**BTCH90009**

**Assignment 3: Computational Gene Prediction**

Name: Jiayu Wang

Student ID: 1039580

Word Count: 1459 (excluding tables and figures)

# Introduction

Gene prediction via computational approaches is one of the bioinformatics hotspots. Gene prediction, also called gene finding, generally refers to locating genes in a genome as well as annotation of the biological functions of the genes [1].

In this research, a 20-kb unknown DNA sequence is allocated to me and I will perform gene prediction and annotation on this DNA sequence using two types of commonly-used methods: *Ab initio* methods and homology-based methods [2].

*Ab initio* methods are considered as intrinsic methods since they only use DNA sequences themselves as input and only based on the gene structure. This kind of prediction is derived from certain genomic features including: signal sensors such as certain codons and splice sites, content sensors like the species-specific patterns of codon usage. Homology-based methods rely on the alignment of the subject sequence with the known protein, mRNA or cDNA sequence and use the similarity to infer the gene structure or function.

By comparison, homology-based methods are more suitable for prokaryotes since prokaryotic genomes are highly condensed and the recognition of translational start sites can be hard by *Ab initio* methods. For eukaryotic genes, *ab initio* methods are more suitable due to the complexity of eukaryotic genomes such as introns and non-coding regions [3]. In practice, researchers usually combine these two methods to achieve better performance though.

# Methods

## Identifying the organism

I used BLASTn to identify the organism by using the entire 20-kb assigned DNA sequence (No. 35) and used non-redundant database [4]. Since BLASTn is more efficient to analyze the homology between the input DNA sequence and genomic sequences from different species regardless these sequences encode proteins or not. Thus, I chose BLASTn to identify the sourced species.

## *ab initio* method of gene prediction

In this research, I used FGENSH program to perform *ab initio* gene prediction[5]. FGENESH is a trained online gene prediction tool based on hidden Markov model (HMM) that supports prediction in both chains of DNA sequences [6]. FGENESH has overall better performance regarding both specificity and sensitivity compared with other *ab initio* methods according to previous researches [7].

I used the entire assigned DNA sequence as input on FGENESH and specified the organism as *Talaromyces marneffei* as previously identified by BLASTn. To identify the type of genes, I used BLASTx with full-length mRNA in the result of FGENESH. This is because genes are expressed into proteins and hence we need to focus on the alignment with protein product.

### Homology-based method of gene prediction

I used BLASTx with the entire assigned DNA sequence as input and used refseq_protein database to avoid unwanted job exit because the large size of the query sequence can cause overwhelming computational demand [8].

I limited the organism to only *Talaromyces marneffei* otherwise the potential true gene with not exact matching might be ranked too behind in the BLASTx results and might be missed [8].

## Results

### Identifying the organism

The top five results of identifying the organism in BLASTn are as shown in **Table 1.** Except for the fourth sequence, all the other sequences share a common sourced organism: *Talaromyces marneffei.* Also, the identity of the fourth sequence (81.77%) is much lower than the other sequences. Hence, I identified the target organism as *Talaromyces marneffei.*

**Table 1. Top five BLASTn results showing the source organism**

| Sequence ID | Name | Sourced organism | Identity | Scores | E-value |
|---|---|---|---|---|---|
| CP015873.1 | *Talaromyces marneffei* strain TM4 chromosome | *[Talaromyces marneffei]* | 99.69% | 36622 | 0 |
| CP045654.1 | *Talaromyces marneffei* isolate 11CN-20-091 chromosome 2, | *[Talaromyces marneffei]* | 99.65% | 36583 | 0 |
| XM_002151891.1 | *Talaromyces marneffei* amidohydrolase, putative (PMAA_047390), partial mRNA | *[Talaromyces marneffei]* | 100% | 2344 | 0 |
| CP017349.1 | *Talaromyces pinophilus* strain 1-95 chromosome 6, complete sequence | *[Talaromyces pinophilus]* | 81.77% | 1860 | 0 |
| XM_002151890.1 | *Talaromyces marneffei* ATCC 18224 uncharacterized protein (PMAA_047380), partial mRNA | *[Talaromyces marneffei]* | 100% | 1766 | 0 |

### Gene prediction by *ab initio* method

#### (i) The chosen *ab initio* method: FGENESH

Compared with Augustus, FGENESH is more suitable in this research. This is because there is no specific organism option in the Augustus for the target organism *Talaromyces marneffei.* I have run Augustus by setting the organism species as *Aspergillus terreus*, which is the closest species to *Talaromyces marneffei* [9]. Only 3 genes were reported in this Augustus result and all these 3 genes are including in the FGENESH result. Besides, recent evidence shows that FGENESH has an overall better performance in both specificity and sensitivity [3]. Hence, I chose the FGENESH result as the most accurate *ab initio* annotation.

#### (ii)Summary diagram of genes predicted by FGENESH

The summary diagram is shown below in **Fig.1** where the positions of start codons, stop codons, gene directions, exons and introns are labeled.

In the FGENESH result, the start codons and stop codons of all five genes are with typical sequences such as ATG, TGA, TAA. All five genes have long ORF. These two features are signs of reliability.

Although the GC contents of Gene4 exon11 and Gene5 exon2 are shown slightly higher than 50%, which indicates a potential possibility of long false-positive.

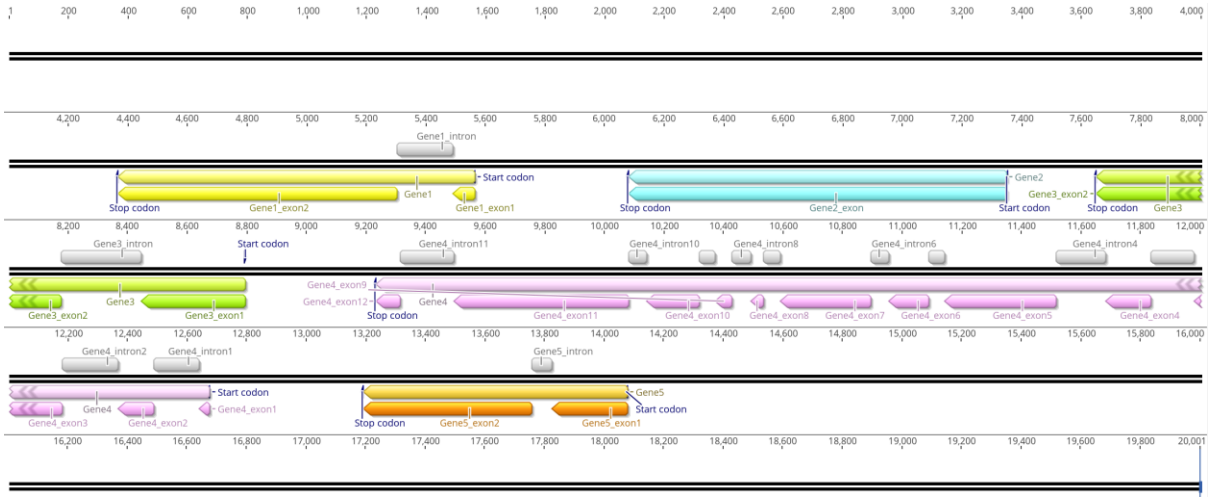


**Fig.1**. **Summary diagram of gene prediction by FGENESH.** 5 genes are inhibited in different colors and the arrow direction stands for the gene direction. Exons and introns of each gene are represented above and below the gene indicator bars respectively.

### (iii) Polypeptide sequences of the genes predicted by FGENESH

Polypeptide sequences of the predicted genes are shown in **Table 2** respectively.

**Table 2. Polypeptide sequences of the predicted genes by FGENESH**

| Gene | Polypeptide sequence |
|---|---|
| **Gene1** | MEKKTKVELRNYIPQLISKAEWRKRRTTAEVHKGLSLTRQRSVTGDEWHATAEAKECGVK IQFSSIKGNSTQELRLNYLLSVKTSSNHGAVIERMKNTISQIPKPIYNDRITGFVKSERF RAWITSKSSDALLVNNDSVNIYRGANTVSLLTELTLLLYEKLSILPKAEKVFTLMYPCAE YPATIGDMTTVRRMLRHLTSQYYDNDREHWDKQKFVRKYISTLKSIREEKPEKEQEKTNL LSDLFIGMLEEAQPDQVIYILIDGFDLVEMDEPEANKSDFWWLLTCFYSLIKDLRKLKKP HPVIKLLITYHGTCQDETRAVWKDYLMDISKPNGL |
| **Gene2** | MAQAHELSDILKNASLDLSPYEDLYKYFHANPELSRQEKSTSEKIAAHLSALKAYEIHTN IGGYGLVGVLKNGVGKTILLRADMDALPVKELTGLPYASSITMSDAEGVEKPVMHACGHD MHITCLLAAAELLANTQHAWSGTLIVLFQPDEERGGGAQAMVDDGLYSKIPVPDYCFGQH VMRMRAGTIGSRPGTIMAAADSMKVTVFGRGGHGSLPHQTVDPALLAAHIVVRLQSIVSR EIDPTDLGVVTVGSLQAGQTENIIADRAEIGLDFRTVKLETREKILSAIRRIVEAECMAS GSPKPPIFTPTRRFPPTNNDRDVASKLAASFETHFGDSFDGDISRSNVAEDFSTLATSKG VPSCFWFLGGIDPDLWDRVSKEENPTEEIPTNHSAFFAPVIQPTMKAGVDALCIAALTFL RK |
| **Gene3** | MVFAVPEQPKTELGRYRILSSTAGVRVSPIALGALSLGKSWSLIMGTVDEEQSFELLDAY TDAGGNFIDTANNYQNEDSEKYIGKWMAARGNRDLLFIATKFTNSYRTHELGTGKTGRWN VMRRDFEREIIPMALHFGMALCPWDVLGGGQFKTARQLEAAENSNQGSRSGVQSEEARQV SAALERVVSGHGINSVQQIALAYIMQKARNVFPVVGGRMVEHLHDNIGALSIYLTDEQIR YLDSIKDFDLGFPLNFIGDDPKEIGTQPRMMEILTGAKIAWQKSPKPIGHD |
| **Gene4** | MTFVLMTLSKFRIEQYPSFDFTSVIDLRDNRYPSTIKFLFKAPNQGVAWGTVTFKFPCYL |

| | |
|---|---|
| | GDKYDLHLMPFIDIFMSTKYRVESDLVFKVTQLRTKELERSLASNYRATTSRQLLRVRRS<br>LLRLHPGIPLAGETRLSKQQSVVILPCGQVDILSKKAETQRGWTGFQAVGNGALFLMMRN<br>SISSHISYLLLLLFKSFNIATSSMARLSEAFYASIINSLHDHTMLDILTPSLKWWLSCIL<br>LYIVALVVYRLKFHHAAHIPGPILAKVTFLYEWYYDLYLGGQYAFKLKELHKIYGPIIRI<br>SPETVHIDDPDYFEVVHNQKNGREVKPPHVAETFGPYPALLGTGPHELHRVRRSALSPFF<br>SRKAVLDLVPTIQRPTAILCRRLQEASLTGETMNLKYIFAAVTLDIINDYCFAREPTTTL<br>EPDFGKKETDNVAYFLKVSLLNTHIPWLMRLTYSLPDRVSKTFTPAVAKILDFRMELSRQ<br>VEAIRNGQDTSHQTAAHRTVFHELLDSTLPPDELKPARLRDEAFSLVAAGSDTTAHVLRG<br>TTYHISANPTIRLQLFNELKTAIPDPDNLPSLPELEKLPYLSAVIHEGLRLANPVTHRII<br>RQYPDKAFDYQGYLIPPNTIVGMSPNLVHFNEDKFPEPQVFKPERWLGPDNARLQRYLIS<br>FSRGPRACLGIALARAELFIILASVFRQFELDVSEIERSRDIDFSRDFILGAQADDSPGA<br>LVKVKLAGYKTIDEVLSMENEGEQGGSLVQIDRRG |
| **Gene5** | MPQVLLLGGHGKIALYLTPLLLARSWNVTSVIRNAEHEVEILALGNGKPGKVSVLLSSLD<br>DVKSQADAQAVLDKVKPDYVVWAAGAGGKGAPQRTYAIDQDAAKYFIASSFAAPSVTKFL<br>LISHMGSRKVKPSWLSEDGWKRTQWLWSVGLPDYCKAKWEADQYQTALAAITKQREPTRK<br>FQSISLRPGHLIDDPATGKVALGHVAAAEGSVTREDVAIVADRLLARDDTEGWFDLLGGE<br>ELIDEAVERVAKEKIDSIGDEDVDAMIKKFNL |

**(iiii) Type of the proteins encoded by the genes predicted by FGENESH**

The BLASTx results showing the type of encoded protein are shown in **Table 3**. All five alignment results are with extremely low E-values as well as identities and similarities at 100% or close to 100%, which means these alignments are well-matched and less possible to be false.

**Table 3. BLASTx results showing the type of the proteins encoded by the genes predicted by FGENESH**

| Genes | Sequence ID | Name | Sourced organism | Identity, Similarity | Scores | E-value |
|---|---|---|---|---|---|---|
| **Gene1** | XP_002151926.1 | uncharacterized protein *PMAA_047380* | *[Talaromyces marneffei]* | 98%, 97% | 625 | 0 |
| **Gene2** | XP_002151927.1 | amidohydrolase, putative | *[Talaromyces marneffei]* | 100%, 100% | 778 | 0 |
| **Gene3** | XP_002151928.1 | norsolorinic acid reductase, putative | *[Talaromyces marneffei]* | 95%, 94% | 310 | 3e-108 |
| **Gene4** | XP_002151930.1 | benzoate 4-monooxygenase cytochrome P450 | *[Talaromyces marneffei]* | 100%, 100% | 1040 | 0 |
| **Gene5** | XP_002151931.1 | conserved hypothetical protein | *[Talaromyces marneffei]* | 100%, 100% | 499 | 0 |

## Gene prediction by homology-based method

### (i) Genes predicted by BLASTx

Here I used BLASTx to perform homology-based gene prediction. The alignment results are shown in **Table 4**. All 5 genes are of high similarity and identities, high scores and low E-values (close to 0), which means the alignment results are well-matched and less suspicious.

**Table 4. BLASTx results showing homology-based gene prediction**

| Genes | Sequence ID | Name | Sourced organism | Identity, Similarity | Scores | E-value |
|---|---|---|---|---|---|---|
| **Gene1** | XP_002151926.1 | *uncharacterized protein* PMAA_047380 | *[Talaromyces marneffei]* | 99%, 100% | 600 | 0 |
| **Gene2** | XP_002151927.1 | amidohydrolase, putative | *[Talaromyces marneffei]* | 100%, 100% | 778 | 0 |
| **Gene3** | XP_002151928.1 | norsolorinic acid reductase, putative | *[Talaromyces marneffei]* | 95%, 94% | 307 | 3e-96 |
| **Gene4** | XP_002151930.1 | benzoate 4-monooxygenase cytochrome P450 | *[Talaromyces marneffei]* | 88%, 87% | 529 | 0 |
| **Gene5** | XP_002151931.1 | conserved hypothetical protein | *[Talaromyces marneffei]* | 100%, 100% | 383 | 0 |

### (ii)Type of encoded proteins

All the five identified types of encoded proteins by homology-based method here are the same as those predicted by FGENESH above as shown in **Table 4**.

### (iii)Summary diagram of genes predicted by BLASTx

A summary diagram showing the top 5 predicted genes is shown below in **Fig.2** where the positions of start codons, stop codons, gene directions, exons and introns are labeled. All five genes have long ORFs with typical start and stop codons, which means these results are less likely to be false-positive. GC content in Gene5 exon2 is slightly higher than other exons. Gene1, Gene2, Gene3, Gene5 are the same as those by FGENESH above. Only Gene4 is different from Gene4 by FGENESH regarding gene length and the number of exons.

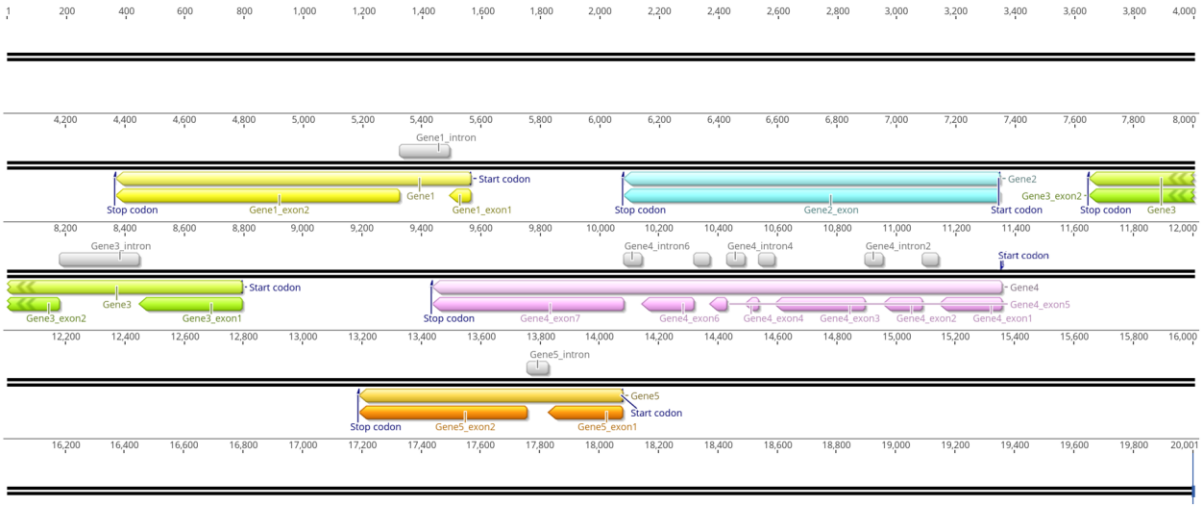


**Fig.2**. **Summary diagram of gene prediction by BLASTx**. 5 genes are inhibited in different colors and the arrow direction stands for the gene direction. Exons and introns of each gene are represented above and below the gene indicator bars respectively.

**(iiii) Polypeptide sequences of the genes predicted by BLASTx**

The polypeptide sequences of the genes predicted by BLASTx are shown below in **Table 5**. Only the product of gene4 is different from the results by FGENESH above in length and some peptide segments, polypeptide sequences of the other four genes are the same.

**Table 5. Polypeptide sequences of the predicted genes by BLASTx**

| Gene | Polypeptide sequence |
|---|---|
| **Gene1** | MEKKTKVELRNYIPQLISKAEWRKRRTTAEVHKGLSLTRQRSVTGDEWHATAEAKECGVK<br>IQFSSIKGNSTQELRLNYLLSVKTSSNHGAVIERMKNTISQIPKPIYNDRITGFVKSERF<br>RAWITSKSSDALLVNNDSVNIYRGANTVSLLTELTLLLYEKLSILPKAEKVFTLMYPCAE<br>YPATIGDMTTVRRMLRHLTSQYYDNDREHWDKQKFVRKYISTLKSIREEKPEKEQEKTNL<br>LSDLFIGMLEEAQPDQVIYILIDGFDLVEMDEPEANKSDFWWLLTCFYSLIKDLRKLKKP<br>HPVIKLLITYHGTCQDETRAVWKDYLMDISKPNGL |
| **Gene2** | MAQAHELSDILKNASLDLSPYEDLYKYFHANPELSRQEKSTSEKIAAHLSALKAYEIHTN<br>IGGYGLVGVLKNGVGKTILLRADMDALPVKELTGLPYASSITMSDAEGVEKPVMHACGHD<br>MHITCLLAAAELLANTQHAWSGTLIVLFQPDEERGGGAQAMVDDGLYSKIPVPDYCFGQH<br>VMRMRAGTIGSRPGTIMAAADSMKVTVFGRGGHGSLPHQTVDPALLAAHIVVRLQSIVSR<br>EIDPTDLGVVTVGSLQAGQTENIIADRAEIGLDFRTVKLETREKILSAIRRIVEAECMAS<br>GSPKPPIFTPTRRFPPTNNDRDVASKLAASFETHFGDSFDGDISRSNVAEDFSTLATSKG<br>VPSCFWFLGGIDPDLWDRVSKEENPTEEIPTNHSAFFAPVIQPTMKAGVDALCIAALTFL<br>RK |
| **Gene3** | MVFAVPEQPKTELGRYRILSSTAGVRVSPIALGALSLGKSWSLIMGTVDEEQSFELLDAY<br>TDAGGNFIDTANNYQNEDSEKYIGKWMAARGNRDLLFIATKFTNSYRTHELGTGKTGRWN<br>VMRRDFEREIIPMALHFGMALCPWDVLGGGQFKTARQLEAAENSNQGSRSGVQSEEARQV<br>SAALERVVSGHGINSVQQIALAYIMQKARNVFPVVGGRMVEHLHDNIGALSIYLTDEQIR<br>YLDSIKDFDLGFPLNFIGDDPKEIGTQPRMMEILTGAKIAWQKSPKPIGHD |
| **Gene4** | MLDILTPSLKWWLSCILLYIVALVVYRLKFHHAAHIPGPILAKVTFLYEWYYDLYLGGQYAFKLKELHKI<br>YGPIIRISPETVHIDDPDYFEVVHNQKNGREVKPPHVAETFGPYPALLGTGPHELHRVRRSALSPFFSR<br>K |

| | |
|---|---|
| | AVLDLVPTIQRPTAILCRRLQEASLTGETMNLKYIFAAVTLDIINDYCFAREPTTTLEPDFGKKETDNVA<br>YFLKVSLLNTHIPWLMRLTYSLPDRVSKTFTPAVAKILDFRMELSRQVEAIRNGQDTSHQTAAHRTV<br>FHE<br>LLDSTLPPDELKPARLRDEAFSLVAAGSDTTAHVLRGTTYHISANPTIRLQLFNELKTAIPDPDNLPSLP<br>ELEKLPYLSAVIHEGLRLANPVTHRIIRQYPDKAFDYQGYLIPPNTIVGMSPNLVHFNEDKFPEPQVFK<br>P<br>ERWLGPDNARLQRYLISFSRGPRACLGIALARAELFIILASVFRQFELDVSEIERSRDIDFSRDFILGAQ<br>ADDSPGALVKVKLAG |
| **Gene5** | MPQVLLLGGHGKIALYLTPLLLARSWNVTSVIRNAEHEVEILALGNGKPGKVSVLLSSLD<br>DVKSQADAQAVLDKVKPDYVVWAAGAGGKGAPQRTYAIDQDAAKYFIASSFAAPSVTKFL<br>LISHMGSRKVKPSWLSEDGWKRTQWLWSVGLPDYCKAKWEADQYQTALAAITKQREPTRK<br>FQSISLRPGHLIDDPATGKVALGHVAAAEGSVTREDVAIVADRLLARDDTEGWFDLLGGE<br>ELIDEAVERVAKEKIDSIGDEDVDAMIKKFNL |

**One commonly found gene**

By checking with my colleagues' results, gene cytochrome P450 is the common type of gene found in all DNA segments.

## Discussions

**(i) Comparison between *ab initio* and homology-based methods in this research**

For my assigned sequences, these two methods show highly consensus predictions except only Gene4, which means the four consensus genes are very likely to be true. Gene4 predicted by FGENESH is longer and has more exons. It is difficult to determine which prediction is more accurate since long genes with multiple exons are naturally not easy to be accurately predicted. In this case, I tend to believe the prediction by *ab initio* since it has better BLASTx alignment results with 100% identities and similarities to the reference protein. FGENESH also involves Besides, as discussed in the introduction section, *ab initio* methods are more suitable for eukaryotes [3].

**(ii) The features of BLASTx**

BLASTx is to translate the nucleotide sequence queries into polypeptides and then align to the reference protein, which is a protein-to-protein comparison. BLASTn is to perform nucleotide-nucleotide comparison. BLASTp is also protein-to-protein comparison while the queries are polypeptide sequences.

In this research, BLASTx is the most suitable for homology-based prediction since our purpose is to find genes. Genes have their protein product while non-coding sequences do not. Hence BLASTx is better than BLASTn since BLASTn includes comparison for both coding and non-coding nucleotide sequences. My assigned sequence is nucleotide sequence instead of polypeptide sequence so BLASTp is less suitable.

**(iii) Nature of the commonly-predicted gene: cytochrome P450**

Gene cytochrome P450 encodes cytochrome P450 enzymes (P450s), which are important haem-thiolate monooxygenase superfamily [10]. P450s are commonly found in all organism forms from prokaryotes to eukaryotes and are involved in diverse biological roles [11]. P450s typically catalyze the conversion of lipophilic intermediates to become more hydrophilic by oxidation reaction in different metabolic pathways [12].

In fungi, different P450s catalyze electron transfer in different forms and several typical catalyzation are shown in **Fig.3** [13]. P450s are essential for fungi to adapt to extreme environmental conditions by detoxifying pollutants and toxic chemicals [14].

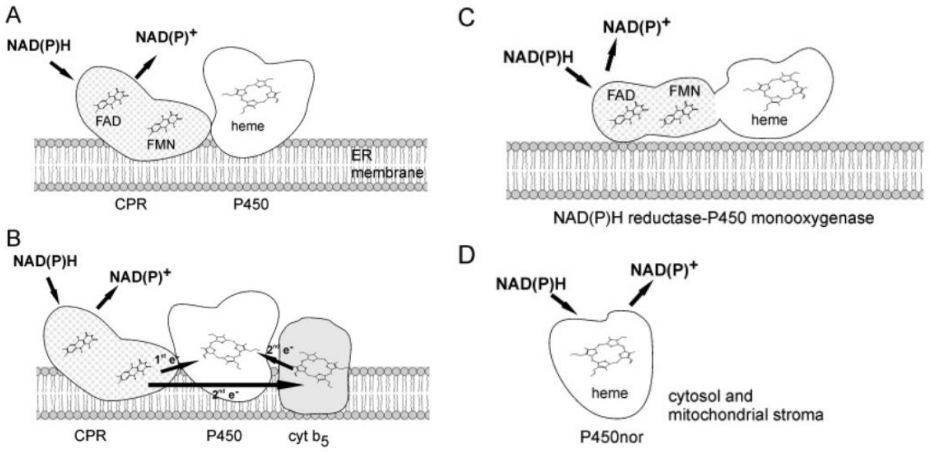


**Fig.3.** Schematic demonstration of different systems of P450s in fungi [13]. **a,b** microsomal system, class II; **c** fungal and bacterial fusion system, class VIII; **d** fungal soluble system, class IX.

The topography in P450s is highly conserved. P450s have two typical signature motifs, CXG and EXXR in their polypeptide sequences, as shown in **Fig.4** [15]. These two motifs are included in the haem-binding loop, a highly conserved structure that contains active site heme-iron center that performs binding and catalysis functions [16].

The homology between the predicted Gene4 and P450s is of 100% identity according to BLASTx results.

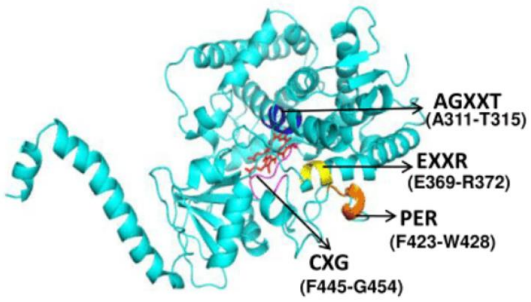


**Fig.4.** Schematic demonstration of conserved motifs in P450s [15]

## Conclusions

In summary, five genes are predicted and are consensus in both *ab initio* method and homology-based method, though the length and exons of Gene4 are slightly between the two methods. All predicted genes show acceptable reliability and are supported with strong evidence.

## Reference


1. Putluri S, Mohammad ZUR, Smieee, Amara C, Putluri Nr: **New Exon Prediction Techniques Using Adaptive Signal Processing Algorithms for Genomic Analysis.** *IEEE Access* 2019, **7:**1-1.
2. Wang Z, Chen Y, Li Y: **A brief review of computational gene prediction methods.** *Genomics, proteomics & bioinformatics* 2004, **2:**216-221.
3. Nasiri J, Naghavi M, Rad SN, Yolmeh T, Shirazi M, Naderi R, Nasiri M, Ahmadi S: **Gene identification programs in bread wheat: a comparison study.** *Nucleosides Nucleotides Nucleic Acids* 2013, **32:**529-554.
4. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL: **NCBI BLAST: a better web interface.** *Nucleic acids research* 2008, **36:**W5-W9.
5. Salamov AA, Solovyev VV: **Ab initio gene finding in Drosophila genomic DNA.** *Genome research* 2000, **10:**516-522.
6. Kulp D, Haussler D, Reese MG, Eeckman FH: **A generalized hidden Markov model for the recognition of human genes in DNA.** *Proc Int Conf Intell Syst Mol Biol* 1996, **4:**134-142.
7. Yao H, Guo L, Fu Y, Borsuk LA, Wen TJ, Skibbe DS, Cui X, Scheffler BE, Cao J, Emrich SJ, et al: **Evaluation of five ab initio gene prediction programs for the discovery of maize genes.** *Plant Mol Biol* 2005, **57:**445-460.
8. Pruitt KD, Tatusova T, Maglott DR: **NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucleic acids research* 2007, **35:**D61-D65.
9. Lau S, Tsang C-C, Woo P: **Talaromyces marneffei Genomic, Transcriptomic, Proteomic and Metabolomic Studies Reveal Mechanisms for Environmental Adaptations and Virulence.** *Toxins* 2017, **9:**192.
10. George HL, Hirschi KD, VanEtten HD: **Biochemical properties of the products of cytochrome P450 genes (PDA) encoding pisatin demethylase activity in Nectria haematococca.** *Archives of microbiology* 1998, **170:**147-154.
11. Lisowska K, Szemraj J, Różalska S, Długoński J: **The expression of cytochrome P-450 and cytochrome P-450 reductase genes in the simultaneous transformation of corticosteroids and phenanthrene by Cunninghamella elegans.** *FEMS microbiology letters* 2006, **261:**175-180.
12. Sono M, Roach MP, Coulter ED, Dawson JH: **Heme-containing oxygenases.** *Chemical reviews* 1996, **96:**2841-2888.
13. Črešnar B, Petrič Š: **Cytochrome P450 enzymes in the fungal kingdom.** *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 2011, **1814:**29-35.
14. Sutherland JB: **Detoxification of polycyclic aromatic hydrocarbons by fungi.** *Journal of industrial microbiology and biotechnology* 1992, **9:**53-61.
15. Hussain R, Kumari I, Sharma S, Ahmed M, Khan T, Akhter Y: **Catalytic diversity and homotropic allostery of two Cytochrome P450 monooxygenase like proteins from Trichoderma brevicompactum.** *JBIC Journal of Biological Inorganic Chemistry* 2017, **22**.
16. Nelson DR: **Progress in tracing the evolutionary paths of cytochrome P450.** *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 2011, **1814:**14-18.