

# FUNGAL COMPARATIVE GENOMICS

Jason Stajich  
Duke University  
UC Berkeley (Aug 2006)



# Evolutionary genomics

## Evolutionary & Organismal Biology

Phylogeny  
Population genetics and structure  
Phenotype  
Ecological adaptations

## Comparative Genomics

Molecular evolution  
Gene order  
Gene families  
Gene and genome structure  
Gene content  
Conserved elements  
Rates of molecular evolution  
Gene function inference

## Model Systems

Genetic tools  
Gene function & expression  
Regulatory networks  
Pathways  
Molecular & cellular biology  
Disease models

# Industrial uses of fungi

---

- Bread, beer, wine - *Saccharomyces cerevisiae*
- Sake and soy sauce - *Aspergillus oryzae*
- Dairy - *Penicillium roqueforti*, *Kluyveromyces lactis*
- Citric acid - *Aspergillus niger*
- Riboflavin - *Ashbya gossypii*
- Stonewashed jeans - *Trichoderma reesei*
- Penicillin antibiotic - *Penicillium notatum*
- Button Mushrooms - *Agaricus bisporus*

# Agricultural impact of fungi

---

- Two-thirds of plant disease is caused by fungi
  - Wheat blight (*Fusarium*)
  - Strawberry grey mold (*Botrytis*)
  - Leaf rusts (*Puccinia*)
  - Wheat and maize smuts (*Ustilago*).
  - Also deposit mycotoxins - e.g. ergot
- Mycorrhizal fungi provide nutrient exchange and nitrogen fixation



USDA



A.G. Bölker



# FUNGAL COMPARATIVE GENOMICS

- Problems
  - Many fungal genomes
  - No central place for annotations, interlinking homolog information
  - Want to visual gene structures and genome context
  - Need system for good database system for scripting genome questions



# GETTING THE DATA IN

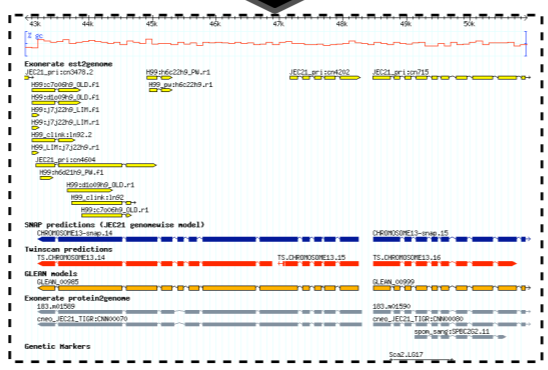
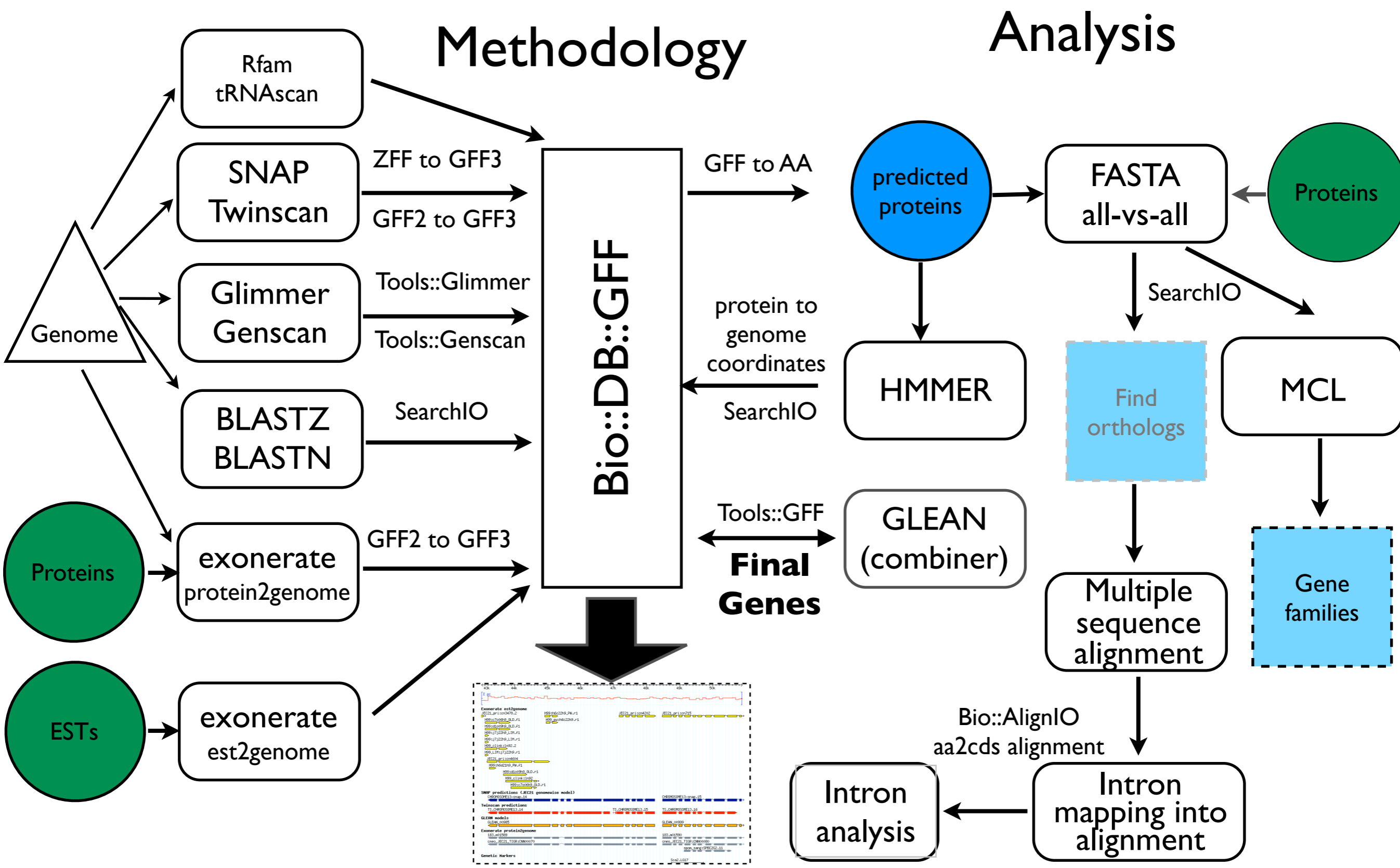
- GFF3 as the data transfer format
- Write GenBank -> GFF3 scripts
- Read in data from genome Centers (Broad, Sanger, WashU, JGI, SGD)
- Pipeline for Genome Annotation







**51 + More funded and  
in progress world-wide**





# WHAT I NEEDED

- Database for storing and querying genome annotations
  - Bio::DB::GFF (BioPerl & Gbrowse)
- Visualization - Gbrowse
- Analyses
  - Ability to query for a gene's exon-intron structure and sequences
  - Are gene families clustered on chromosome?
  - Are functional classes of genes clustered on chromosome?



# GBrowse

- Visualization of annotation data
- Does not have to be for whole / finished genomes
  - Most projects are unfinished so many contigs (100s - 1000s)
  - BLAST interface with link to Gbrowse view allows user to start with query sequence and get to the genomic location



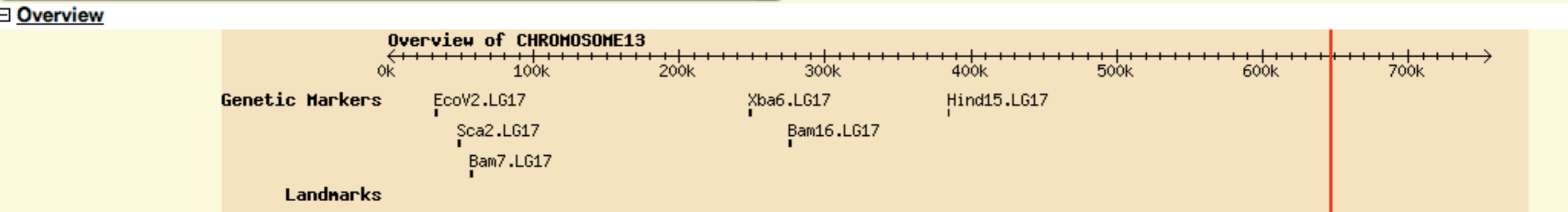
# GBrowse View

Landmark or Region: CHROMOSOME13:646044..6469; Search

Reports & Analysis: Annotate Restriction Sites Configure... Go

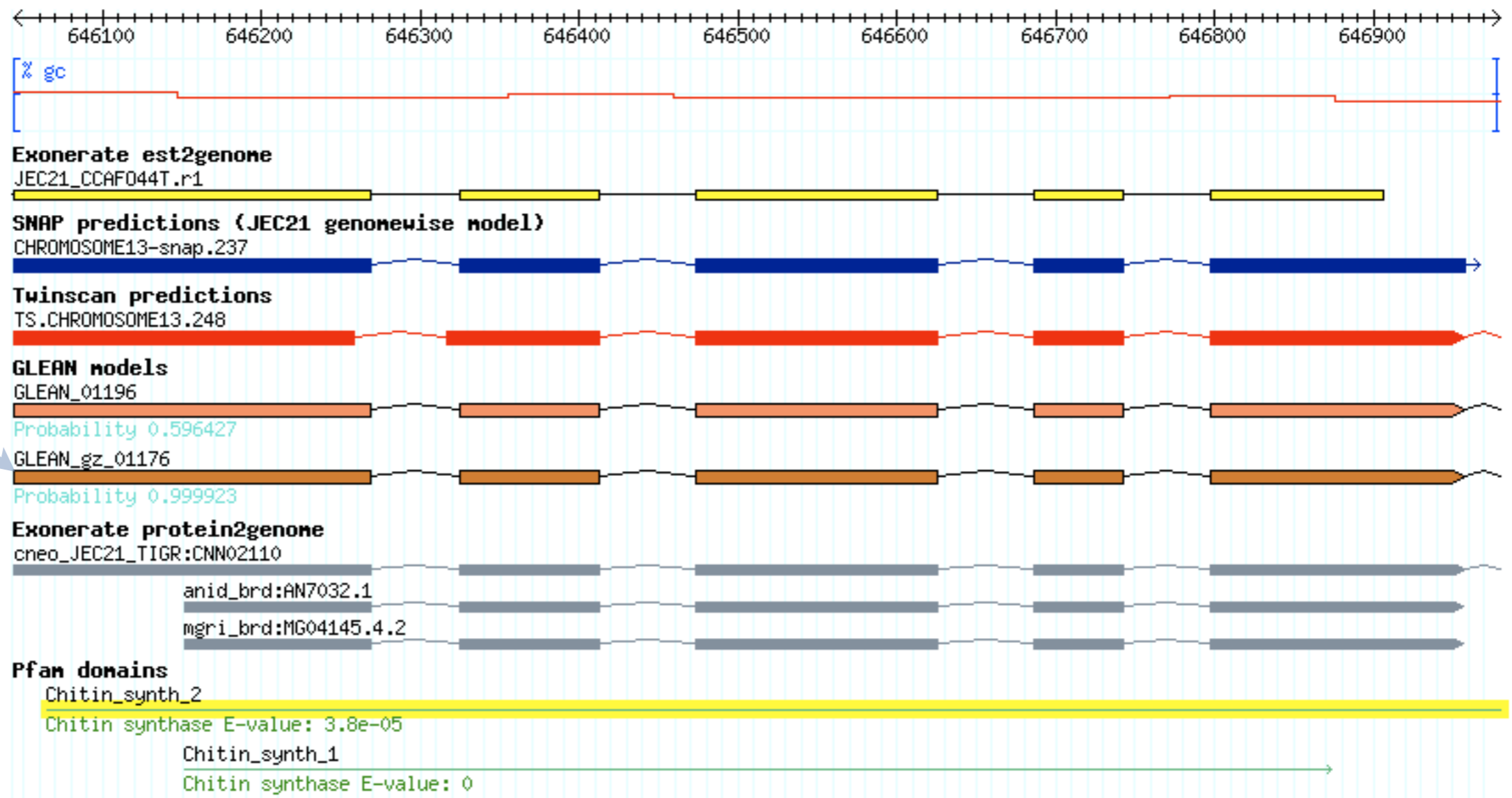
Data Source: Cryptococcus neoformans var grubii serotype A, strain H99 (Duke 2004-10-30 assembly)

Scroll/Zoom: <<< - + >>> Show 936 bp

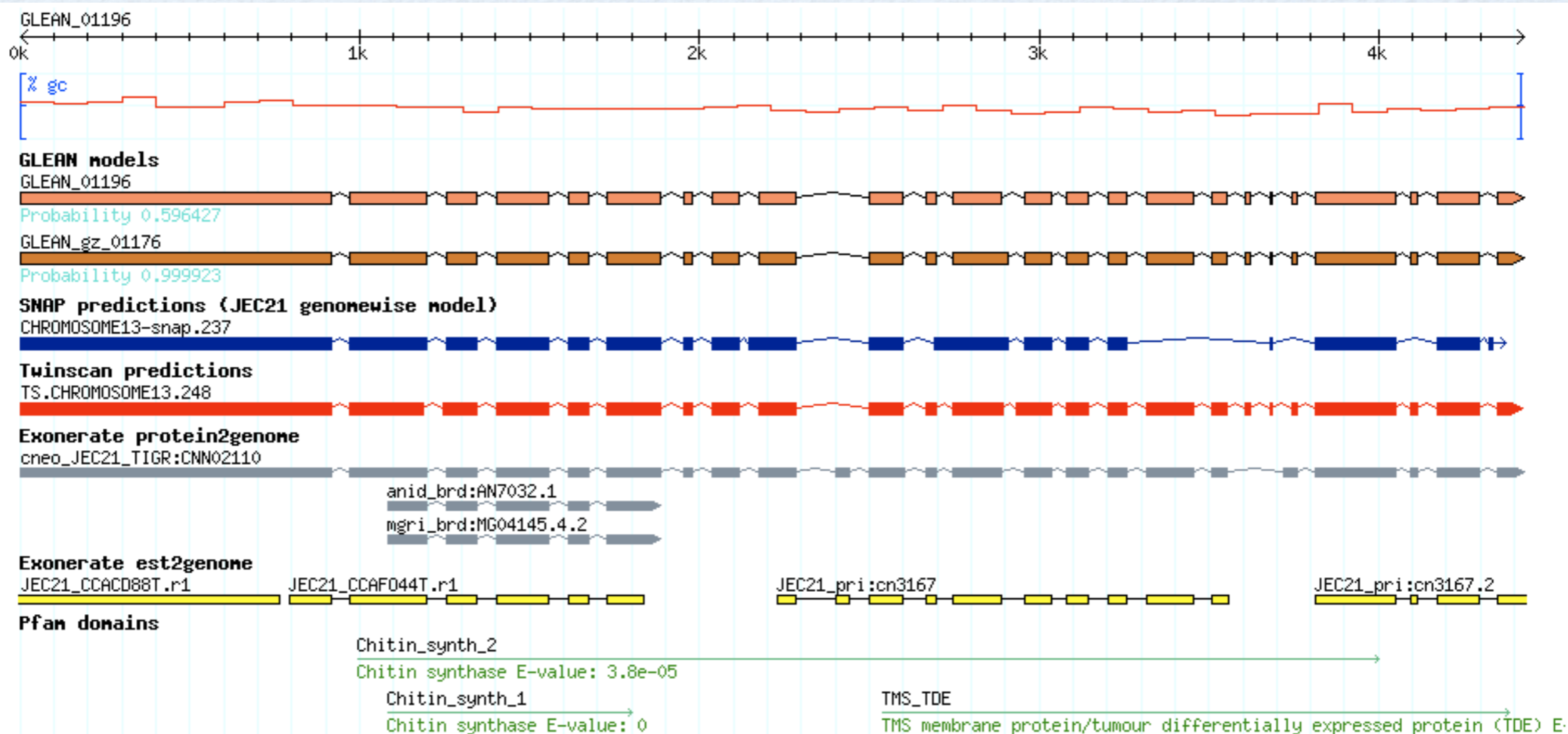


Region Details

Click Here



# GENE PAGE (1)



[Gbrowse Details](#)

## CDS Statistics

## Peptide Statistics

**Locus Length:** 4428 bp  
**CDS Length:** 2952 bp  
**Exon count:** 24 exons

**Protein Length:** 984 residues  
**Molecular Weight:** 108489.9



# GENE PAGE (2)

## Coding Sequence

```
>cneo_H99:GLEAN_01196
ATGCCAAACATATCACGCAAACCTCCTCCGCGcttctactctccttctcactcccccttcgccttcacttt
ATGCTCCCATACAATCACCCCCGGCCCCCTCTTATGACTACCACGCCAACCCAGGACGTTGAATCCATT
CTCCGACGCACGTGAAGTCGGTGGATATGCTCAACTCCAAGGGGAAGATCAAATGACTGGCGCACCTTTA
TACCAGCCTCCGTATGCTCCTCAACTACTCGTTGCTCAACCAActcctgtttcttcccgcctcccgtttt
tcGAAGCTGCGCTTGCCCGCGCGCGGATCCAAACGCCAGCTTACCAGAAGCCCCGACTCCAGCTTA
TGcccaacctttaccttctacctcccgcctcccGACCCAATCACCTGATCTATCTGTTGGTTTACT
CAGGCCAATACAGTTTCGCTACGCCATCAATCCAAGATCTCAGTTGAAGGAAGGATCCCCTTCGCCTTAC
CGTTCATGGATGACAGTTTCGTTTACAATGATGCTGCTCACCTCTATAACGTTGAGCCGGACGTCGAAAA
AGCTTTGCTCGGAAGTGGACTGGGTTATGAATCGGAGAAGCGTGTGCAATCTTCGATGGGCTTCAATGAC
AATGATGGCGACCTCTCAGTCCCTCAGTCTGTTGGTGGCCGACCTCCTTCATGGGAGCCGAGCGGCATAT
TGGATGAAAAAGGGGAAATGTGACTACAAAGCATTTTGGGCCTGCACCTGCGGGTTCGAGTGGTTCGGCG
AGCGCACAACGCTGCAGGGTACCGCAGGATCAAACAATCAGCGACCTTCGATGAGAATGGTTTCTTTGCT
ATCGAGATGAACATCCCCACCCGACTGGCGCAATTTCTACCCATCAAGGGAGTTGAAGAGCAAAAAGACTA
CAAGGTATACTGCGATTACCACCGACCCAGATGATGTCACAGCAGCTGGCTTCCGTCTTCGCCAGAACAT
GACTTCTCCGCCCCGACAGACTGAACTTTTATCGTGATCACTATGTACAATGAGAACGCCGAGCTCTTT
TGTCGAACACTTTATGGTGTGATGAAGAATATAGCCACCTATGTGGGCGTAAGAACTCAAGGGTCTGGG
GCAAGGATGGTTGGCAAAGGTTGTCGTTGTCATTGTCGCGGACGGACGTAAGGCGGTTAACCCCCGCGT
CCTCGATTGTTTAGCAGCTCTTGGAGTTTACCAAGAAGGCGCAATGACGAACACAGTAAAGGATCGACCG
GTCACAGCGCATGTTTTCGAATACACGACCAGCTTTGCTCTTGACGGTGTATTTACACTTCAAATATCCAG
ACAAAGGCATTGTCCCCTGCCAGATTATCTTCTGCATGAAAGAGAAAAATGCCAAAAAGATCAACTCCCA
TCGATGGTTTTTCAACGCCCTTCGCGCCCTTGCTATCACCAAATGTCTGCATTCTTCTTGATGTGGGAACC
CAGCCAGCTCCGAAATCCATCTATCATCTTTGGAAAGCATTTGATGTCAATTTCTAATGTTGGTGGTGCCT
GTGGAGAAATTGCGACCTTCAAGGGCAAACCTTGAGGAGTATTATTGAACCCCTTGTCGCGGCCCAAGC
CTTTGAGTACAAGATGTCCAACATCCTCGACAAACCTTTGGAGAGTCTCTTCGGATACTGCACTGTGTTG
CCTGGTGCCTTCTCGGCTTACAGGTGGATCGCTTTGCAAAAACATGGGGATGGGAGAACGGGACCTTTGG
CGAGTTATTTTGGTGGTGAACAGCTCAATACTGGAAAGGCAGACACATTCACTGGTAATATGGCCAAACC
CAAGGCCAACTGGGTGCTGAAATTCGTTAAGGCTGCTGTTGGAGAAACAGATTGCCCTGATACCATCCCA
GAGTTTATTGCTCAAAGAAGAAGATGGCTTAAACGGTTCCTTCTTTGCAGCTGTCTATGCGTTGATGCACA
CGAACCAAATTTGGCGATCCGACCATTGCTTCGCGAGAAAGTCAGCCCTGATGTTGGAATCAGTGTACAA
CTTTCTGAACCTGATATCTCGTGGTTCGCTTTGGCAAACCTTTACATTTTCTTTGTCATCCTTACGAGC
GCTTTGGAGGGCAGCGCTTCAATGTCCCTCATATCGATGTGCTCAATACTATTGCACGATATGGTTACC
ttggtgctttggttggttggttCATCTTCGCAATGGGAAACAGGCCACAAGGTTCCGCTTGGAAAGTATAA
AGCAGCAATCTACTTTTTCGCCCTTTTACTACCTATATGCTGGTTCGAGCAGTGTCTTGTACGGTACAG
GCAATCAAAAATATAAACAGCCCAATTTTGGCAAAGATGGTAGTATCACTCATATCAACCTATGGTATTT
ATGTGATTTCCAGTTTCTTGCCCTTGACCCTTGGCACATCTTACTTGCTTTATTCAATATGTTCTCTT
CTCACCTACTTATATCAATGTTcttaatggtTATGCCATTCCAACCTTCACGACTTGTCATGGGGTACA
AAAGGCTCTGATGCAACCCAGGCGTCGGATTTGGGTGCTGTTCCGGAGTGGGAAAGCACGTCGAAGTGG
AACTTGTAACCTGCCAGCAAGACATTGATATTGCCTATCAGGATGCTTTGGACAATATTAGATTAAGAGG
ATCAAAAAGTTGACTCTGCTGAATCTGAGCCCAAAAAGGAGCAATCTGAACAAGCCCAGAAGGATACTTAT
GCCAACTTTCGTACCAATTTACTTTTGGTCTGGTTCGCTGTCAAACGCCCTTCTCGCAAGTGTATCCTTA
CAGGCAACAATTCGGAGCGTTTGACGAGGGTTCCGGCAGTTCAAAAGCCACAATATACATGCTTGTGAT
```

## Translation

```
>cneo_H99:GLEAN_01196
MPNISRKPPPRFYSPSHSPSPSLYAPIQSPAPSYDYHANPRTLNPFSDAREVGGYAQLQGEDQMTGAPL
YQPPYAPQLLVAQPTPVSSRLPFFEAALARARGIQTPSLPEAPTPAYAQPLPSYLPDPDPNHPDLSVGLT
QANTVRYAINPRSQLKEGSRSPSPFMDDSFVYNDAAHLYNVEPDVEKALLGSGLYESEKRVESMGMFND
NDGDLSVPQSFQGRPPSWEPGILDEKEMSTTKHFGPAPAGRVGRRAHNAAGYRRIKQSATLDENGFFA
IEMNIPTRLAQFLPIKGVVEQKTTRYTAITDTPDDVPAAGFRLRQNMTPPRQTELFIVITMYNENAELF
CRTLYGVMKNIAHLCGRKNSRVWKGKDWQKVVVCIIVADGRKAVNPRVLDCLAALGVYQEGAMTNTVKDRP
VTAHVFEYTTSFALDGLHFKYPDKGIVPCQIIFCMKEKNAKKINSHRWFNFAFAPLLSPNVCILLDVGT
QPAPKSIYHLWKAFDVNSNVGGACGEIATFKGKTWRSLLNPLVAAQAFEYKMSNILDKPLESLFGYCTVL
PGAFSAYRWIALQNNGDGRTGPLASYFAGEQLNTGKADFTTGNMAKPKANWVLKFKVKA AVGETDCPDTIP
EFIAQRRRWLNGSFFAAVYALMHTNQIWRSDHSFARKSALMLESVYNFLNLI FSWFALANFYIFFVILTS
ALEGSAFNVPHIDVLNTIARYGYL GALVGCIFAMGNRPQGS PWKYKAAIYFFALLTTYMLVA AVLCTVQ
AIKNINSPIFAKMVVS LISTYGIYVISSFLALDPWHIFTCFIQYVLF SPTYINVLNVYAYS NLHDL SWGT
KGS DATQASDLGAVSGVGKHVEVELVTAQQDIDIAYQDALDNIRLRGSKVDSA ESEPKKEQSEQAQKDTY
ANFRNLLLWVSLSNALLASVILTGNNSGAFDEGSGSSKATIYMLVILIFVAGMSIFRFICSTLYLVISL
FTG*
```



# GENE PAGE (3)

## Intron sequences

```
>cneo_H99:GLEAN_01196_intron1 CHROMOSOME13:645986..646042
GTAGGGCAGCGACTTTTGCAAGCTTGAGACCCCGTCTATTCGCTGACTCACACATAG
>cneo_H99:GLEAN_01196_intron2 CHROMOSOME13:646269..646324
GTGGGTGATCTTATATCTCCGCATATGTGTTTCAATATTGATGTCAAGGTATTCAG
>cneo_H99:GLEAN_01196_intron3 CHROMOSOME13:646413..646472
GTCAGGTTATCTTCAACAGTACAAAGCGCATTGCTGACATTTGATACTTTGACAAATAG
>cneo_H99:GLEAN_01196_intron4 CHROMOSOME13:646625..646685
GTGAGCTTGGGGCTCCATCATGTATGCGTCATGTATTCGTAGCTGATTACTTCTTCTTTAG
>cneo_H99:GLEAN_01196_intron5 CHROMOSOME13:646743..646797
GTATGTATCACTCATTCGTTTCGACCCAATGAACGCATCACTGACCATGTATGTAG
>cneo_H99:GLEAN_01196_intron6 CHROMOSOME13:646958..647023
GTAAGTCAGCAACCGCCGCACTAATATTCTACATGGTCAGCTAAACGCTGGTTTGTTCCTACTAG
>cneo_H99:GLEAN_01196_intron7 CHROMOSOME13:647052..647107
GTGAGTACAACATGCAAATTTATTTGTGTCGAATCTGACGCTGATAATGACCATAG
>cneo_H99:GLEAN_01196_intron8 CHROMOSOME13:647187..647247
GTGAGTTGCAACAGCTGAGCATCTTTAGTTTCCAGGACTCACAGCAGACGGTGATATGCAG
>cneo_H99:GLEAN_01196_intron9 CHROMOSOME13:647355..647569
GTGAGTCAAAAATTGTAGAACTCAAGCGTTTTACTGACTGCCTCGTTCTTCATCAGGGTATGTTTCATAG
CGGCGCTTGATGTTTCTTTTTGATTATCTTCTAATCAGCTGCTATAGTACCTGGCAGAAGATAGAATCCT
GTGTTTCGAAATCGTGTAGGCACCACTTCACGGTAATAATACATGTCATTGTTGCTGATCCAACACGGCG
TATAG
>cneo_H99:GLEAN_01196_intron10 CHROMOSOME13:647672..647737
GTGAGTAATTCTTTTTATCggggaaaaaagaaaaagggggggTATCTGACATTCTTCGTCTTCTAG
>cneo_H99:GLEAN_01196_intron11 CHROMOSOME13:647766..647815
GTGAGCGCCGCACGGATTTGGCATTGAGTTTACGCTTACATATCTTCAG
>cneo_H99:GLEAN_01196_intron12 CHROMOSOME13:647960..648030
GTGGGTACAGCCACTGTTGTATGTTTATACGGATACCCTAATAAGCaaaaaaaaaaaaaaaaaaaaaCA
G
>cneo_H99:GLEAN_01196_intron13 CHROMOSOME13:648106..648151
GTACGCTGTTTTTCATCCGTATAAGACATTAGCTCATTTCGATGTTAG
>cneo_H99:GLEAN_01196_intron14 CHROMOSOME13:648214..648274
GTATGTGTTTCAATTTCTCTGGACAAGAGGGACAGCCAGCCGACGCTTTTCATCTTTCTTCAG
>cneo_H99:GLEAN_01196_intron15 CHROMOSOME13:648329..648389
GTGAGTCGTCACAAGTGGTGCTCAGGGGTCAAGGATAATCACTAAACGTTTTTTAACACAG
```



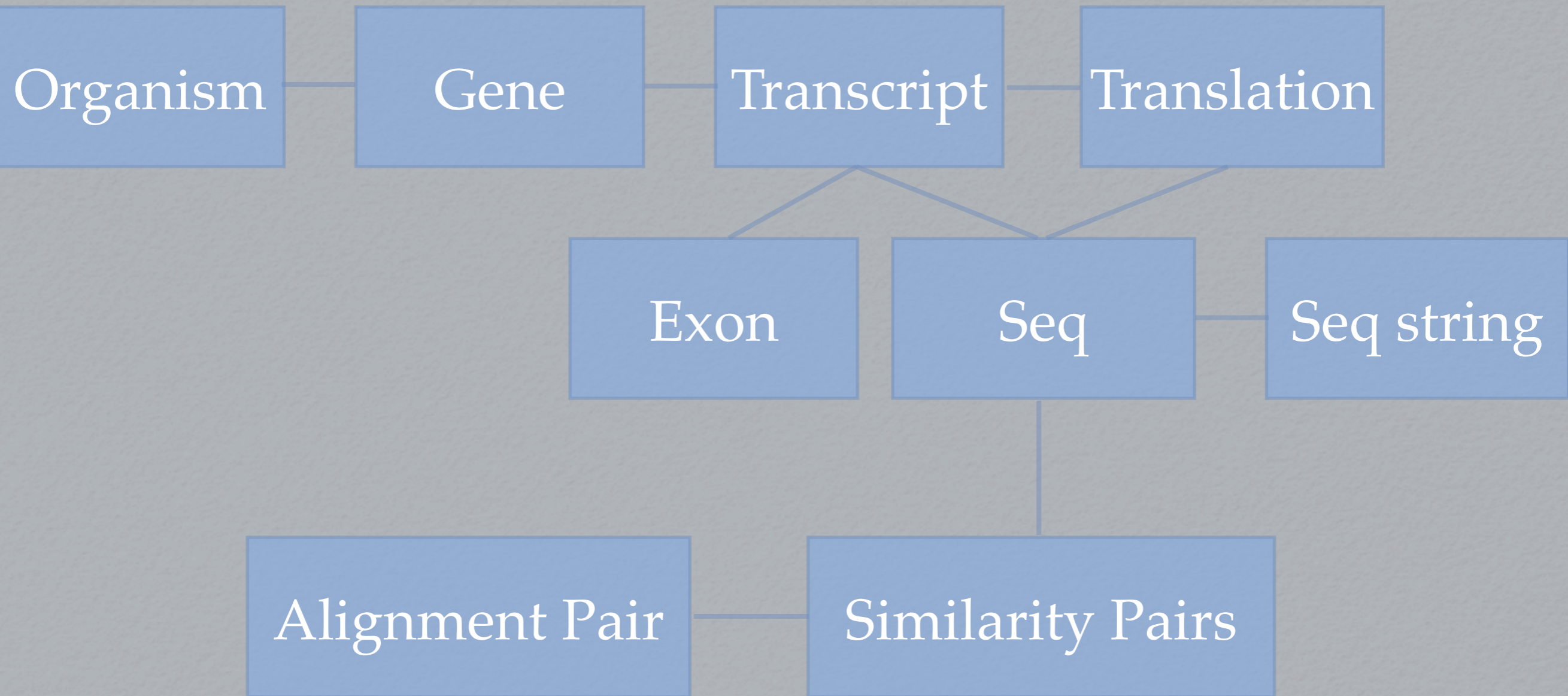
# GENE PAGE (4)

## Homologs from FASTA

| Hit   | Hit len | Bits | E-value  | % sim | % id  | % Query aligned |
|---|---------|------|----------|-------|-------|-----------------|
| <a href="#">Cryptococcus neoformans - cneo_H99:GLEAN_01196</a>    | 983     | 6060 | 0        | 100.0 | 100.0 | 100.0           |
| <a href="#">Cryptococcus gattii - cneo_WM276:GLEAN_00366</a>      | 984     | 5747 | 0        | 98.1  | 94.3  | 100.0           |
| <a href="#">Cryptococcus gattii - cneo_R265:GLEAN_gz_05631</a>    | 1009    | 3586 | 0        | 95.7  | 91.9  | 100.0           |
| <a href="#">Cryptococcus neoformans - cneo_JEC21:CNN02110</a>     | 996     | 3523 | 0        | 95.4  | 90.9  | 100.0           |
| <a href="#">Phanerochaete chrysosporium - pchr:GLEAN_gz_10814</a> | 903     | 1775 | 1.6e-116 | 64.8  | 42.5  | 99.6            |
| <a href="#">Uncinocarpus reesii - uree:GLEAN_05059</a>            | 845     | 1590 | 1.7e-103 | 69.0  | 43.0  | 77.9            |
| <a href="#">Phanerochaete chrysosporium - pchr:GLEAN_gz_12555</a> | 1004    | 1575 | 2.2e-102 | 74.2  | 52.9  | 83.9            |
| <a href="#">Coprinus cinereus - ccin:GLEAN_gz2_06353</a>          | 840     | 1514 | 3.7e-98  | 67.3  | 41.0  | 82.6            |
| <a href="#">Coprinus cinereus - ccin:GLEAN_gz2_11986</a>          | 1026    | 1495 | 9.7e-97  | 70.6  | 49.2  | 97.2            |
| <a href="#">Botrytis cinerea - bcin:BC1G_11533</a>                | 1168    | 1478 | 1.7e-95  | 65.4  | 42.0  | 99.5            |
| <a href="#">Trichoderma reesei - tree:12480</a>                   | 955     | 1458 | 3.7e-94  | 68.4  | 45.6  | 88.2            |
| <a href="#">Fusarium verticillioides - fver:GLEAN_09145</a>       | 1180    | 1448 | 2.2e-93  | 66.2  | 41.8  | 98.9            |
| <a href="#">Neurospora crassa - ncra:NCU05239.1</a>               | 926     | 1436 | 1.3e-92  | 68.8  | 44.5  | 86.9            |
| <a href="#">Histoplasma capsulatum - hcap_186R:GLEAN_05323</a>    | 1149    | 1433 | 2.5e-92  | 58.4  | 38.2  | 99.3            |
| <a href="#">Coprinus cinereus - ccin:GLEAN_gz2_06575</a>          | 941     | 1430 | 3.4e-92  | 65.4  | 43.4  | 99.6            |
| <a href="#">Uncinocarpus reesii - uree:GLEAN_08490</a>            | 1210    | 1427 | 6.9e-92  | 65.8  | 41.8  | 99.0            |
| <a href="#">Coccidioides immitis - cimm:anid_cimm_1.72-g26.1</a>  | 1244    | 1423 | 1.3e-91  | 65.2  | 41.4  | 98.2            |
| <a href="#">Phanerochaete chrysosporium - pchr:GLEAN_gz_04887</a> | 647     | 1420 | 1.3e-91  | 68.9  | 44.0  | 65.3            |



# SIMILARITY DATABASE





# OTHER TOOLS

- BLAST interface
  - Search your sequence and get marked up results with links to Gbrowse
  - “Yeast protein to genomic visualization of locus in your organism of interest”



# BLAST TOOL

**Database and Program Options**

Program:  Database:

Enter sequence below (most standard FASTA suggested). Maximum 1000000 characters.  
>gi|1302185|emb|CAA96086.1  
MSDQNNRSRNEYHSNRKNPSYEL  
TNMLYNGDDGNNNTINDNERDIY  
VIQTTPELIHNGSQTMATPIERPFF  
IPQYHDQPFQYNGYHGLQAKDY  
EYLHDDSRPVNDGKEELDSVKSGY  
KESDIIVSNDNLTANRALKRSGTEIF  
VTCEPNQLAEKNFTVRQLKYLTPR  
KKIVVCIISDGRSKINERSLALLSSLC  
GTVPIQLLFCLKEQNQKKINSHRW  
IRTDLGKRFVKLLNPLVASQNFYK  
ENEGFHFFSSNMYLAEDRILCFEVV

Or load it from disk

Set subsequence: From

The query sequence is filtered for low complexity. Filter  Low complexity

Post Process with Smith-Waterman


Expect:  Matrix:

Powered by the [WU-Blast Program](#)

- nt Archeascomycota
- nt Basidiomycota
- nt Cryptococcus
- nt Euscomycota**
- nt Hemiascomycota
- nt Zygomycota
- nt ashbya\_gossypii
- nt aspergillus\_fumigatus
- nt aspergillus\_nidulans
- nt aspergillus\_terreus
- nt botrytis\_cinerea
- nt candida\_albicans
- nt candida\_glabrata
- nt candida\_guilliermondii
- nt candida\_lusitaniae
- nt candida\_tropicalis
- nt chaetomium\_globosum
- nt coccidioides\_immitis
- nt coprinus\_cinereus
- nt cryptococcus\_neoformans\_H99
- nt cryptococcus\_neoformans\_JEC21
- nt cryptococcus\_neoformans\_R265
- nt cryptococcus\_neoformans\_WM276
- nt debaryomyces\_hansenii
- nt fusarium\_graminearum
- nt fusarium\_verticillioides
- nt histoplasma\_capsulatum\_186R
- nt kluveromyces\_lactis
- nt kluveromyces\_waltii
- nt magnaporthe\_grisea
- nt neurospora\_crassa
- nt phanerochaete\_chrysosporium
- nt pneumocystis\_carnii
- nt podospira\_anserina
- nt saccharomyces\_bayanus
- nt saccharomyces\_castellii
- nt saccharomyces\_cerevisiae\_rm11-1a\_1
- nt saccharomyces\_cerevisiae\_s288c
- nt saccharomyces\_cerevisiae\_yjm789
- nt saccharomyces\_kluveri
- nt saccharomyces\_kudriavzevii

Overlay Hits over Genome Image

0 seqs at a time





# RE-FORMATTED BLAST

## TBLASTN Query of GI|1302185|EMB|CAA96086.1| against nt Euascomycota

TBLASTN 2.0MP-WashU [10-May-2005] [linux24-i686-ILP32F64 2005-05-10T21:16:37]

Copyright (C) 1996-2000 Washington University, Saint Louis, Missouri USA.  
All Rights Reserved.

**Reference:** Gish, W. (1996-2000) <http://blast.wustl.edu>

**Query=** GI|1302185|EMB|CAA96086.1| CHS1 [SACCHAROMYCES CEREVISIAE]

(1,131 letters)

**Database:** uncinocarpus\_reesii.2.nt; chaetomium\_globosum.20041105.nt; coccidioides\_immitis.20040311.nt; fusarium\_gra  
magnaporthe\_grisea.20031031.nt; neurospora\_crassa.20020212.nt; podospora\_anserina.20040122.nt; aspergillus\_fumiga  
stagonospora\_nodorum.20050205.nt; aspergillus\_terreus.1.nt; fusarium\_verticillioides.2.nt; sclerotinia\_sclerotiorum.1.nt

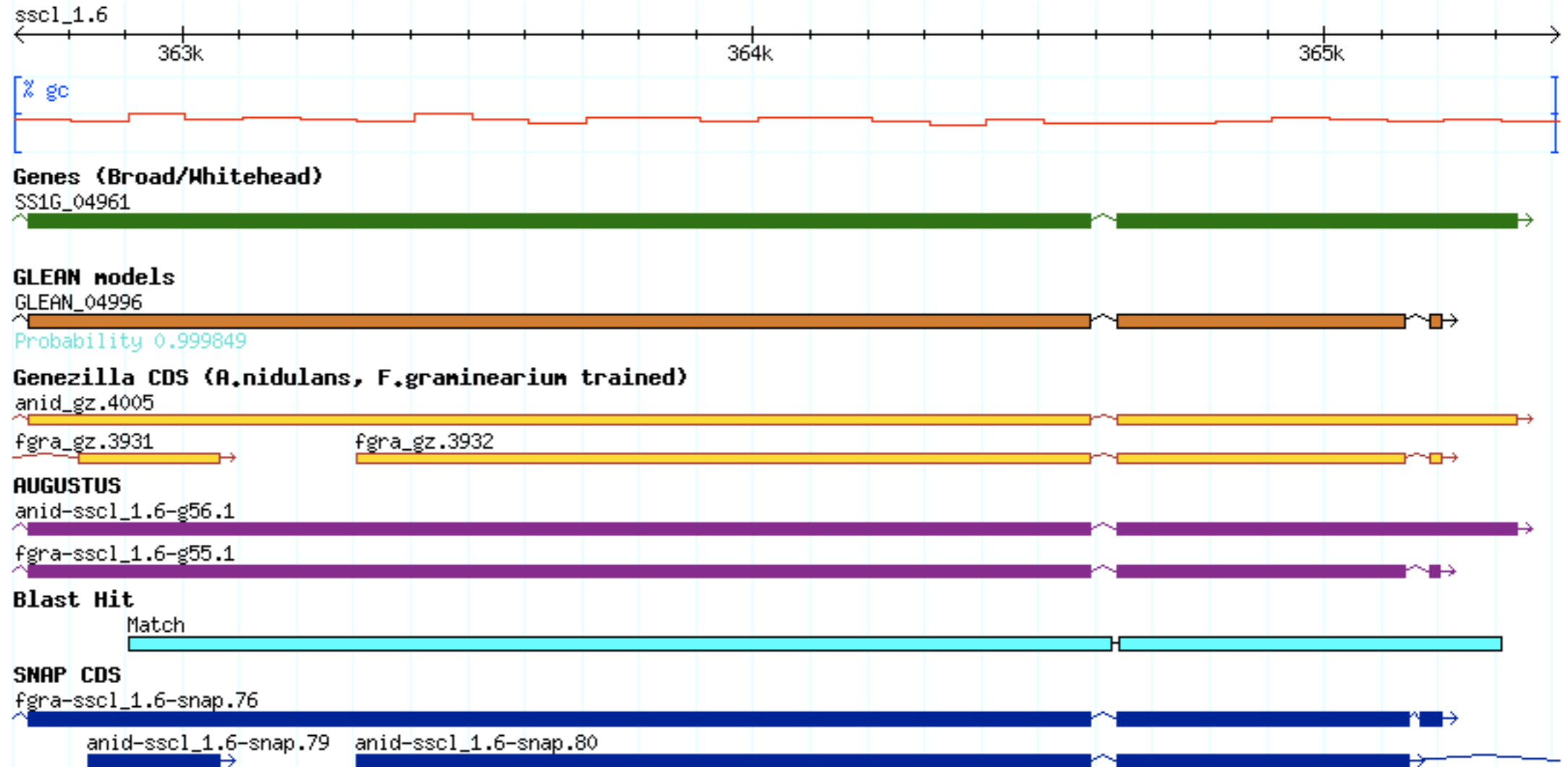
12,142 sequences; 504,395,971 total letters

| <b>Sequences producing significant alignments:</b> | <b>Score</b>  | <b>E</b>     |
|--|---------------|--------------|
|  | <b>(bits)</b> | <b>value</b> |
| sscl_1.6   | 1413          | 5.6e-143     |
| snod_1.8   | 1381          | 1.4e-139     |
| ncra:ncra_3.221                                    | 1366          | 5.2e-138     |
| tree_50  | 1361          | 1.7e-137     |
| fgra:fgra_1.425                                    | 1353          | 1.2e-136     |
| fver_2.7   | 1347          | 5.5e-136     |
| anid:anid_1.78                                     | 1312          | 2.8e-132     |
| uree_2.4   | 1303          | 2.5e-131     |
| hcap_186R:hcap-186R_17.30                          | 1299          | 3.2e-131     |
| cimm:cimm_1.106                                    | 1297          | 1.1e-130     |
| afum:afum_57                                       | 1289          | 7.7e-130     |
| ater_1.8   | 1286          | 1.6e-129     |



# WITH LINKS

sscl\_1.6 Link\_group:2



Length = 2,040,374

Score = 502.5 bits (1413), Expect = 5.6e-143, P = 5.6e-143  
Identities = 290/586 (49%), Positives = 386/586 (65%), Gaps = 23/586 (3%), Frame = +3  
Links = (1)

Query: 314 KDDFSRDDEYDDLNTIDKLFQANGVPASSSVSSIGSKESDIIIVSNDNLTANRALKRSGT 373  
+D + +D+ DD I LQ P + S K D + + T AL+R T  
Sbjct: 362907 QDPYGYND- DDHQPI--LQSHEPYGPDPTASGAEYKGYDGAGHSPSSTPIPALRRYKT 363077



# ADDITIONAL DATA TO INTEGRATE

- Curated life-history information about sequenced fungi (with Anne Pringle, Harvard)
- Expression data...
- Mart-enabled?



# WHAT'S MISSING

- Homolog / Ortholog / Paralog capturing
  - Pairwise focused summary statistics
  - Multiway ortholog summaries
  - Ensembl Compara --> GMOD Compara?
  - Linking to gene trees



# QUERIES TO ADDRESS

- “All the genes in closely related pathogenic fungi not present in non-pathogenic outgroup”
  - Species-tree defined unique genes, etc
- “Rapidly evolving cell-surface associated genes”
- Gene family size change (paralogous expansions)