



Universidade de São Paulo
Instituto de Química



Dicas para realizar projeto Origem de Contigs

João Carlos Setubal

2020

disciplina QBQ2507-IBI5035


Invstigue seus contigs com


- **BLAST** (NCBI)
 - blastn, blastx
- ORFfinder

ORFfinder

- Identificação de ORFs nos contigs
 - <https://www.ncbi.nlm.nih.gov/orffinder/>
- Para cada contig, analise tantas ORFs quanto necessário para chegar a uma conclusão robusta
 - use tamanho mínimo de 300 bp
 - Faça **blastp** nas ORFs promissoras
 - Rode **Interproscan** nas ORFs promissoras (para prever função)

<https://www.ebi.ac.uk/interpro/>

EMBL-EBI  Services Research Training About us

 **InterPro**
Protein sequence analysis & classification

Search InterPro...
Examples: IPR020405, kinase, P51587, PF02932, GO:0007165


[Home](#) [Search](#) [Release notes](#) [Download](#) [About InterPro](#) [Help](#) [Contact](#)

InterPro: protein sequence analysis & classification

InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites. We combine protein signatures from a number of member databases into a single searchable resource, capitalising on their individual strengths to produce a powerful integrated database and diagnostic tool. [Read more about InterPro](#)

Analyse your protein sequence

| |

 **InterPro 58.0**
9th June 2016

Features include:

- New member database: Conserved Domain Database (CDD) version 3.14.
- Updated to HAMAP version 201605.11.
- Integration of 542 new methods from the CDD (318), Pfam (1), SUPERFAMILY (1), HAMAP (43) and PANTHER (179) databases.

| [Read more](#)

IDA 
Domain architecture search

Interproscan 5
[Learn more >>](#)

Documentation

[About InterPro](#): core concepts, update frequency, how to cite, team and consortium members.

[FAQs](#): what are entry types and why are they important, interpreting results, downloading InterPro?

[Web services documentation](#)

Protein focus

[Zika Virus and Microcephaly](#)
Feb 2016

You have probably been as horrified and saddened as me to see the shocking abnormality that affects newborn babies whose ...

[HTML](#) | [All protein focus](#)


Publications




[The InterPro protein families database: the classification resource after 15 years](#)

Our latest paper describing new developments on the InterPro website (*Nucleic Acids Research*, Jan 2015).
[HTML](#) | [PDF \(3,4Mb\)](#) | [All publications](#)


Tools | InterProScan

 InterProScan is a sequence analysis application (nucleotide and protein ...)


Tools | IDA


 The InterPro Domain Architecture (IDA) tool allows you to search the ...

Blog

 [A new InterPro member database](#)
Jun 2016
CDD joins InterPro We are pleased ...

Tweets by @InterProDB

 **InterPro**
@InterProDB
InterProScan 5 (version 5.19-58.0) is now available. For more details please visit github.com/ebi-pf-team/interproscan

 **ebi-pf-team/interproscan**
Contribute to interproscan development.
github.com

Artemis (opcional)

- **browser** de genomas (ou de contigs) que pode ser rodado localmente
- <http://www.sanger.ac.uk/science/tools/artemis>
- Tutorial disponibilizado na página da disciplina

Critérios de identificação de organismos com base em BLAST

- Se a comparação foi de **nucleotídeos**
 - Para espécie, no mínimo **97%** de identidade e no mínimo **100 nt** de alinhamento
 - Entre **90 e 96%** pode-se fazer uma associação com o gênero (“parente”)
 - Abaixo de 90% não é possível fazer afirmações categóricas; **mas é possível especular**
- Se a comparação foi de **aminoácidos**
 - *Em geral não é possível inferir o organismo correspondente, **mas é possível especular, especialmente levando em conta também os resultados de blastn***

Espero ver informações sobre os micro-organismos identificados

- **Pesquisar a literatura** sobre organismos positivamente identificados (mas também os parentes) e apresentar os resultados dessa pesquisa no relatório (com referências)
- Há alguma **coerência** entre as diversas identificações dos diferentes contigs?

CDSs podem dar dicas adicionais

- Há uma **coerência** entre bons hits de blastp e bons hits de blastn?
- Há funções identificadas (dos produtos dos genes) que são coerentes com **o ambiente** da sua hipótese?

Outras ferramentas

- Existem diversas ferramentas de análises genômicas disponíveis na internet, além das aqui mencionadas
- Todos os alunos são encorajados a usar ferramentas não mencionadas e apresentar seus resultados, caso estes ajudem na geração da hipótese final

Apresente suas evidências!

- Exemplos
 - Mostre os alinhamentos principais que ajudaram vc a chegar em sua conclusão
 - Indique
 - % de identidade
 - E-value