

SSD1 motifs in Aspergillus Nidulans & similarities with Aspergillus Fumigatus

Tristan

04/07/2023

Summary

In this doc we look for the *CNYTCNYT* Ssd-1 binding motif in Aspergillus Nidulans. We also look for similarities between those genes with significant motifs and the genes with significant motifs in Aspergillus Fumigatus.

Aspergillus Nidulans motif analysis

Load 1000nt UTRs from Aspergillus Nidulans

```
#load data
# define fasta filename
As_up1000_file <- paste(here::here(),
                        "motif_conservation",
                        "data",
                        "Asp_Ni.fasta",
                        sep= "/")

# load fasta file as DNA string set
As_up1000_DSS <- readDNASTringSet(As_up1000_file)

# remove truncated sequences, which we don't need.
As_up1000_DSS <- As_up1000_DSS[width(As_up1000_DSS) == 1000]

# assign just the ORF id to the name, but keep all the info.
As_namesinfo <-
  tibble(everything = names(As_up1000_DSS)) %>%
  tidyr::separate(everything,into = c("id","strain","info","seqtype","seq","length"), sep = " \\| ") %>%
  dplyr::select(id, info,seq)
```

Count occurrences of *CNYTCNYT* motif in Aspergillus Nidulans

```
CNYTCNYT_count_df <- tibble(id = As_namesinfo$id,
                             info = As_namesinfo$info,
                             count_up1000 = vcountPattern(pattern = DNASTring("CNYTCNYT"),
                                                             subject = As_up1000_DSS,
```

```

        fixed = "subject"),
count_up200 = vcountPattern(pattern = DNASTring("CNYTCNYT"),
  subject = subseq(As_up1000_DSS,start = 801L, end = 1000L),
  fixed = "subject"),
count_up100 = vcountPattern(pattern = DNASTring("CNYTCNYT"),
  subject = subseq(As_up1000_DSS,start = 901L, end = 1000L),
  fixed = "subject")
)
arrange(CNYTCNYT_count_df,desc(count_up100)) %>%
  head(n = 20) %>%
  knitr::kable()

```

id	info	count_up1000	count_up200	count_up100
AN8881	Predicted siderophore transporter	12	8	8
AN5241	Endo-beta-1,4-glucanase B	9	8	8
AN4882	Putative pectin lyase involved in degradation of pectin	6	6	6
AN7229	Transcript induced in response to calcium dichloride in a CrzA-dependent manner	8	6	6
AN8319	Protein of unknown function	6	6	6
AN5602	Ortholog(s) have role in peroxisome organization and integral component of peroxisomal membrane, plasma membrane localization	6	6	6
AN1602	Putative beta-1,4-endoglucanase	9	7	6
AN8327	Protein with polygalacturonase activity, involved in degradation of pectin	7	6	6
AN6656	Putative endopolygalacturonase	7	5	5
AN4821	Has domain(s) with predicted DNA binding, zinc ion binding activity, role in transcription, DNA-templated and nucleus localization	7	5	5
AN7501	Chitinase 2	5	5	5
AN1106	Ortholog(s) have mitochondrion localization	7	5	5
AN4438	Ortholog of A. nidulans FGSC A4 : AN4246, AN8484, A. fumigatus Af293 : Afu1g06480, Afu2g05150/mp2, A. niger CBS 513.88 : An18g05510 and A. oryzae RIB40 : AO090001000450	11	10	5
AN1088	Ortholog(s) have proteasome binding activity, role in cellular response to arsenic-containing substance, proteasome-mediated ubiquitin-dependent protein catabolic process and cytoplasm, nucleus localization	5	5	5
AN0938	Dual specificity protein kinase, putative	7	5	5
AN5931	Putative ATP-dependent RNA helicase	10	10	5
AN7873	Putative multifunctional enzyme with a predicted role in cytosolic fatty acid formation	7	6	5
AN6286	Ortholog(s) have cytoplasm localization	5	5	5
AN1081	Has domain(s) with predicted oxidoreductase activity and role in metabolic process	4	4	4
AN4668	Putative mitogen activated protein kinase (MAPK)	5	4	4

This is a table of the genes with the highest counts in the 1000nt UTRs, 200nt UTRs and 100nt UTRs, sorted by the count in the 100nt UTRs.

100nt upstream

```

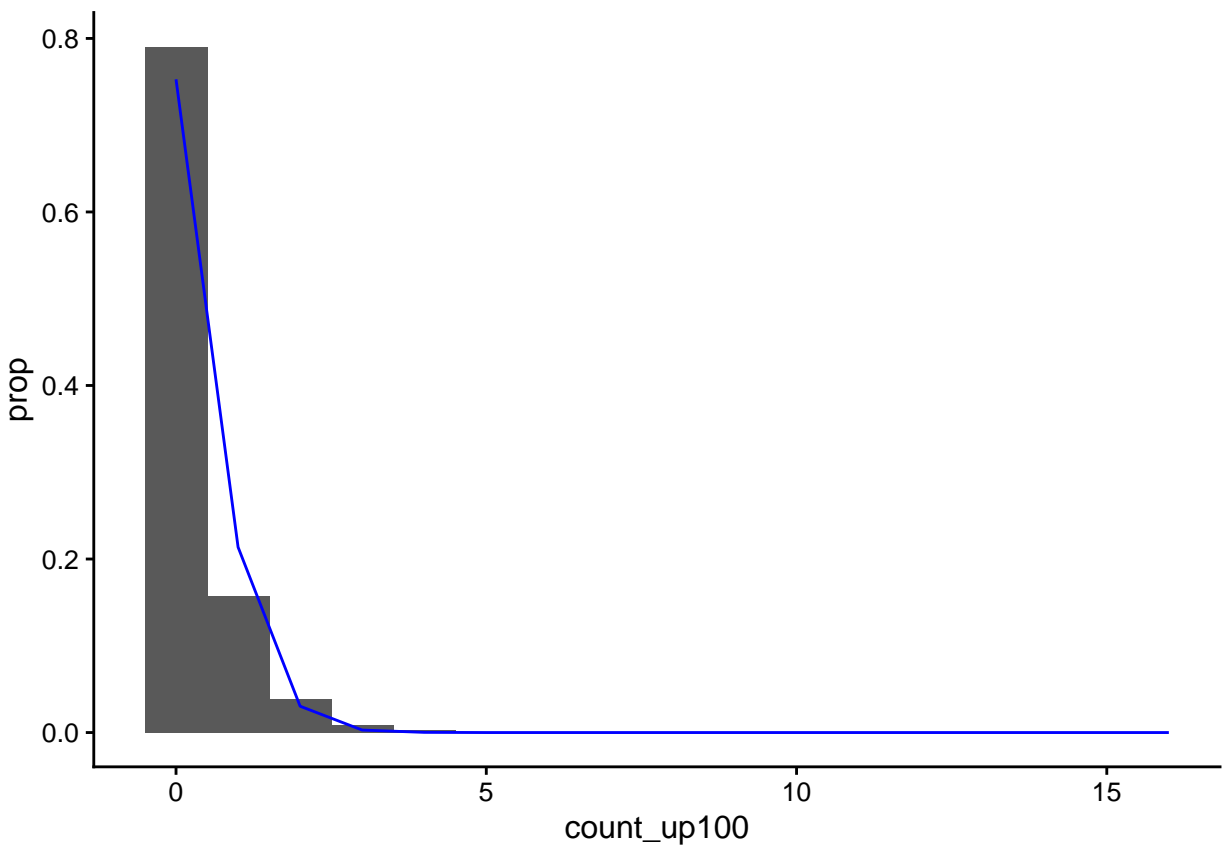
# plots of counts to 100bp upstream
CNYT_up100_reducedmean <- CNYTCNYT_count_df %>%

```

```

pull(count_up100) %>%
mean()
CNYT_up100_reducedvar <- CNYTCNYT_count_df %>%
pull(count_up100) %>%
var()
ggplot(data = CNYTCNYT_count_df, aes(x = count_up100)) +
  geom_bar(width = 1, aes(y = after_stat(prop)), stat = "count", group = 1) +
  geom_line(data = tibble(),
            aes(x = 0:20, y = dpois(0:20, lambda=CNYT_up100_reducedmean)),
            colour = "blue") +
  scale_x_continuous(limits = c(-0.5,16), oob = scales::squish)

```



Comparison of Aspergillus Nidulans and Fumigatus GO terms

Following this, I did a GO term search on fungidb using the genelists of genes with ≥ 2 *CNYTCNYT* motifs in the 100nt UTRs of Aspergillus Nidulans and Aspergillus Fumigatus.

```

dat_file <- paste(here::here(),
                  "motif_conservation",
                  "data",
                  sep= "/")

#load data

```

```
ASP_N_GO <- read_csv(paste(dat_file, "ASP_N.csv", sep= "/")) %>%
  mutate(From = "Aspergillus Nidulans")
```

```
## Rows: 12 Columns: 11
## -- Column specification -----
## Delimiter: ","
## chr (3): ID, Name, Result gene list
## dbl (8): Bgd count, Result count, Pct of bgd, Fold enrichment, Odds ratio, P...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
ASP_F_GO <- read_csv(paste(dat_file, "ASP_F.csv", sep= "/")) %>%
  mutate(From = "Aspergillus Fumigatus")
```

```
## Rows: 11 Columns: 11
## -- Column specification -----
## Delimiter: ","
## chr (3): ID, Name, Result gene list
## dbl (8): Bgd count, Result count, Pct of bgd, Fold enrichment, Odds ratio, P...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

After loading in both GO Enrichment files, you can see that *Aspergillus Nidulans* has 12 GO terms with a p-value of < 0.05 , whereas *Aspergillus Fumigatus* has 11 GO terms with a p-value of < 0.05 . They have an overlap of 9 GO terms.

```
#find overlap in GO names
```

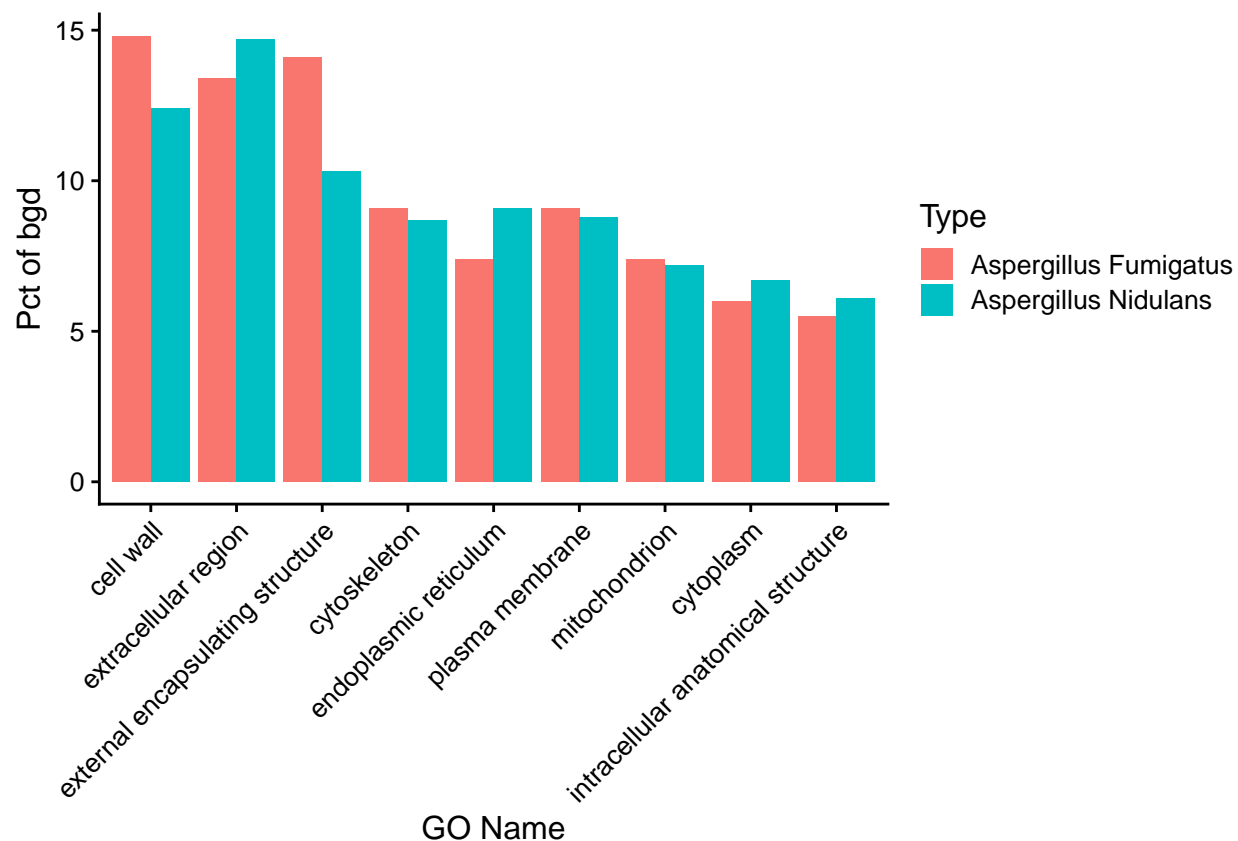
```
overlap <- intersect(ASP_N_GO$Name, ASP_F_GO$Name)
```

```
#plot bar chart of "Pct of bgd" for each GO name in overlap from both ASP_N_GO and ASP_F_GO
```

```
combined <- rbind(ASP_F_GO, ASP_N_GO)
```

```
combined <- combined[combined$Name %in% overlap,]
```

```
ggplot(combined, aes(x = reorder(Name, -`Pct of bgd`, FUN = min), y = `Pct of bgd`, group = From, fill =
  geom_bar(stat = "identity", position = "dodge") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(x = "GO Name", y = "Pct of bgd", fill = "Type")
```



This shows that the GO terms of the targeted genes in *Aspergillus Nidulans* and *Aspergillus Fumigatus* are similar, with cell wall having lower representation in *Aspergillus Nidulans* but extracellular region having higher representation.

```
sessionInfo()
```

```
## R version 4.2.3 (2023-03-15)
## Platform: aarch64-apple-darwin20.0.0 (64-bit)
## Running under: macOS Ventura 13.4
##
## Matrix products: default
## BLAS: /Users/tristansones-dykes/mambaforge/envs/R/lib/R/lib/libRblas.dylib
## LAPACK: /Users/tristansones-dykes/mambaforge/envs/R/lib/R/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats4 stats graphics grDevices utils datasets methods
## [8] base
##
## other attached packages:
## [1] Biostrings_2.66.0 GenomeInfoDb_1.34.9 XVector_0.38.0
## [4] IRanges_2.32.0 S4Vectors_0.36.2 BiocGenerics_0.44.0
## [7] cowplot_1.1.1 lubridate_1.9.2 forcats_1.0.0
## [10] stringr_1.5.0 dplyr_1.1.2 purrr_1.0.1
```

```

## [13] readr_2.1.4          tidyr_1.3.0          tibble_3.2.1
## [16] ggplot2_3.4.2        tidyverse_2.0.0     here_1.0.1
##
## loaded via a namespace (and not attached):
## [1] tidyselect_1.2.0      xfun_0.39            colorspace_2.1-0
## [4] vctrs_0.6.3          generics_0.1.3       htmltools_0.5.5
## [7] yaml_2.3.7           utf8_1.2.3           rlang_1.1.1
## [10] pillar_1.9.0         glue_1.6.2           withr_2.5.0
## [13] bit64_4.0.5          GenomeInfoDbData_1.2.9 lifecycle_1.0.3
## [16] zlibbioc_1.44.0      munsell_0.5.0        gtable_0.3.3
## [19] evaluate_0.21        labeling_0.4.2       knitr_1.43
## [22] tzdb_0.4.0           fastmap_1.1.1        parallel_4.2.3
## [25] fansi_1.0.4          highr_0.10           scales_1.2.1
## [28] vroom_1.6.3          bit_4.0.5            farver_2.1.1
## [31] hms_1.1.3            digest_0.6.31        stringi_1.7.12
## [34] grid_4.2.3           rprojroot_2.0.3      cli_3.6.1
## [37] tools_4.2.3          bitops_1.0-7         magrittr_2.0.3
## [40] Rcurl_1.98-1.12      crayon_1.5.2         pkgconfig_2.0.3
## [43] timechange_0.2.0     rmarkdown_2.23       rstudioapi_0.14
## [46] R6_2.5.1             compiler_4.2.3

```