18	1.8e-10
11	GO:0046872	metal ion binding	2638	145	94.43	8.3e-10
12	GO:0005518	collagen binding	48	13	1.72	7.7e-09

<b>13</b>	GO:0004806	triglyceride lipase activity	30	10	1.07	4.8e-08
<b>14</b>	GO:0046914	transition metal ion binding	724	55	25.91	5.3e-08
<b>15</b>	GO:0004607	phosphatidylcholine-sterol O-acyltransferase...	5	5	0.18	5.7e-08

BP

<b>GO.ID</b>	<b>Term</b>	<b>Annotated</b>	<b>Significant</b>	<b>Expected</b>	<b>classicFisher</b>
--------------	-------------	------------------	--------------------	-----------------	----------------------

<b>1</b>	GO:0030198	extracellular matrix organization	195	40	5.93	1.4e-22
<b>2</b>	GO:0043062	extracellular structure organization	205	40	6.23	1.0e-21

3	GO:0030574	collagen catabolic process	72	24	2.19	3.2e-19
4	GO:0032963	collagen metabolic process	100	26	3.04	1.1e-17
5	GO:0001867	complement activation, lectin pathway	10	9	0.30	1.9e-13
6	GO:0001868	regulation of complement activation, lec...	10	9	0.30	1.9e-13
7	GO:0010185	regulation of cellular defense response	10	9	0.30	1.9e-13
8	GO:0002920	regulation of humoral immune response	26	12	0.79	3.3e-12
9	GO:1900075	positive regulation of	9	8	0.27	5.8e-12

		neuromu scular syn...				
<b>10</b>	GO:0030 449	regulatio n of complem ent activation	13	9	0.40	1.3e-11
<b>11</b>	GO:1901 889	negative regulatio n of cell junction ass...	44	14	1.34	2.1e-11
<b>12</b>	GO:0006 956	complem ent activation	14	9	0.43	3.4e-11
<b>13</b>	GO:0002 684	positive regulatio n of immune system pro...	230	28	6.99	2.9e-10
<b>14</b>	GO:1900 073	regulatio n of neuromu scular synaptic tra...	12	8	0.36	2.9e-10
<b>15</b>	GO:0099 174	regulatio n of	29	11	0.88	3.6e-10

		presynapse organization				
--	--	-------------------------	--	--	--	--

Gains siboglinidae  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

1	GO:0038023	signaling receptor activity	606	23	5.42	1.5e-09
2	GO:0060089	molecular transducer activity	606	23	5.42	1.5e-09
3	GO:0004888	transmembrane signaling receptor activit...	492	19	4.40	4.2e-08
4	GO:0005344	oxygen carrier activity	35	6	0.31	5.6e-07
5	GO:0005215	transporter activity	978	25	8.75	6.8e-07

<b>6</b>	GO:0005509	calcium ion binding	533	18	4.77	7.6e-07
<b>7</b>	GO:0005216	ion channel activity	331	14	2.96	1.1e-06
<b>8</b>	GO:0022836	gated channel activity	263	12	2.35	3.3e-06
<b>9</b>	GO:0015267	channel activity	367	14	3.28	3.8e-06
<b>10</b>	GO:0030594	neurotransmitter receptor activity	105	8	0.94	4.0e-06
<b>11</b>	GO:0022803	passive transmembrane transporter activity	369	14	3.30	4.1e-06
<b>12</b>	GO:0015464	acetylcholine receptor activity	17	4	0.15	1.3e-05
<b>13</b>	GO:0008227	G protein-coupled amine receptor activity	35	5	0.31	1.3e-05

<b>14</b>	GO:0004937	alpha1-adrenergic receptor activity	6	3	0.05	1.4e-05
<b>15</b>	GO:0004252	serine-type endopeptidase activity	61	6	0.55	1.6e-05

BP

<b>GO.ID</b>	<b>Term</b>	<b>Annotated</b>	<b>Significant</b>	<b>Expected</b>	<b>classicFisher</b>
--------------	-------------	------------------	--------------------	-----------------	----------------------

<b>1</b>	GO:0008015	blood circulation	212	13	1.49	1.9e-09
<b>2</b>	GO:0003013	circulatory system process	271	14	1.91	4.0e-09
<b>3</b>	GO:0014061	regulation of norepinephrine secretion	5	4	0.04	1.1e-08
<b>4</b>	GO:0006	muscle contracti	124	10	0.87	1.3e-08

	936	on				
<b>5</b>	GO:0048243	norepinephrine secretion	6	4	0.04	3.3e-08
<b>6</b>	GO:0035296	regulation of tube diameter	79	8	0.56	6.8e-08
<b>7</b>	GO:0097746	regulation of blood vessel diameter	79	8	0.56	6.8e-08
<b>8</b>	GO:0003008	system process	908	22	6.39	1.0e-07
<b>9</b>	GO:0035150	regulation of tube size	89	8	0.63	1.7e-07
<b>10</b>	GO:0003012	muscle system process	165	10	1.16	1.9e-07
<b>11</b>	GO:0035556	intracellular signal transduction	1203	25	8.46	2.0e-07



<b>12</b>	GO:0006954	inflammatory response	168	10	1.18	2.3e-07
<b>13</b>	GO:0008016	regulation of heart contraction	63	7	0.44	2.5e-07
<b>14</b>	GO:0050727	regulation of inflammatory response	94	8	0.66	2.7e-07
<b>15</b>	GO:0007165	signal transduction	2377	36	16.72	3.3e-07

## Losses

Riftia

MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0003827	alpha-1,3-mannosyl glycoprotein	5	3	0.03	2.5e-06
----------	------------	---------------------------------	---	---	------	---------

		2-beta-N-...				
2	GO:0003677	DNA binding	2064	31	13.15	3.6e-06
3	GO:0004867	serine-type endopeptidase inhibitor acti...	163	7	1.04	8.3e-05
4	GO:0033971	hydroxyisourate hydrolase activity	4	2	0.03	0.00024
5	GO:0004866	endopeptidase inhibitor activity	215	7	1.37	0.00046
6	GO:0005539	glycosaminoglycan binding	286	8	1.82	0.00049
7	GO:0061135	endopeptidase regulator activity	220	7	1.40	0.00052
8	GO:0030414	peptidase inhibitor activity	237	7	1.51	0.00081

<b>9</b>	GO:0008201	heparin binding	179	6	1.14	0.00101
<b>10</b>	GO:0046982	protein heterodimerization activity	257	7	1.64	0.00130
<b>11</b>	GO:0003676	nucleic acid binding	3209	34	20.44	0.00131
<b>12</b>	GO:0003810	protein-glutamine gamma-glutamyl transferase activity	9	2	0.06	0.00141
<b>13</b>	GO:0061134	peptidase regulator activity	273	7	1.74	0.00183
<b>14</b>	GO:1901681	sulfur compound binding	326	7	2.08	0.00489
<b>15</b>	GO:0001054	RNA polymerase I activity	1	1	0.01	0.00637

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0006334	nucleosome assembly	43	5	0.27	7.7e-06
<b>2</b>	GO:0006049	UDP-N-acetylglucosamine catabolic process...	4	2	0.03	0.00024
<b>3</b>	GO:0034728	nucleosome organization	92	5	0.59	0.00031
<b>4</b>	GO:0031497	chromatin assembly	97	5	0.62	0.00039
<b>5</b>	GO:0009227	nucleotide-sugar catabolic process	5	2	0.03	0.00040
<b>6</b>	GO:0065004	protein-DNA complex assembly	102	5	0.65	0.00049
<b>7</b>	GO:0006335	DNA replication-dependent	6	2	0.04	0.00060

		nt nucleoso me ass...				
<b>8</b>	GO:0034 723	DNA replicatio n- depende nt nucleoso me org...	6	2	0.04	0.00060
<b>9</b>	GO:0006 333	chromati n assembly or disassem bly	113	5	0.72	0.00079
<b>10</b>	GO:1901 029	negative regulatio n of mitochon drial out...	7	2	0.04	0.00083
<b>11</b>	GO:0007 596	blood coagulati on	245	7	1.56	0.00099
<b>12</b>	GO:0007 599	hemostas is	247	7	1.58	0.00104
<b>13</b>	GO:0032 309	icosanoid secretion	33	3	0.21	0.00120
<b>14</b>	GO:0050	coagulati	261	7	1.67	0.00143

	817	on				
<b>15</b>	GO:0071 103	DNA conforma tion change	193	6	1.23	0.00150

Oasisia  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0005 525	GTP binding	434	19	4.19	4.4e-08
<b>2</b>	GO:0032 550	purine ribonucle oside binding	437	19	4.22	4.9e-08
<b>3</b>	GO:0001 883	purine nucleosid e binding	438	19	4.23	5.1e-08
<b>4</b>	GO:0032 549	ribonucle oside binding	441	19	4.26	5.7e-08
<b>5</b>	GO:0001 882	nucleosid e binding	444	19	4.29	6.3e-08
<b>6</b>	GO:0019	guanyl nucleotid	464	19	4.48	1.3e-07

	001	e binding				
<b>7</b>	GO:0032561	guanylyl ribonucleotide binding	464	19	4.48	1.3e-07
<b>8</b>	GO:0050051	leukotriene-B4 20-monooxygenase activity	8	4	0.08	5.7e-07
<b>9</b>	GO:0043167	ion binding	8450	108	81.67	4.4e-05
<b>10</b>	GO:0004139	deoxyribose-phosphate aldolase activity	2	2	0.02	9.3e-05
<b>11</b>	GO:0034722	gamma-glutamyl-peptidase activity	2	2	0.02	9.3e-05
<b>12</b>	GO:0050544	arachidonic acid binding	2	2	0.02	9.3e-05
<b>13</b>	GO:0004181	metallocarboxypeptidase	51	5	0.49	0.00013

		activity				
<b>14</b>	GO:000147	transcription termination site sequence -...	3	2	0.03	0.00028
<b>15</b>	GO:0008391	arachidonic acid monooxygenase activity	35	4	0.34	0.00035

BP

<b>GO.ID</b>	<b>Term</b>	<b>Annotated</b>	<b>Significant</b>	<b>Expected</b>	<b>classicFisher</b>
--------------	-------------	------------------	--------------------	-----------------	----------------------

<b>1</b>	GO:0036100	leukotriene catabolic process	8	4	0.08	6.5e-07
<b>2</b>	GO:0036101	leukotriene B4 catabolic process	8	4	0.08	6.5e-07



3	GO:0036102	leukotriene B4 metabolic process	8	4	0.08	6.5e-07
4	GO:0042361	menaquinone catabolic process	8	4	0.08	6.5e-07
5	GO:0042374	phylloquinone metabolic process	8	4	0.08	6.5e-07
6	GO:0042376	phylloquinone catabolic process	8	4	0.08	6.5e-07
7	GO:0042377	vitamin K catabolic process	8	4	0.08	6.5e-07
8	GO:1901662	quinone catabolic process	8	4	0.08	6.5e-07
9	GO:0042373	vitamin K metabolic process	12	4	0.12	4.4e-06

<b>10</b>	GO:0006691	leukotriene metabolic process	47	6	0.47	6.9e-06
<b>11</b>	GO:0007597	blood coagulation, intrinsic pathway	14	4	0.14	8.8e-06
<b>12</b>	GO:0009233	menaquinone metabolic process	15	4	0.15	1.2e-05
<b>13</b>	GO:0042363	fat-soluble vitamin catabolic process	16	4	0.16	1.6e-05
<b>14</b>	GO:0046039	GTP metabolic process	33	5	0.33	1.8e-05
<b>15</b>	GO:0042758	long-chain fatty acid catabolic process	17	4	0.17	2.1e-05

Paraescarpia  
MF

GO.ID	Term	Annotated	Significant	Expected	
-------	------	-----------	-------------	----------	--

1	GO:0003947	(N-acetylneuraminyl)-galactosylglucosylc...	15	15	0.85	1.7e-19
2	GO:0008528	G protein-coupled peptide receptor activ...	292	55	16.47	1.8e-15
3	GO:0001653	peptide receptor activity	334	57	18.84	5.0e-14
4	GO:0001517	N-acetylglucosamine 6-O-sulfotransferase...	48	20	2.71	3.3e-13
5	GO:0004423	iduronate-2-sulfatase activity	12	10	0.68	1.9e-11
6	GO:0004	G protein-	860	94	48.50	4.1e-10

	930	coupled receptor activity				
<b>7</b>	GO:0045 130	keratan sulfotran sferase activity	25	12	1.41	2.6e-09
<b>8</b>	GO:0008 188	neuropep tide receptor activity	181	29	10.21	3.4e-07
<b>9</b>	GO:0050 513	glycoprot ein 2-beta- D-xylosyl transfera se...	6	5	0.34	3.2e-06
<b>10</b>	GO:0008 376	acetylgal actosami nyltransf erase activity	85	17	4.79	4.4e-06
<b>11</b>	GO:0086 038	calcium:s odium antiporte r activity invol...	4	4	0.23	1.0e-05
<b>12</b>	GO:0030 551	cyclic nucleotid e binding	66	14	3.72	1.5e-05

<b>13</b>	GO:0000026	alpha-1,2-mannosyl transferase activity	12	6	0.68	2.2e-05
<b>14</b>	GO:0035438	cyclic-di-GMP binding	8	5	0.45	2.7e-05
<b>15</b>	GO:0046556	alpha-L-arabinofuranosidase activity	8	5	0.45	2.7e-05

BP

<b>GO.ID</b>	<b>Term</b>	<b>Annotated</b>	<b>Significant</b>	<b>Expected</b>	<b>classicFisher</b>
--------------	-------------	------------------	--------------------	-----------------	----------------------

<b>1</b>	GO:0001573	ganglioside metabolic process	35	18	2.11	1.6e-13
<b>2</b>	GO:0030259	lipid glycosylation	24	15	1.44	3.5e-13
<b>3</b>	GO:0001574	ganglioside	25	15	1.50	8.3e-13

		biosynthetic process				
4	GO:0042339	keratan sulfate metabolic process	51	19	3.07	4.2e-11
5	GO:0019915	lipid storage	70	22	4.21	6.1e-11
6	GO:0006044	N-acetylglucosamine metabolic process	58	20	3.49	6.8e-11
7	GO:0007621	negative regulation of female receptivity ...	22	11	1.32	1.4e-08
8	GO:0030203	glycosaminoglycan metabolic process	279	41	16.79	1.0e-07
9	GO:0030148	sphingolipid biosynthetic process	155	28	9.33	1.5e-07

<b>10</b>	GO:0045924	regulation of female receptivity	35	12	2.11	4.9e-07
<b>11</b>	GO:0060180	female mating behavior	35	12	2.11	4.9e-07
<b>12</b>	GO:0045434	negative regulation of female receptivity	19	9	1.14	5.3e-07
<b>13</b>	GO:0046008	regulation of female receptivity, post-m...	19	9	1.14	5.3e-07
<b>14</b>	GO:0006040	amino sugar metabolic process	137	25	8.25	5.8e-07
<b>15</b>	GO:1903510	mucopolysaccharide metabolic process	205	32	12.34	6.6e-07

Lamellibrachia  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

1	GO:0097602	cullin family protein binding	19	6	0.19	2.0e-08
2	GO:0003677	DNA binding	2064	38	20.18	8.3e-05
3	GO:0019961	interferon binding	3	2	0.03	0.00028
4	GO:0019964	interferon-gamma binding	3	2	0.03	0.00028
5	GO:0015280	ligand-gated sodium channel activity	34	4	0.33	0.00033
6	GO:0008201	heparin binding	179	8	1.75	0.00038
7	GO:0003676	nucleic acid binding	3209	49	31.37	0.00066



<b>8</b>	GO:0042 393	histone binding	162	7	1.58	0.00108
<b>9</b>	GO:0004 867	serine- type endopept idase inhibitor acti...	163	7	1.59	0.00112
<b>10</b>	GO:0008 526	phosphat idylinosit ol transfer activity	6	2	0.06	0.00139
<b>11</b>	GO:0035 064	methylat ed histone binding	52	4	0.51	0.00166
<b>12</b>	GO:0140 034	methyлатi on- depende nt protein binding	52	4	0.51	0.00166
<b>13</b>	GO:0099 094	ligand- gated cation channel activity	227	8	2.22	0.00179
<b>14</b>	GO:0015 252	proton channel activity	25	3	0.24	0.00180

<b>15</b>	GO:0043169	cation binding	6146	79	60.08	0.00197
-----------	------------	----------------	------	----	-------	---------

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0042395	ecdysis, collagen and cuticulin-based cu...	4	3	0.04	4.3e-06
<b>2</b>	GO:0007596	blood coagulation	245	11	2.52	4.9e-05
<b>3</b>	GO:0007599	hemostasis	247	11	2.54	5.2e-05
<b>4</b>	GO:0050817	coagulation	261	11	2.69	8.6e-05
<b>5</b>	GO:1990390	protein K33-linked ubiquitination	3	2	0.03	0.00031
<b>6</b>	GO:0050	regulation	369	12	3.80	0.00046

	878	n of body fluid levels				
<b>7</b>	GO:0070447	positive regulation of oligodendrocyte p...	4	2	0.04	0.00062
<b>8</b>	GO:0007006	mitochondrial membrane organization	69	5	0.71	0.00072
<b>9</b>	GO:0097345	mitochondrial outer membrane permeabiliz...	18	3	0.19	0.00078
<b>10</b>	GO:1902108	regulation of mitochondrial membrane per...	18	3	0.19	0.00078
<b>11</b>	GO:1902110	positive regulation of	18	3	0.19	0.00078

		mitochon drial mem...				
<b>12</b>	GO:0060669	embryoni c placenta morphog enesis	19	3	0.20	0.00092
<b>13</b>	GO:1902686	mitochon drial outer membran e permeabi liz...	19	3	0.20	0.00092
<b>14</b>	GO:0048844	artery morphog enesis	73	5	0.75	0.00094
<b>15</b>	GO:0001892	embryoni c placenta developm ent	76	5	0.78	0.00112

Osedax  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
					er

<b>1</b>	GO:0030158	protein xylosyltransferase activity	24	19	2.32	1.3e-15
<b>2</b>	GO:0030151	molybdenum ion binding	22	17	2.13	8.6e-14
<b>3</b>	GO:0140096	catalytic activity, acting on a protein	2675	367	258.55	1.2e-13
<b>4</b>	GO:0043546	molybdopterin cofactor binding	28	19	2.71	1.4e-13
<b>5</b>	GO:0004842	ubiquitin-protein transferase activity	473	94	45.72	7.6e-12
<b>6</b>	GO:0019787	ubiquitin-like protein transferase activ...	485	95	46.88	1.4e-11
<b>7</b>	GO:0003824	catalytic activity	8502	947	821.77	1.9e-10

<b>8</b>	GO:0004022	alcohol dehydrogenase (NAD+) activity	13	11	1.26	4.3e-10
<b>9</b>	GO:0016616	oxidoreductase activity, acting on the C...	199	49	19.23	5.7e-10
<b>10</b>	GO:0018455	alcohol dehydrogenase [NAD(P)+] activity	14	11	1.35	1.8e-09
<b>11</b>	GO:0045499	chemorepellent activity	14	11	1.35	1.8e-09
<b>12</b>	GO:0004423	iduronate-2-sulfatase activity	12	10	1.16	3.8e-09
<b>13</b>	GO:0001517	N-acetylglucosamine 6-O-sulfotransferase...	48	20	4.64	5.3e-09
<b>14</b>	GO:0008	5'-nucleotid	16	11	1.55	1.8e-08

	253	ase activity				
<b>15</b>	GO:0008502	melatoni n receptor activity	30	15	2.90	2.2e-08

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0019321	pentose metaboli c process	44	26	4.34	1.1e-15
<b>2</b>	GO:1903510	mucopoly saccharid e metaboli c process	205	60	20.24	4.4e-15
<b>3</b>	GO:0042732	D-xylose metaboli c process	30	20	2.96	7.9e-14
<b>4</b>	GO:0030203	glycosam inoglycan metaboli c process	279	70	27.54	1.2e-13

5	GO:0072523	purine-containing compound catabolic pro...	71	31	7.01	1.3e-13
6	GO:0016567	protein ubiquitination	719	133	70.97	4.4e-13
7	GO:0070647	protein modification by small protein co...	858	151	84.70	6.7e-13
8	GO:0032446	protein modification by small protein co...	768	139	75.81	6.9e-13
9	GO:0006807	nitrogen compound metabolic process	8960	1021	884.46	4.0e-12
10	GO:1901565	organonitrogen compound catabolic proces...	1422	219	140.37	5.0e-12



<b>11</b>	GO:0006281	DNA repair	485	95	47.88	4.4e-11
<b>12</b>	GO:0006022	aminoglycan metabolic process	344	73	33.96	2.0e-10
<b>13</b>	GO:0071704	organic substance metabolic process	10758	1183	1061.95	2.2e-10
<b>14</b>	GO:0044257	cellular protein catabolic process	715	124	70.58	2.3e-10
<b>15</b>	GO:0051603	proteolysis involved in cellular protein...	678	119	66.93	2.5e-10

Siboglinidae  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

--	--	--	--	--	--	--

1	GO:0016491	oxidoreductase activity	1424	193	70.70	1e-30
2	GO:0016705	oxidoreductase activity, acting on paired...	416	84	20.65	5.4e-29
3	GO:0004497	monooxygenase activity	318	67	15.79	2.2e-24
4	GO:0052689	carboxylic ester hydrolase activity	240	53	11.91	1.7e-20
5	GO:0005506	iron ion binding	422	72	20.95	1.8e-20
6	GO:0016795	phosphoric triester hydrolase activity	19	16	0.94	1.0e-18
7	GO:0047862	diisopropyl-fluorophosphatase activity	19	16	0.94	1.0e-18

<b>8</b>	GO:0008131	primary amine oxidase activity	17	14	0.84	3.0e-16
<b>9</b>	GO:0051213	dioxygenase activity	193	41	9.58	1.6e-15
<b>10</b>	GO:0004508	steroid 17-alpha-monooxygenase activity	29	17	1.44	1.7e-15
<b>11</b>	GO:0047442	17-alpha-hydroxypregesterone aldolase ac...	29	17	1.44	1.7e-15
<b>12</b>	GO:0009881	photoreceptor activity	56	22	2.78	6.8e-15
<b>13</b>	GO:0004930	G protein-coupled receptor activity	860	98	42.70	7.1e-15
<b>14</b>	GO:0001872	(1->3)-beta-D-glucan binding	17	13	0.84	2.0e-14

<b>15</b>	GO:0003824	catalytic activity	8502	532	422.09	3.7e-14
-----------	------------	--------------------	------	-----	--------	---------

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0009308	amine metabolic process	166	41	7.40	1.1e-19
<b>2</b>	GO:0016042	lipid catabolic process	375	59	16.71	1.9e-17
<b>3</b>	GO:0006629	lipid metabolic process	1649	146	73.49	2.2e-16
<b>4</b>	GO:0018298	protein-chromophore linkage	53	21	2.36	2.7e-15
<b>5</b>	GO:0019752	carboxylic acid metabolic process	1255	118	55.93	4.1e-15
<b>6</b>	GO:0044242	cellular lipid catabolic process	256	44	11.41	9.9e-15

<b>7</b>	GO:0006631	fatty acid metabolic process	502	64	22.37	2.7e-14
<b>8</b>	GO:0048149	behavioral response to ethanol	45	18	2.01	2.2e-13
<b>9</b>	GO:0032787	monocarboxylic acid metabolic process	785	82	34.99	4.4e-13
<b>10</b>	GO:0015881	creatine transmembrane transport	38	16	1.69	1.9e-12
<b>11</b>	GO:0044281	small molecule metabolic process	2573	185	114.68	6.9e-12
<b>12</b>	GO:0007567	parturition	13	10	0.58	7.4e-12
<b>13</b>	GO:0006694	steroid biosynthetic process	169	31	7.53	1.6e-11
<b>14</b>	GO:0043436	oxoacid metabolic process	1505	123	67.08	2.0e-11

<b>15</b>	GO:0006082	organic acid metabolic process	1545	125	68.86	2.6e-11
-----------	------------	--------------------------------	------	-----	-------	---------

Vestimentifera  
MF

<b>GO.ID</b>	<b>Term</b>	<b>Annotated</b>	<b>Significant</b>	<b>Expected</b>	<b>classicFisher</b>
--------------	-------------	------------------	--------------------	-----------------	----------------------

<b>1</b>	GO:0001225	RNA polymerase II transcription coactiva...	4	3	0.01	8.3e-08
<b>2</b>	GO:0001224	RNA polymerase II transcription cofactor...	6	3	0.02	4.1e-07
<b>3</b>	GO:0000981	DNA-binding transcription factor activit...	504	10	1.41	1.1e-06
<b>4</b>	GO:0140110	transcription	1100	14	3.08	1.2e-06

		regulator activity				
<b>5</b>	GO:0001 228	DNA- binding transcript ion activator acti...	220	7	0.62	2.5e-06
<b>6</b>	GO:0003 700	DNA- binding transcript ion factor activit...	838	12	2.35	2.5e-06
<b>7</b>	GO:0001 216	DNA- binding transcript ion activator acti...	222	7	0.62	2.7e-06
<b>8</b>	GO:0000 976	transcript ion regulator y region sequence ...	660	10	1.85	1.2e-05
<b>9</b>	GO:0001 067	regulator y region nucleic acid binding	668	10	1.87	1.4e-05
<b>10</b>	GO:1990	sequence	690	10	1.93	1.8e-05

	837	-specific double- stranded DNA bi...				
<b>11</b>	GO:0001 223	transcript ion coactivat or binding	21	3	0.06	2.7e-05
<b>12</b>	GO:0003 690	double- stranded DNA binding	743	10	2.08	3.4e-05
<b>13</b>	GO:0043 565	sequence -specific DNA binding	950	11	2.66	5.1e-05
<b>14</b>	GO:0001 221	transcript ion cofactor binding	29	3	0.08	7.2e-05
<b>15</b>	GO:0000 978	RNA polymera se II cis- regulator y region ...	390	7	1.09	1e-04

BP



GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

1	GO:0048856	anatomical structure development	5880	38	16.79	1.4e-09
2	GO:0098742	cell-cell adhesion via plasma-membrane a...	268	10	0.77	3.6e-09
3	GO:0050793	regulation of developmental process	2151	23	6.14	4.3e-09
4	GO:0007275	multicellular organism development	5450	36	15.56	4.3e-09
5	GO:0032502	developmental process	6282	38	17.94	1.1e-08

<b>6</b>	GO:2000026	regulation of multicellular organismal d...	1712	20	4.89	1.6e-08
<b>7</b>	GO:1901262	negative regulation of sorocarp spore ce...	3	3	0.01	2.2e-08
<b>8</b>	GO:0030154	cell differentiation	3781	29	10.80	2.7e-08
<b>9</b>	GO:2000027	regulation of animal organ morphogenesis	175	8	0.50	3.2e-08
<b>10</b>	GO:0048869	cellular developmental process	3851	29	11.00	4.2e-08
<b>11</b>	GO:0044671	sorocarp spore cell differentiation	14	4	0.04	5.8e-08
<b>12</b>	GO:0048837	sorocarp sorus developm	14	4	0.04	5.8e-08

		ent				
<b>13</b>	GO:0098609	cell-cell adhesion	716	13	2.04	7.2e-08
<b>14</b>	GO:0048468	cell development	2301	22	6.57	8.7e-08
<b>15</b>	GO:0031153	slug development involved in sorocarp de...	4	3	0.01	8.8e-08

Vestimentifera\_cl1  
MF

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0005041	low-density lipoprotein particle recepto...	17	5	0.04	3.5e-10
<b>2</b>	GO:0030228	lipoprotein particle receptor activity	22	5	0.05	1.5e-09

3	GO:0030169	low-density lipoprotein particle binding	25	5	0.06	3.0e-09
4	GO:0071813	lipoprotein particle binding	33	5	0.08	1.3e-08
5	GO:0071814	protein-lipid complex binding	33	5	0.08	1.3e-08
6	GO:0005540	hyaluronic acid binding	57	5	0.13	2.2e-07
7	GO:0015035	protein disulfide oxidoreductase activity...	28	4	0.07	5.3e-07
8	GO:0015036	disulfide oxidoreductase activity	49	4	0.12	5.3e-06
9	GO:0005044	scavenger receptor activity	123	5	0.29	1.0e-05
10	GO:0016667	oxidoreductase activity,	64	4	0.15	1.5e-05

		acting on a sul...				
<b>11</b>	GO:0005 539	glycosam inoglycan binding	286	6	0.68	5.3e-05
<b>12</b>	GO:0038 024	cargo receptor activity	189	5	0.45	8.0e-05
<b>13</b>	GO:0005 509	calcium ion binding	1442	12	3.41	8.8e-05
<b>14</b>	GO:0097 367	carbohyd rate derivative binding	2674	15	6.31	0.00078
<b>15</b>	GO:0043 168	anion binding	3427	17	8.09	0.00126

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

<b>1</b>	GO:0007 155	cell adhesion	1675	17	4.31	3.8e-07
<b>2</b>	GO:0022	biological adhesion	1690	17	4.35	4.4e-07

	610					
3	GO:0001944	vasculature development	750	11	1.93	2.3e-06
4	GO:0016525	negative regulation of angiogenesis	104	5	0.27	6.9e-06
5	GO:2000181	negative regulation of blood vessel morph...	107	5	0.28	7.9e-06
6	GO:1901343	negative regulation of vasculature devel...	115	5	0.30	1.1e-05
7	GO:0003017	lymph circulation	58	4	0.15	1.5e-05
8	GO:0048014	Tie signaling pathway	68	4	0.18	2.8e-05
9	GO:0072359	circulatory system development	1170	12	3.01	2.9e-05

<b>10</b>	GO:0036303	lymph vessel morphogenesis	72	4	0.19	3.5e-05
<b>11</b>	GO:0001945	lymph vessel development	73	4	0.19	3.7e-05
<b>12</b>	GO:0009653	anatomical structure morphogenesis	2975	19	7.66	6.5e-05
<b>13</b>	GO:0042742	defense response to bacterium	235	5	0.61	0.00033
<b>14</b>	GO:0045765	regulation of angiogenesis	237	5	0.61	0.00035
<b>15</b>	GO:0120193	tight junction organization	139	4	0.36	0.00044

Vestimentifera\_cl2  
MF

GO.ID	Term	Annotated	Significant	Expected	
-------	------	-----------	-------------	----------	--

					<b>classicFisher</b>
--	--	--	--	--	----------------------

<b>1</b>	GO:0031433	telethonin binding	33	9	0.08	5.4e-17
<b>2</b>	GO:0051371	muscle alpha-actinin binding	50	9	0.12	3.4e-15
<b>3</b>	GO:0008307	structural constituent of muscle	61	9	0.15	2.3e-14
<b>4</b>	GO:0051393	alpha-actinin binding	67	9	0.17	5.6e-14
<b>5</b>	GO:0043621	protein self-association	69	9	0.17	7.5e-14
<b>6</b>	GO:0042805	actinin binding	71	9	0.18	9.8e-14
<b>7</b>	GO:0097493	structural molecule activity conferring ...	20	6	0.05	6.1e-12
<b>8</b>	GO:0004713	protein tyrosine kinase	169	10	0.42	8.7e-12



		activity				
<b>9</b>	GO:0002020	protease binding	131	9	0.32	2.8e-11
<b>10</b>	GO:0051015	actin filament binding	196	9	0.48	1.0e-09
<b>11</b>	GO:0005516	calmodulin binding	259	9	0.64	1.2e-08
<b>12</b>	GO:0005044	scavenger receptor activity	123	7	0.30	2.0e-08
<b>13</b>	GO:0005198	structural molecule activity	626	12	1.55	2.5e-08
<b>14</b>	GO:0038024	cargo receptor activity	189	7	0.47	3.8e-07
<b>15</b>	GO:0003779	actin binding	401	9	0.99	4.9e-07

BP

GO.ID	Term	Annotated	Significant	Expected	classicFisher
-------	------	-----------	-------------	----------	---------------

1	GO:0048	sarcomer	31	9	0.08	5.2e-17

	769	ogenesis				
2	GO:0030241	skeletal muscle myosin thick filament as...	32	9	0.08	7.2e-17
3	GO:0035995	detection of muscle stretch	32	9	0.08	7.2e-17
4	GO:0048739	cardiac muscle fiber development	33	9	0.09	9.9e-17
5	GO:0030240	skeletal muscle thin filament assembly	36	9	0.09	2.4e-16
6	GO:0014866	skeletal myofibril assembly	38	9	0.10	4.1e-16
7	GO:0071688	striated muscle myosin thick filament as...	38	9	0.10	4.1e-16
8	GO:0035994	response to muscle stretch	39	9	0.10	5.4e-16
9	GO:0055	cardiac myofibril	39	9	0.10	5.4e-16

	003	assembly				
<b>10</b>	GO:0031034	myosin filament assembly	44	9	0.12	1.8e-15
<b>11</b>	GO:0031033	myosin filament organization	45	9	0.12	2.2e-15
<b>12</b>	GO:0010737	protein kinase A signaling	87	10	0.23	2.0e-14
<b>13</b>	GO:0055008	cardiac muscle tissue morphogenesis	67	9	0.18	1.0e-13
<b>14</b>	GO:0055013	cardiac muscle cell development	72	9	0.19	2.0e-13
<b>15</b>	GO:0060415	muscle tissue morphogenesis	74	9	0.19	2.6e-13