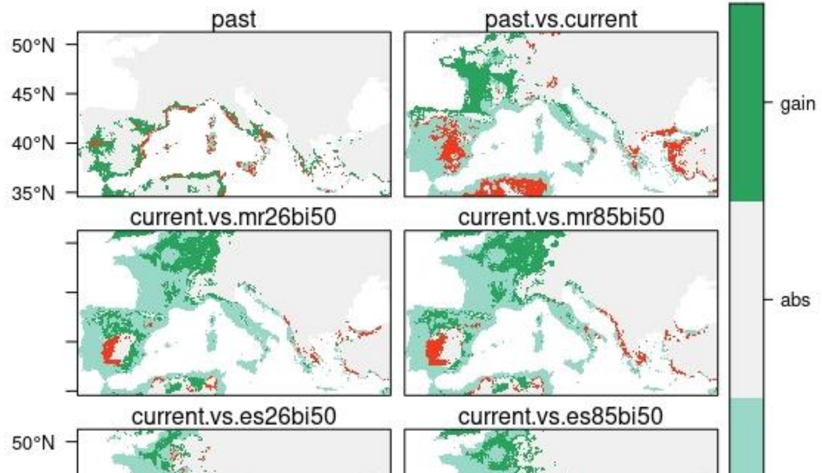
Analisi sequenze

laboratorio Interazioni

0. teachers for a day

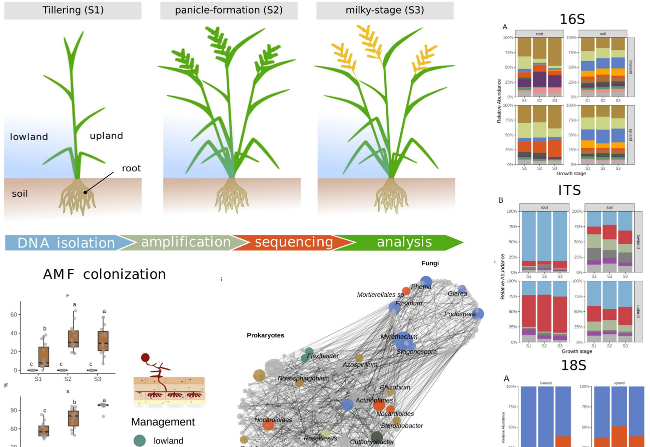
Dr. Martino ADAMO

Postdoc researcher since 2018, I work on plants diversity conservation and orchid mycorrhizas using bio-molecular tools and boring statistical model to study traits influence on species spatial distribution.



Dr. Matteo CHIALVA

He is post-doctoral resercher since 2017 and his research focuses on plant-microbes interactions in model crop species by using multi-omics approaches from transcriptomics to metagenomics. By coupling these tools with systems biology and biostatistics he is interested in linking soil microbiota diversity and functioning to plant responses and ecosystem services.

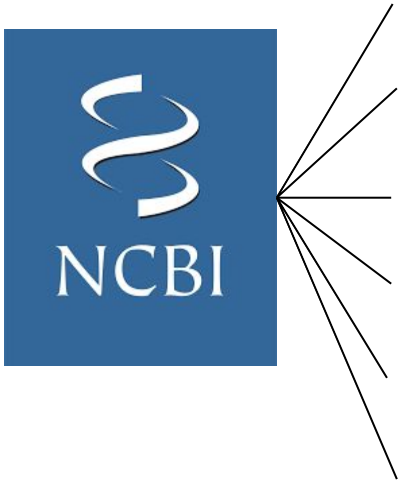


1. Genetic sequence databases

DNA, RNA and protein sequences are stored in dedicated databases. The most known is the National Center for Biotechnology Information (NCBI). European Molecular Biology Laboratory (EMBL) is the European twin. UniProt repository is the most complete protein focused public database. Most of EMBL and UniProt data are automatically uploaded in NCBI.

1. Genetic sequence databases

NCBI (National Center for Biotechnology Information) includes several repositories with different functions. The main focus is to cover all aspects of molecular diversity.



nt :: comprehensive collection of nucleotides from different DBs (87 M sequences)

Protein (nr) :: translated proteins

RefSeq :: representative genomes (including organelles), cDNAs and proteins sequences (high quality)

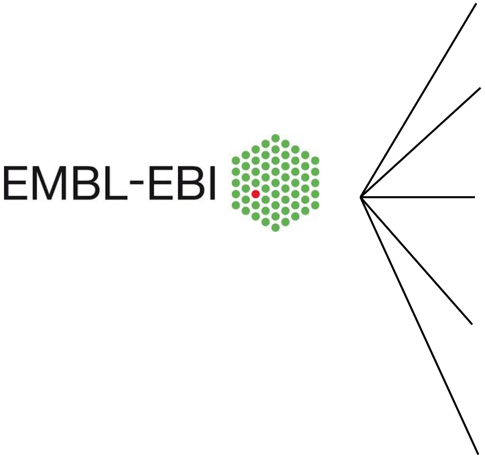
SRA :: raw -omic stuffs

GeneBank :: nucleotide sequence database (NCBI only)

Taxonomy :: taxonomic annotation to all NCBI sequences

1. Genetic sequence databases

EMBL-EBI (European Molecular Biology Laboratory-European Bioinformatics Institute) also includes several repositories. Similar to NCBI database



EMBL-bank :: Nucleotide sequences (DNA/RNA)

SWISS-PROT and TrEMBL :: Protein sequences

ENA :: raw nucleotide sequence data (integrated with NCBI-SRA and DDBJ)

Array Express :: functional genomics data

ENSEMBL Genomes :: Vertebrates, Plants, Metazoa, Fungi, Bacteria

1. How are sequences stored?

Different type of files (mostly simple or compressed **text file**)

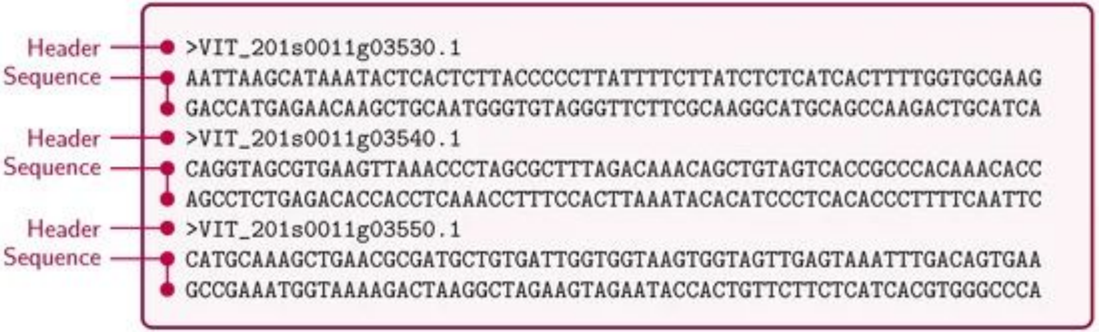
- FASTA (*.fasta/ *.fa/ *.faa/ *.fas): *nucleotides or proteins*

description line (header) →

nucleotide or protein sequence

```
>NG_008679.1:5001-38170 Homo sapiens paired box 6 (PAX6)
ACCCTCTTTTCTTATCATTGACATTTAAACTCTGGGGCAGGTCCTCGCGTAGAACGCGGCTGTCAGATCT
GCCACTTCCCCTGCCGAGCGGCGGTGAGAAGTGTGGGAACCGGCGCTGCCAGGCTCACCTGCCTCCCCGC
CCTCCGCTCCCAGGTAACCGCCCGGGCTCCGGCCCCGGCCCGGCTCGGGGCCCGCGGGGCCTCTCCGCTG
CCAGCGACTGCTGTCCCCAAATCAAAGCCCGCCCCAAGTGGCCCCGGGGCTTGATTTTTGCTTTTAAAAG
GAGGCATACAAAGATGGAAGCGAGTTACTGAGGGAGGGATAGGAAGGGGGGTGGAGGAGGGACTTGTCTT
TGCCGAGTGTGCTCTTCTGCAAAAGTAGCAAAATGTTCCACTCCTAAGAGTGGACTTCCAGTCCGGCCCT
GAGCTGGGAGTAGGGGGCGGGAGTCTGCTGCTGCTGTCTGCTAAAGCCACTCGCGACCGCGAAAAATGCA
GGAGGTGGGGACGCACTTTGCATCCAGACCTCCTCTGCATCGCAGTTCACGACATCCACGCTTGGGAAAG
TCCGTACCCGCGCCTGGAGCGCTTAAAGACACCCTGCCGCGGGTCGGGCGAGGTGCAGCAGAAGTTTCCC
GCGGTTGCAAAGTGCA
```

multiple sequences can be stored in a single file (e.g. a whole genome)

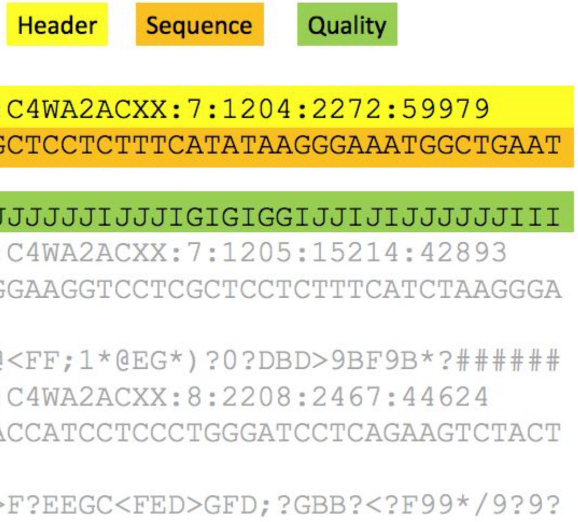


```
>VIT_201s0011g03530.1
AATTAAGCATAAATACTCACTCTTACCCCCTTATTTTCTTATCTCTCATCACTTTTGGTGCGAAG
GACCATGAGAACAAGCTGCAATGGGTGTAGGGTTCTTCGCAAGGCATGCAGCCAAGACTGCATCA
>VIT_201s0011g03540.1
CAGGTAGCGTGAAGTTAAACCCTAGCGCTTTAGACAAACAGCTGTAGTCACCGCCCACAAACACC
AGCCTCTGAGACACCACCTCAAACCTTTCCACTTAAATACACATCCCTCACACCCTTTTCAATTC
>VIT_201s0011g03550.1
CATGCAAAGCTGAACGCGATGCTGTGATTGGTGGTAAGTGGTAGTTGAGTAAATTTGACAGTGAA
GCCGAAATGGTAAAAGACTAAGGCTAGAAGTAGAATACCACTGTTCTTCTCATCACGTGGGCCCA
```

1. How are sequences stored?

- FASTA with quality (*.fastq/ *.fq/*.fastq.gz): *nucleotides from NGS sequencing*

millions of sequences can be stored in a single file

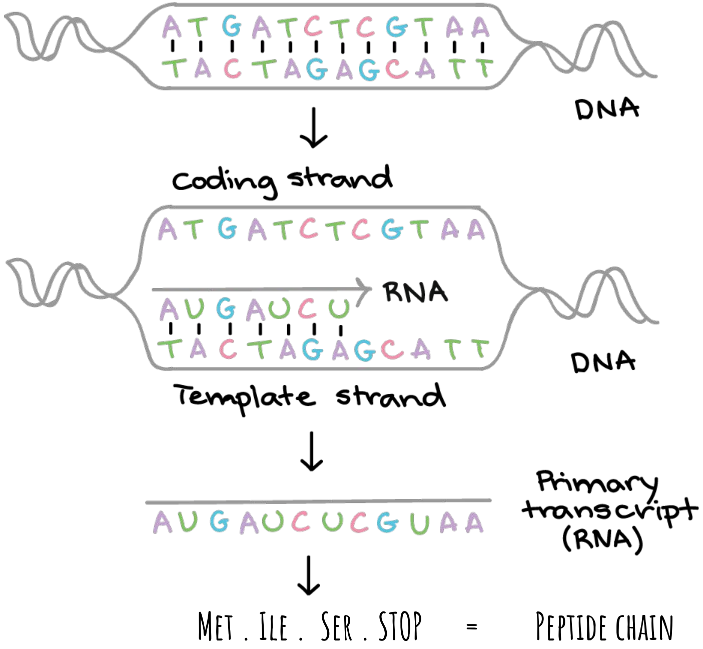


```
@HWI-ST227:389:C4WA2ACXX:7:1204:2272:59979
GGAGGAAGGTCCTCGCTCCTCTTTCATATAAGGGAAATGGCTGAAT
+
FFFFHHHHHHJIJJJJJJJJIJJJIGIGIGGIJJIJIJJJJJJIII
@HWI-ST227:389:C4WA2ACXX:7:1205:15214:42893
GAGGATCCCAGGGAGGAAGGTCCTCGCTCCTCTTTCATCTAAGGGA
+
12BAFB?A:3<AE1@<FF;1*@EG*)?0?DBD>9BF9B*?######
@HWI-ST227:389:C4WA2ACXX:8:2208:2467:44624
AAAGAGGAGAGAGGACCATCCTCCCTGGGATCCTCAGAAGTCTACT
+
BDDA:DB?2AA@FC>F?EEGC<FED>GFD;?GBB?<?F99*/9?9?
```

- Other formats
  - nucleotides electropherograms from sanger sequencing :: *.ab1
  - alignments :: *.bam/ *.sam)
  - variants (SNPs/INDELs) :: *.vcf
  - NCBI sequence file formats :: *.gbf/ *.gbk

2. Sequences similarity

A sequence could be a DNA, an RNA (usually mRNA), or a protein sequence:



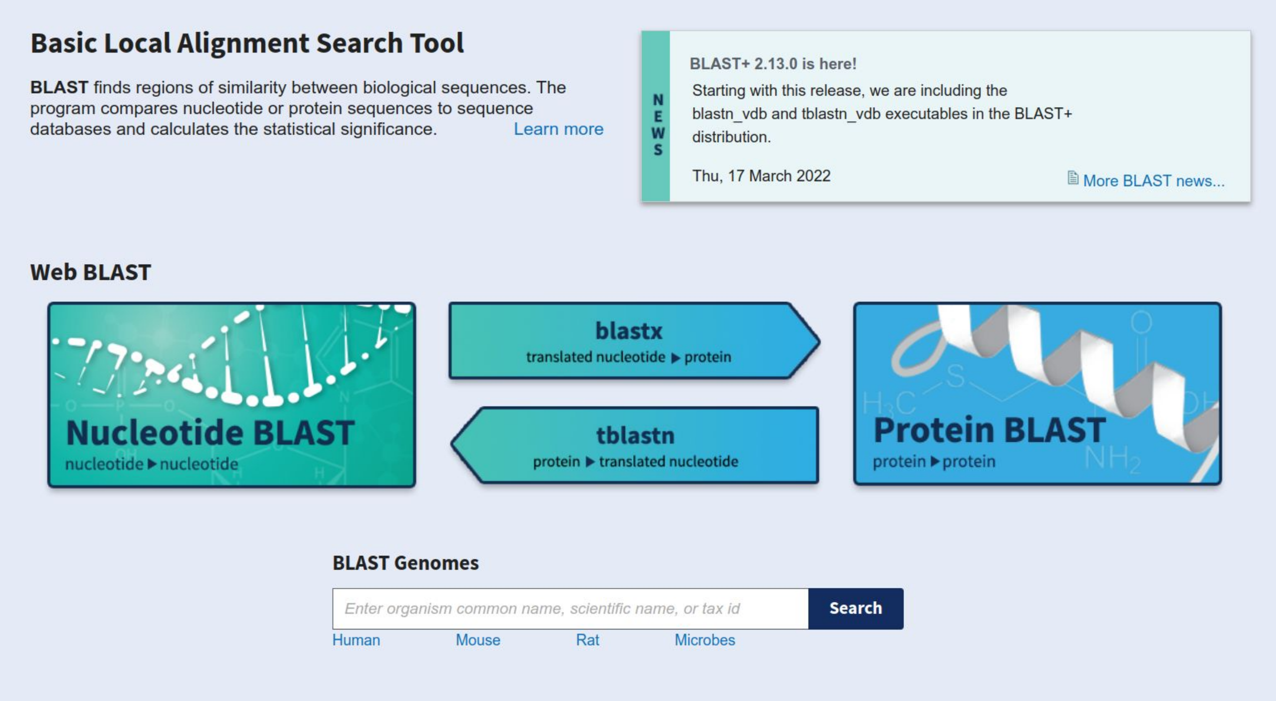
We can statistically compare sequences to measure how many similar they are.

- 2 seqs comparison is intuitive
- multiple seqs comparison is an alignment
- comparison of a sequence *vs* a whole database is known as a BLAST

(**B**asic **L**ocal **A**lignment **S**earch **T**ool)

2. Sequences similarity

https://blast.ncbi.nlm.nih.gov/Blast.cgi



Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. Learn more

NEWS

BLAST+ 2.13.0 is here!

Starting with this release, we are including the blastn_vdb and tblastn_vdb executables in the BLAST+ distribution.

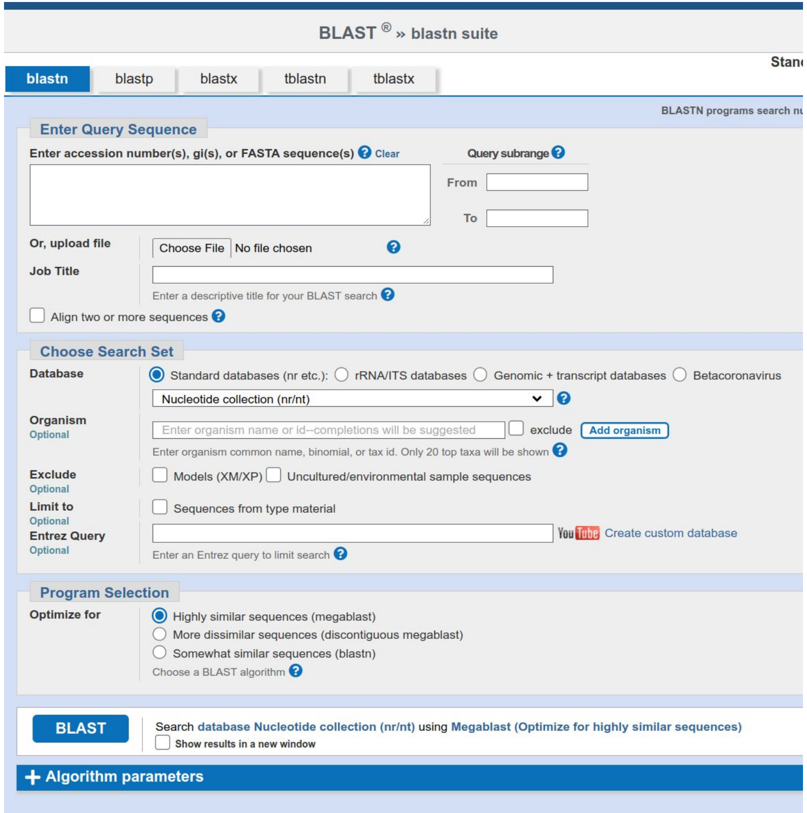
Thu, 17 March 2022

More BLAST news...

you can blast both nucleotide and protein sequences

blast could include translation from nucleotide to protein and *vice-versa*

2. Sequences similarity



BLAST ® » blastn suite

BLASTN programs search nu

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) ? Clear

Query subrange ?

From

To

Or, upload file: Choose File No file chosen ?

Job Title

Enter a descriptive title for your BLAST search ?

□ Align two or more sequences ?

Choose Search Set

Database: ● Standard databases (nr etc.): ○ rRNA/ITS databases ○ Genomic + transcript databases ○ Betacoronavirus

Nucleotide collection (nr/nt) ?

Organism Optional: Enter organism name or id--completions will be suggested □ exclude Add organism

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown ?

Exclude Optional: □ Models (XM/XP) □ Uncultured/environmental sample sequences

Limit to Optional: □ Sequences from type material

Entrez Query Optional: YouTube Create custom database

Enter an Entrez query to limit search ?

Program Selection

Optimize for:
● Highly similar sequences (megablast)
○ More dissimilar sequences (discontiguous megablast)
○ Somewhat similar sequences (blastn)

Choose a BLAST algorithm ?

BLAST — Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)

□ Show results in a new window

+ Algorithm parameters

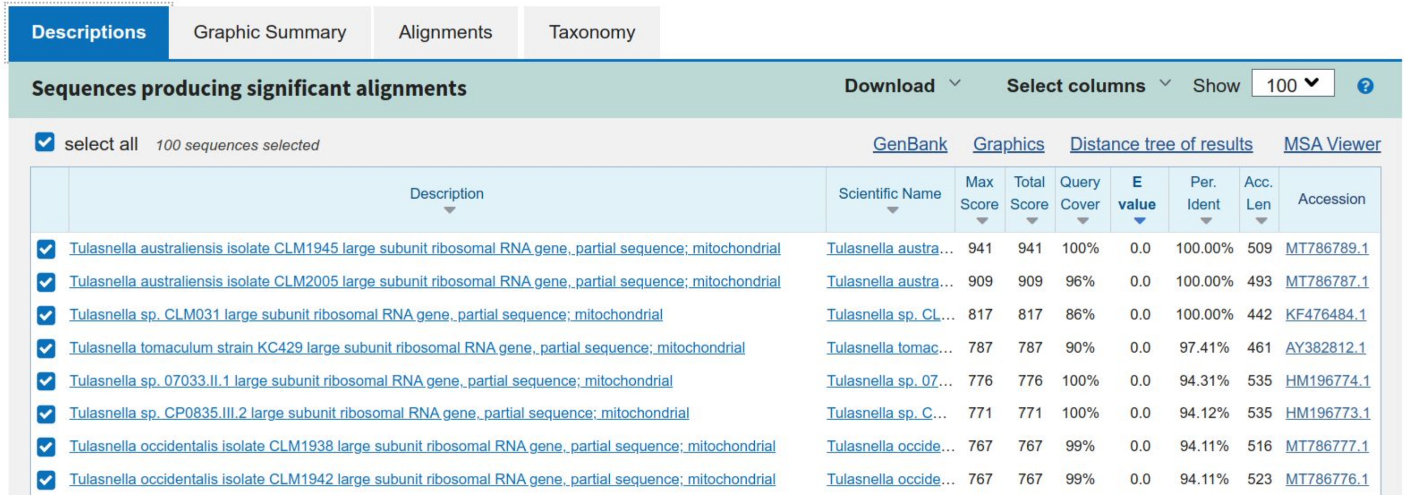
A blast needs few mandatory parameters …

- a sequence (the **QUERY**)
- database selection (the **Subject**)

and too many options

- taxonomic trim
- … and many others

2. Sequences similarity



Sequences producing significant alignments

☑ select all *100 sequences selected*

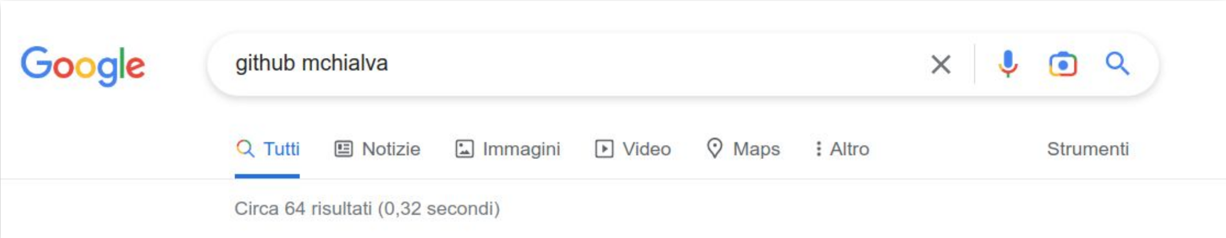
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Tulasnella australiensis isolate CLM1945 large subunit ribosomal RNA gene, partial sequence; mitochondrial	Tulasnella austra...	941	941	100%	0.0	100.00%	509	MT786789.1
☑	Tulasnella australiensis isolate CLM2005 large subunit ribosomal RNA gene, partial sequence; mitochondrial	Tulasnella austra...	909	909	96%	0.0	100.00%	493	MT786787.1
☑	Tulasnella sp. CLM031 large subunit ribosomal RNA gene, partial sequence; mitochondrial	Tulasnella sp. CL...	817	817	86%	0.0	100.00%	442	KF476484.1
☑	Tulasnella tomaculum strain KC429 large subunit ribosomal RNA gene, partial sequence; mitochondrial	Tulasnella tomac...	787	787	90%	0.0	97.41%	461	AY382812.1
☑	Tulasnella sp. 07033.II.1 large subunit ribosomal RNA gene, partial sequence; mitochondrial	Tulasnella sp. 07...	776	776	100%	0.0	94.31%	535	HM196774.1
☑	Tulasnella sp. CP0835.III.2 large subunit ribosomal RNA gene, partial sequence; mitochondrial	Tulasnella sp. C...	771	771	100%	0.0	94.12%	535	HM196773.1
☑	Tulasnella occidentalis isolate CLM1938 large subunit ribosomal RNA gene, partial sequence; mitochondrial	Tulasnella occide...	767	767	99%	0.0	94.11%	516	MT786777.1
☑	Tulasnella occidentalis isolate CLM1942 large subunit ribosomal RNA gene, partial sequence; mitochondrial	Tulasnella occide...	767	767	99%	0.0	94.11%	523	MT786776.1

E-value: expect value is a sort of “significance” of the hits. The lower the E-value, the better the hit. The E-value is dependent on the length of the query sequence and the size of the database.

Percentual identity: percentage of identical bp between query and subject

Accession: univocal sequence identifier click here!

GitHub

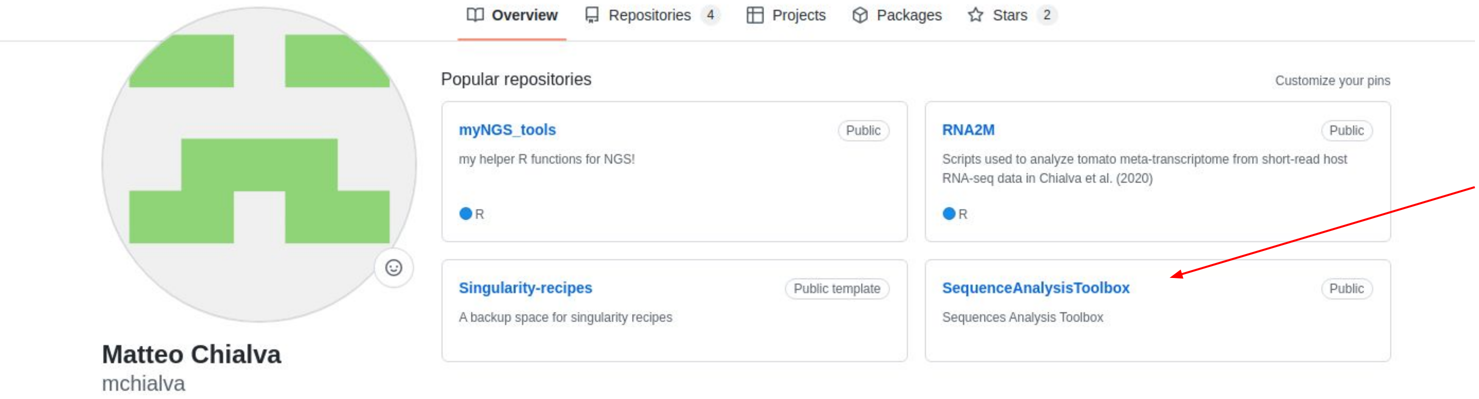


Google

github mchialva

Circa 64 risultati (0,32 secondi)

https://github.com/mchialva



Popular repositories

Customize your pins

myNGS_tools Public
my helper R functions for NGS!
● R

RNA2M Public
Scripts used to analyze tomato meta-transcriptome from short-read host RNA-seq data in Chialva et al. (2020)
● R

Singularity-recipes Public template
A backup space for singularity recipes

SequenceAnalysisToolbox Public
Sequences Analysis Toolbox

Matteo Chialva
mchialva

2. Sequence similarity: a real BLAST example

Task: Given an unknown rRNA marker sequence, blast it and infer its taxonomy

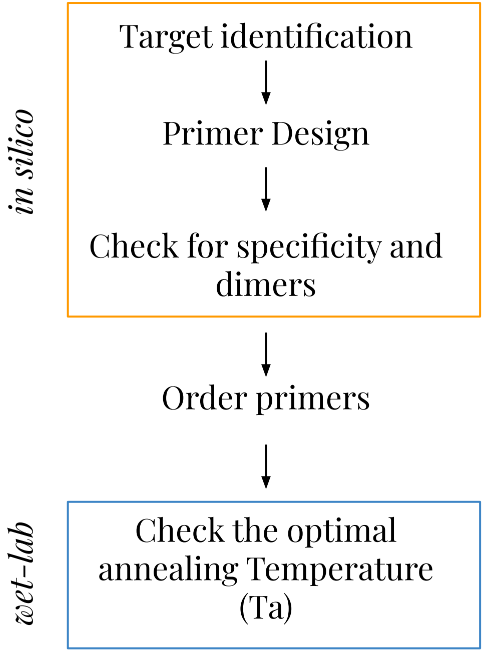
- when isolating microorganisms *in vitro,* a molecular characterization is often required (and integrated with other phenotypic traits) to assign isolates to species/genus.

Dataset: 16S_rRNA.fasta

Activity: BLAST the given nucleotides sequence

3. Primer design

What are primers?
Short synthetic oligonucleotides which targets a portion of DNA (or cDNA/eDNA..)



in silico

Target identification

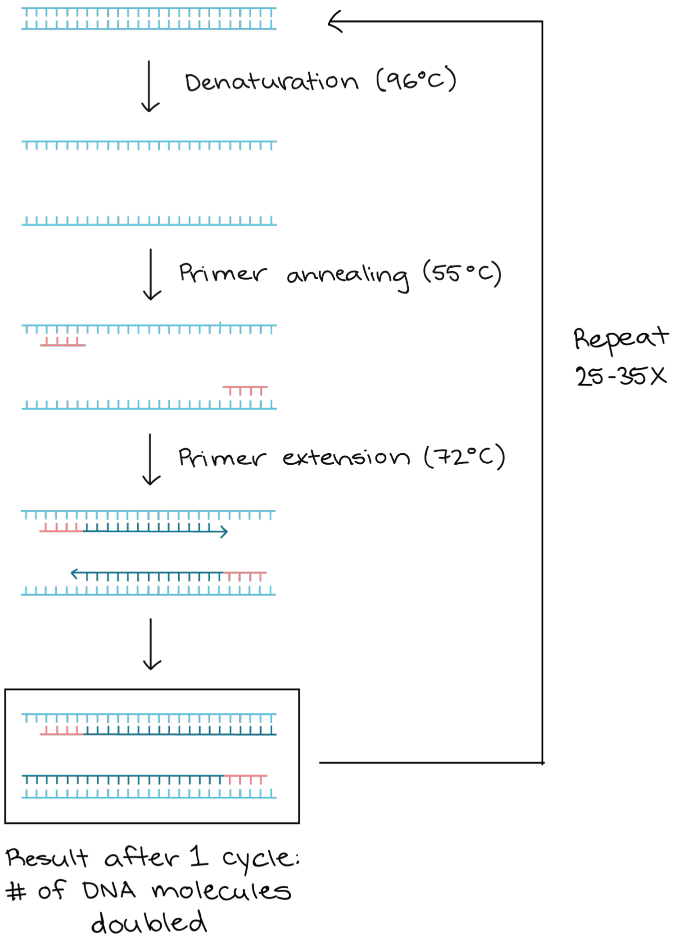
Primer Design

Check for specificity and dimers

Order primers

wet-lab

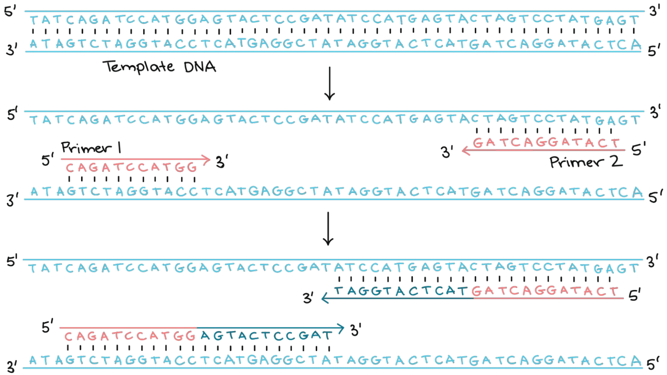
Check the optimal annealing Temperature (Ta)



3. Primer design: oligonucleotides features

Primers and amplicon size

- Avoid too short (lack of specificity) or too long (lack of annealing efficiency) primers
    - optimal at 18-24 nucelotides
    - annealing efficiency is proportional to primer length
- final amplicon size should not exceed 1000-2000 bp (standard Taq polymerases)


5'
TATCAGATCCATGGAGTACTCCGATATCCATGAGTACTAGTCCTATGAGT
3'
3'
ATAGTCTAGGTACCTCATGAGGCTATAGGTACTCATGATCAGGATACTCA
5'
Template DNA
5'
TATCAGATCCATGGAGTACTCCGATATCCATGAGTACTAGTCCTATGAGT
3'
Primer 1
5'
CAGATCCATGG
3'
3'
GATCAGGATACT
5'
Primer 2
3'
ATAGTCTAGGTACCTCATGAGGCTATAGGTACTCATGATCAGGATACTCA
5'
5'
TATCAGATCCATGGAGTACTCCGATATCCATGAGTACTAGTCCTATGAGT
3'
3'
TAGGTACTCATGATCAGGATACT
5'
5'
CAGATCCATGGAGTACTCCGAT
3'
3'
ATAGTCTAGGTACCTCATGAGGCTATAGGTACTCATGATCAGGATACTCA
5'

GC content

primer GC content influence its melting temperature (T_m)*

$$Tm = 2^{\circ}C \times (A+T) + 4^{\circ}C \times (G+C) \text{ [Wallace formula]}$$

- both primers should have a GC content of 40-60%
- The two primers should have a similar Tm (delta of 3-5°C maximum)
- The Tm should be within the range of 55-72°C (optimal at 60°C)
- primers should have a terminal G or C (G/C) clamp to ensure stable primer pairing to their target (GC have a stronger hydrogen bonding)

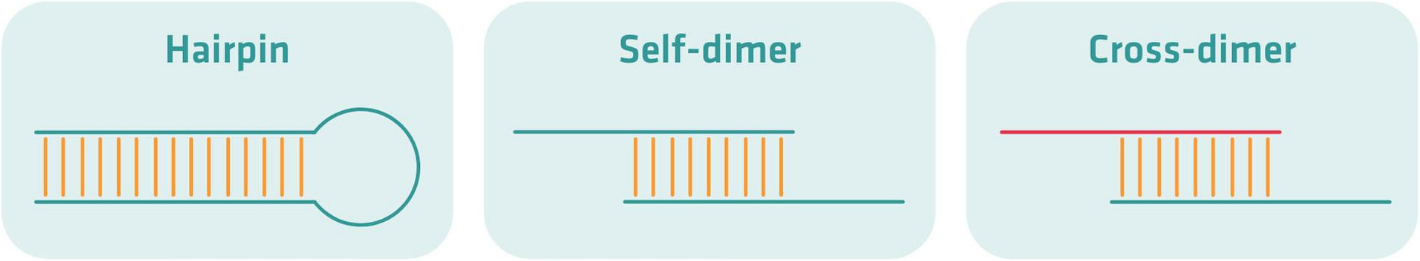
*the temperature at which half of the primers dissociate from the template DNA

3. Primer design: oligonucleotides features

Primers specificity

A primer pair must be complementary only to the target sequence (in most cases…)

- off-targets: primers are homologus to nucelotides outside the region of interest (your target gene)



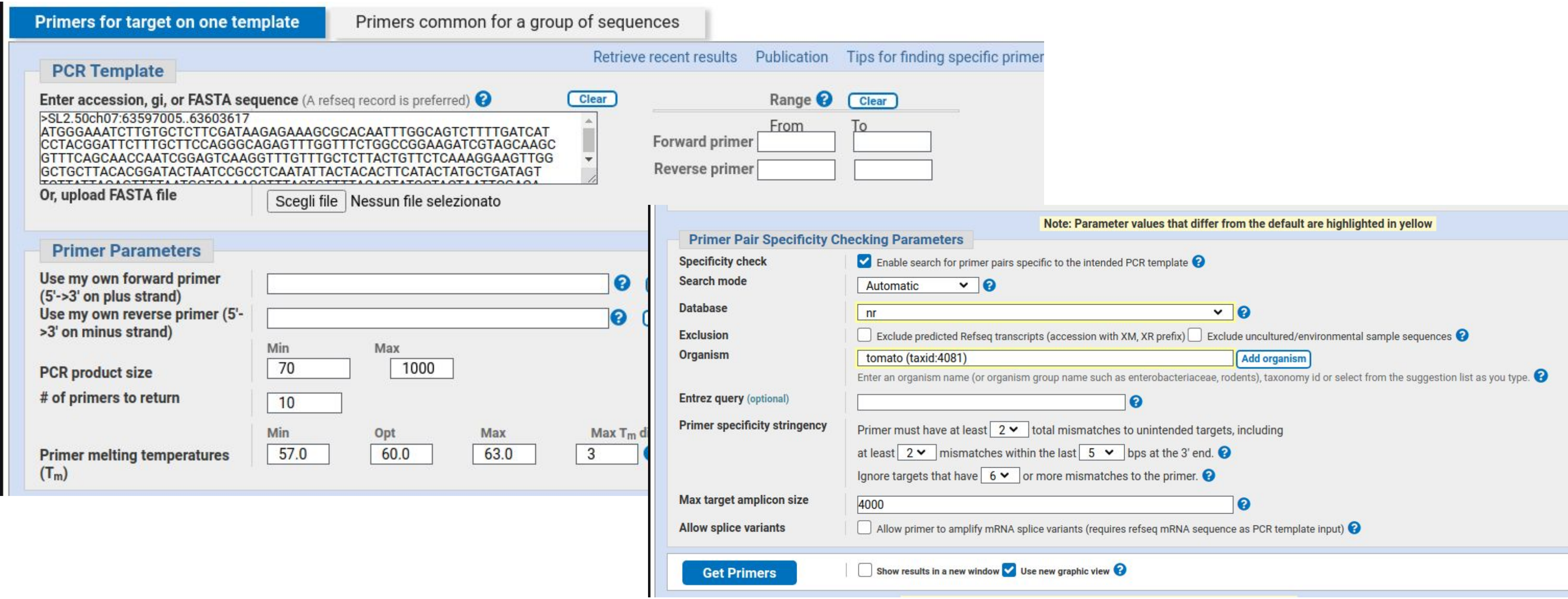
INTRA-primer homology

INTER-primer homology

Effect: decrease in PCR yield!

3. Primer design: PrimerBLAST web-tool

https://www.ncbi.nlm.nih.gov/tools/primer-blast/



Primers for target on one template
Primers common for a group of sequences
Retrieve recent results
Publication
Tips for finding specific primer

PCR Template

Enter accession, gi, or FASTA sequence (A refseq record is preferred) Clear

```
>SL2.50ch07:63597005..63603617
ATGGGAAATCTTGTGCTCTTCGATAAGAGAAAGCGCACAATTTGGCAGTCTTTTGATCAT
CCTACGGATTCTTTGCTTCCAGGGCAGAGTTTGGTTTCTGGCCGGAAGATCGTAGCAAGC
GTTTCAGCAACCAATCGGAGTCAAGGTTTGTTTGCTCTTACTGTTCTCAAAGGAAGTTGG
GCTGCTTACACGGATACTAATCCGCCTCAATATTACTACACTTCATACTATGCTGATAGT
```

Range Clear

	From	To
Forward primer		
Reverse primer		

Or, upload FASTA file — Scegli file Nessun file selezionato

Primer Parameters

Use my own forward primer (5'->3' on plus strand):

Use my own reverse primer (5'->3' on minus strand):

	Min	Opt	Max	Max Tm d
PCR product size	70		1000	
# of primers to return	10			
Primer melting temperatures (T_m)	57.0	60.0	63.0	3

Note: Parameter values that differ from the default are highlighted in yellow

Primer Pair Specificity Checking Parameters

Specificity check: ☑ Enable search for primer pairs specific to the intended PCR template

Search mode: Automatic

Database: nr

Exclusion: □ Exclude predicted Refseq transcripts (accession with XM, XR prefix) □ Exclude uncultured/environmental sample sequences

Organism: tomato (taxid:4081) Add organism

Enter an organism name (or organism group name such as enterobacteriaceae, rodents), taxonomy id or select from the suggestion list as you type.

Entrez query (optional):

Primer specificity stringency: Primer must have at least 2 total mismatches to unintended targets, including at least 2 mismatches within the last 5 bps at the 3' end. Ignore targets that have 6 or more mismatches to the primer.

Max target amplicon size: 4000

Allow splice variants: □ Allow primer to amplify mRNA splice variants (requires refseq mRNA sequence as PCR template input)

Get Primers □ Show results in a new window ☑ Use new graphic view

3. How-to: a real primers design example

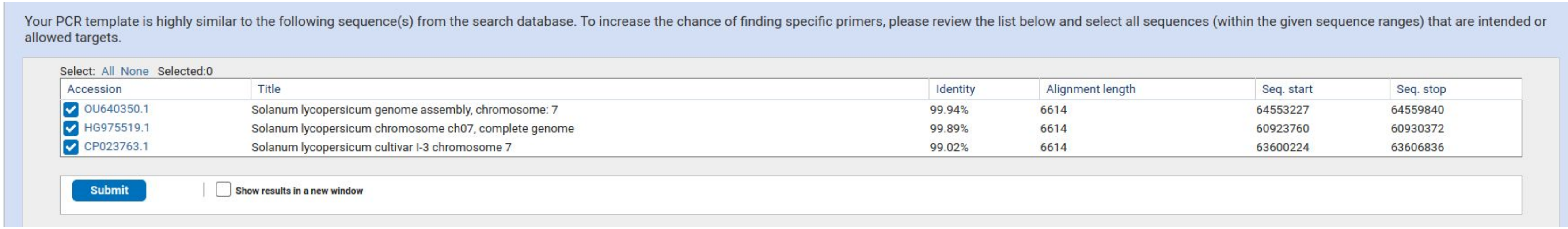
Task: Check presence/absence of a disease-resistance gene in different tomato genotype

- I-3 gene confers resistance against race 3 *Fusarium oxysporum* f. sp. *lycopersici* (Fol) in tomato

Dataset: Tomato_gene1.fasta

Activity: Design standard PCR primers using primerBLAST web-tool and pick the most specific primer pairs

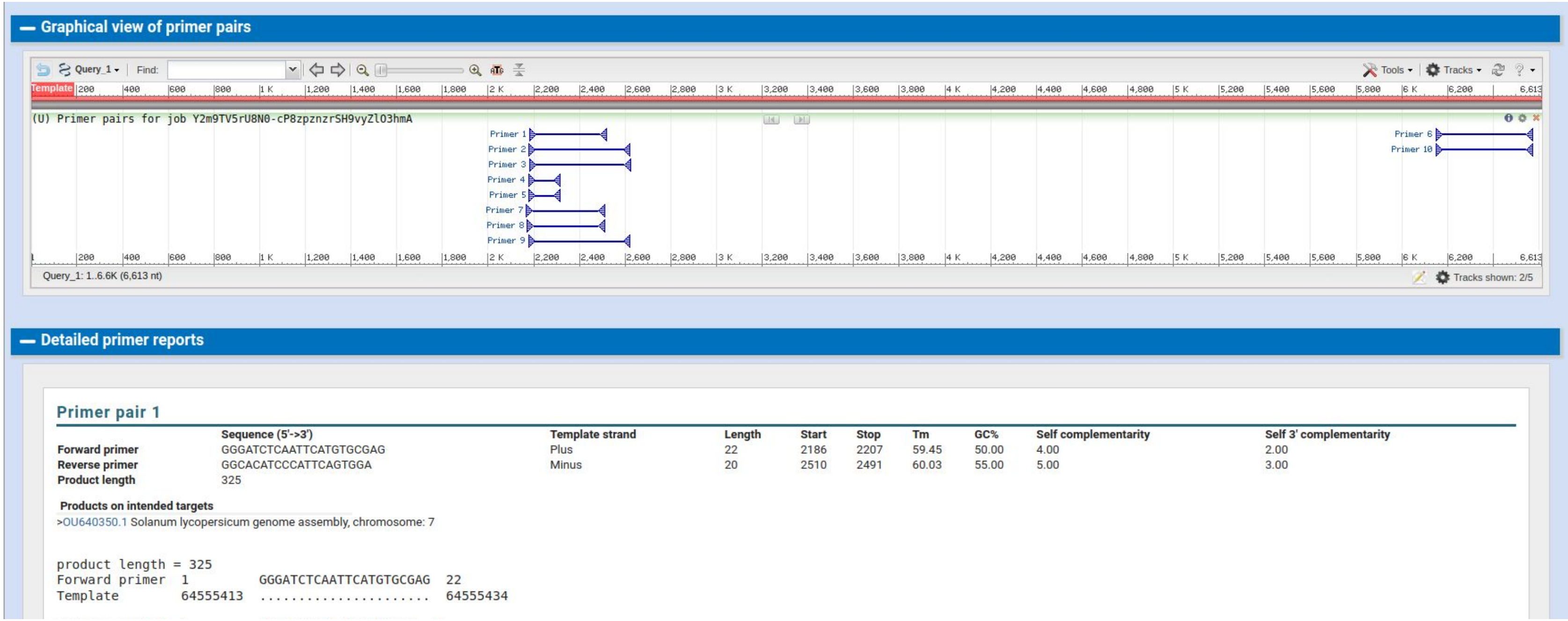
3. Primer design: Primer-BLAST web-tool



Your PCR template is highly similar to the following sequence(s) from the search database. To increase the chance of finding specific primers, please review the list below and select all sequences (within the given sequence ranges) that are intended or allowed targets.

Select: All None Selected:0

Accession	Title	Identity	Alignment length	Seq. start	Seq. stop
☑ OU640350.1	Solanum lycopersicum genome assembly, chromosome: 7	99.94%	6614	64553227	64559840
☑ HG975519.1	Solanum lycopersicum chromosome ch07, complete genome	99.89%	6614	60923760	60930372
☑ CP023763.1	Solanum lycopersicum cultivar I-3 chromosome 7	99.02%	6614	63600224	63606836



— Graphical view of primer pairs

— Detailed primer reports

Primer pair 1

	Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GGGATCTCAATTCATGTGCGAG	Plus	22	2186	2207	59.45	50.00	4.00	2.00
Reverse primer	GGCACATCCCATTCAGTGGA	Minus	20	2510	2491	60.03	55.00	5.00	3.00
Product length	325								

Products on intended targets

>OU640350.1 Solanum lycopersicum genome assembly, chromosome: 7

```
product length = 325
Forward primer  1         GGGATCTCAATTCATGTGCGAG  22
Template        64555413  ......................  64555434
```

4. How to build a Phylogeny

What is Pylogeny? The study of evolutionary relationships of organisms through the comparative analysis of traits (including molecular sequences, but not only..) with the aim to reconstruct genealogical relationships between organisms or gene/proteins families

Different steps:

1) Select sequences by taxonomy/orthology including the outgroup:
   it depends on the biological question posed. Often it is the most time-consuming task in phylogeny

2) Align sequences and generate the multiple sequences alignment file (MSA)
   Different computational strategies (i.e. different software available). MSA should be high quality as possible and often require manual curation

3) Select the best evolutionary model
   Select the model which better describes the MSA i.e. the way and the rate nucleotides/amino acids change across taxa and the

4) Infer Phylogeny
   Different algorithms available: distance-based models (neighbour-joining [NJ]) or heuristics models (maximum likelihood [ML] or Bayesian models)

5) Plot and annotate the tree and draw biological/evolutionary conclusions

3. How-to: a complete phylogenetic reconstruction workflow

Task: Reconstruct phylogeny of 13 different mammalian species using K-casein exon 4 DNA.

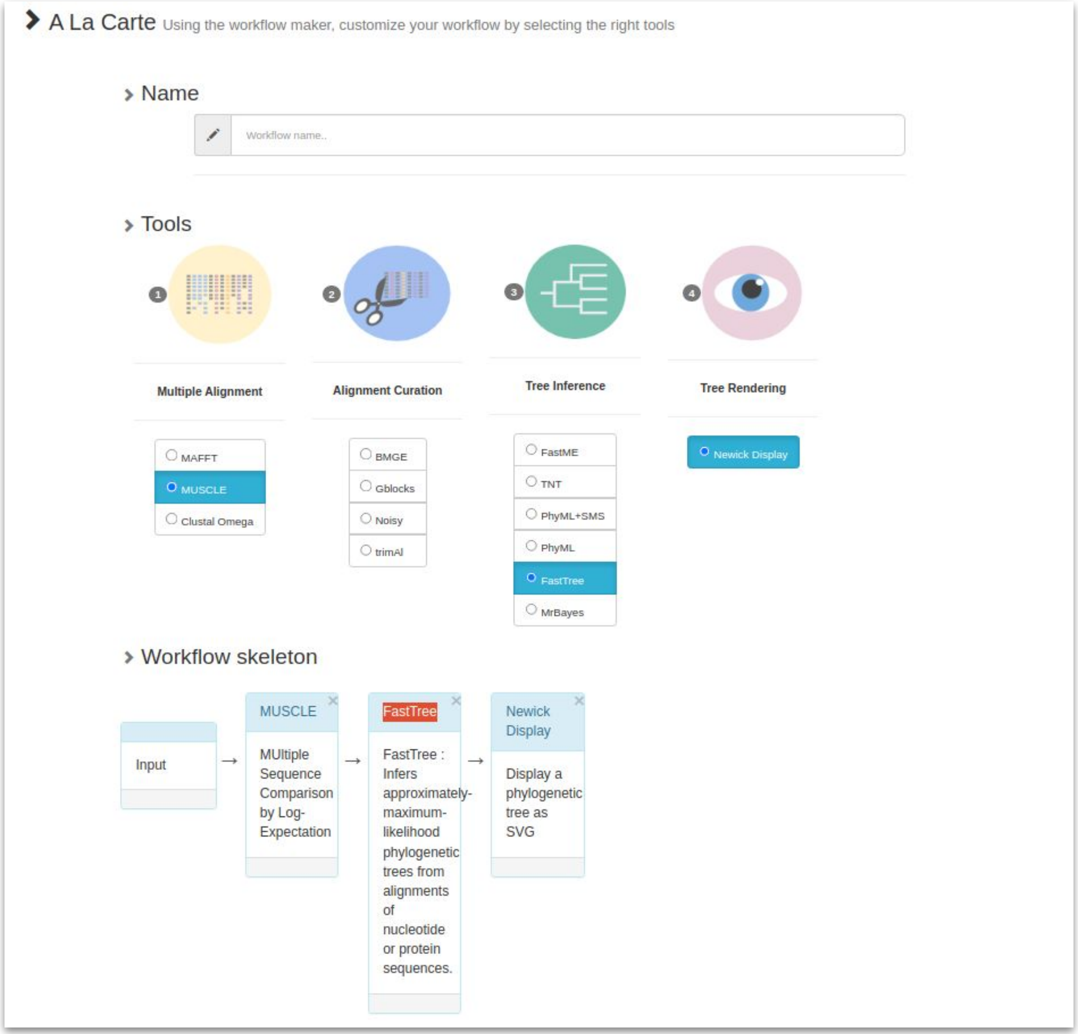
Dataset: KCAS_13_mammals.fasta

from Gatesy et al. (1999)

Activity: Use ngphylogeny.fr web-service to perform all the phylogenetic reconstruction step from alignment to the tree plot.

4. how to build a phylogeny

https://ngphylogeny.fr/



A La Carte Using the workflow maker, customize your workflow by selecting the right tools

Name

Workflow name..

Tools

1 Multiple Alignment

- ○ MAFFT
- ● MUSCLE
- ○ Clustal Omega

2 Alignment Curation

- ○ BMGE
- ○ Gblocks
- ○ Noisy
- ○ trimAl

3 Tree Inference

- ○ FastME
- ○ TNT
- ○ PhyML+SMS
- ○ PhyML
- ● FastTree
- ○ MrBayes

4 Tree Rendering

- ● Newick Display

Workflow skeleton

Input → MUSCLE → FastTree → Newick Display

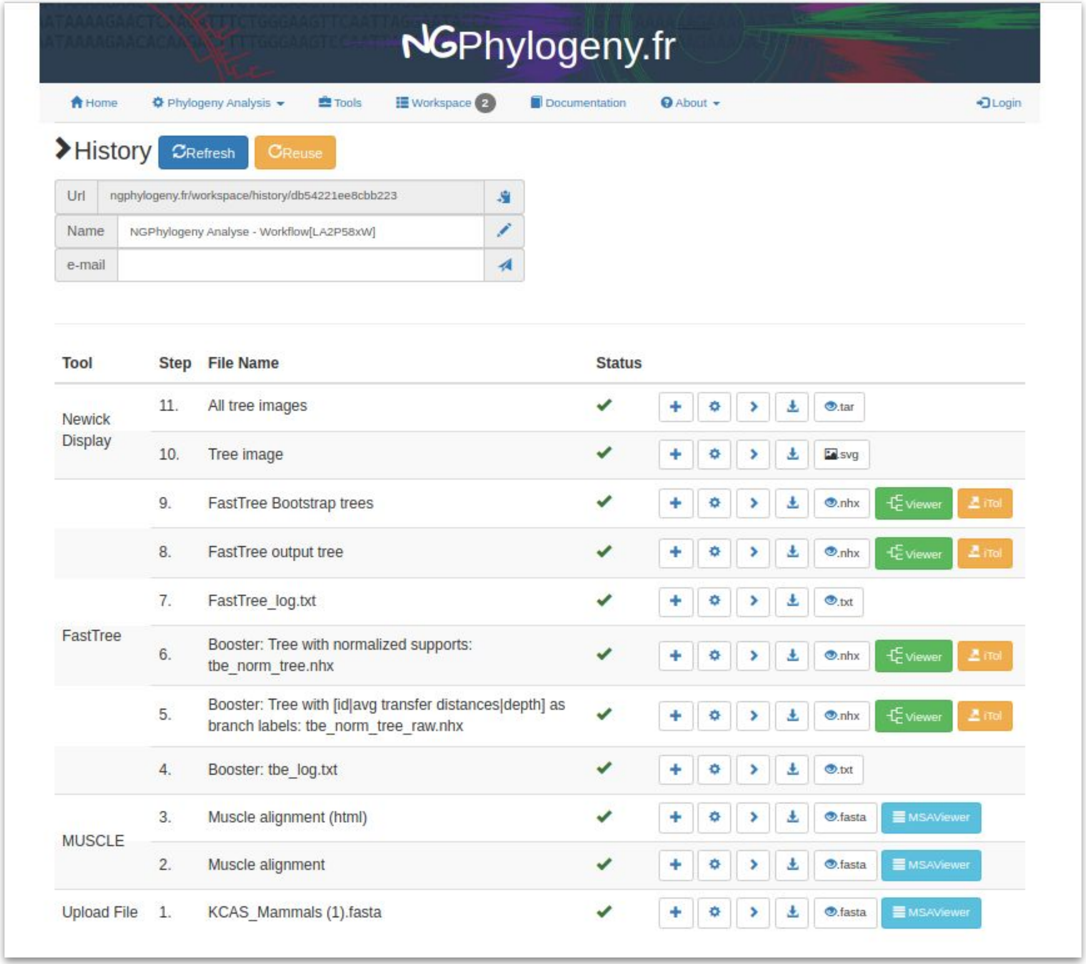
MUSCLE: MUltiple Sequence Comparison by Log-Expectation

FastTree: FastTree : Infers approximately-maximum-likelihood phylogenetic trees from alignments of nucleotide or protein sequences.

Newick Display: Display a phylogenetic tree as SVG

4. how to build a phylogeny

https://ngphylogeny.fr/



History

Url ngphylogeny.fr/workspace/history/db54221ee8cbb223

Name NGPhylogeny Analyse - Workflow[LA2P58xW]

e-mail

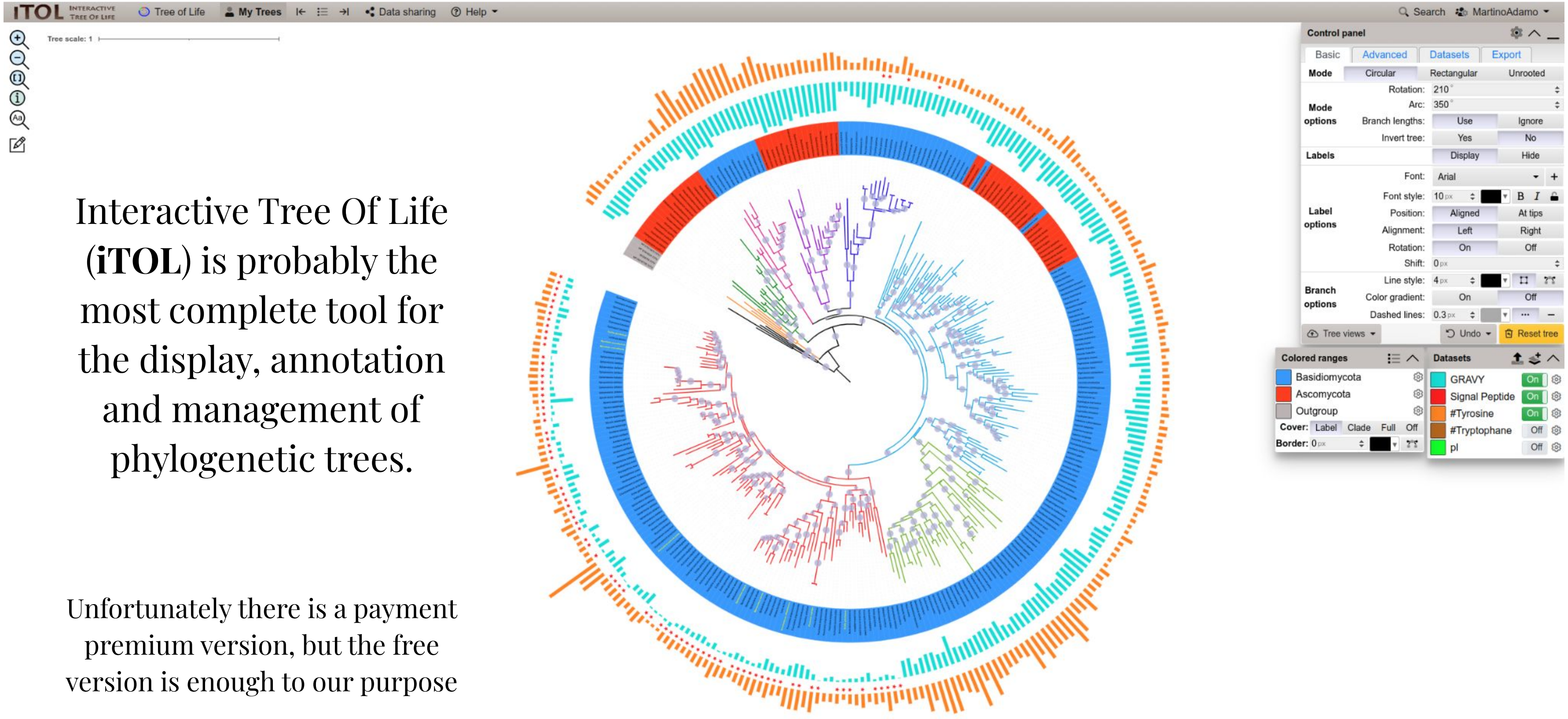
Tool	Step	File Name	Status	
Newick Display	11.	All tree images	✓	.tar
	10.	Tree image	✓	svg
FastTree	9.	FastTree Bootstrap trees	✓	.nhx Viewer iTol
	8.	FastTree output tree	✓	.nhx Viewer iTol
	7.	FastTree_log.txt	✓	.txt
	6.	Booster: Tree with normalized supports: tbe_norm_tree.nhx	✓	.nhx Viewer iTol
	5.	Booster: Tree with [id|avg transfer distances|depth] as branch labels: tbe_norm_tree_raw.nhx	✓	.nhx Viewer iTol
	4.	Booster: tbe_log.txt	✓	.txt
MUSCLE	3.	Muscle alignment (html)	✓	.fasta MSAViewer
	2.	Muscle alignment	✓	.fasta MSAViewer
Upload File	1.	KCAS_Mammals (1).fasta	✓	.fasta MSAViewer

4. how to build a phylogeny

https://itol.embl.de/

Interactive Tree Of Life (**iTOL**) is probably the most complete tool for the display, annotation and management of phylogenetic trees.

Unfortunately there is a payment premium version, but the free version is enough to our purpose



Mode: [✓ Circular] Rectangular Unrooted

Mode options
Rotation: 210°
Arc: 350°
Branch lengths: [✓ Use] Ignore
Invert tree: Yes [✓ No]

Labels: [✓ Display] Hide

Label options
Font: Arial
Font style: 10 px
Position: [✓ Aligned] At tips
Alignment: [✓ Left] Right
Rotation: [✓ On] Off
Shift: 0 px

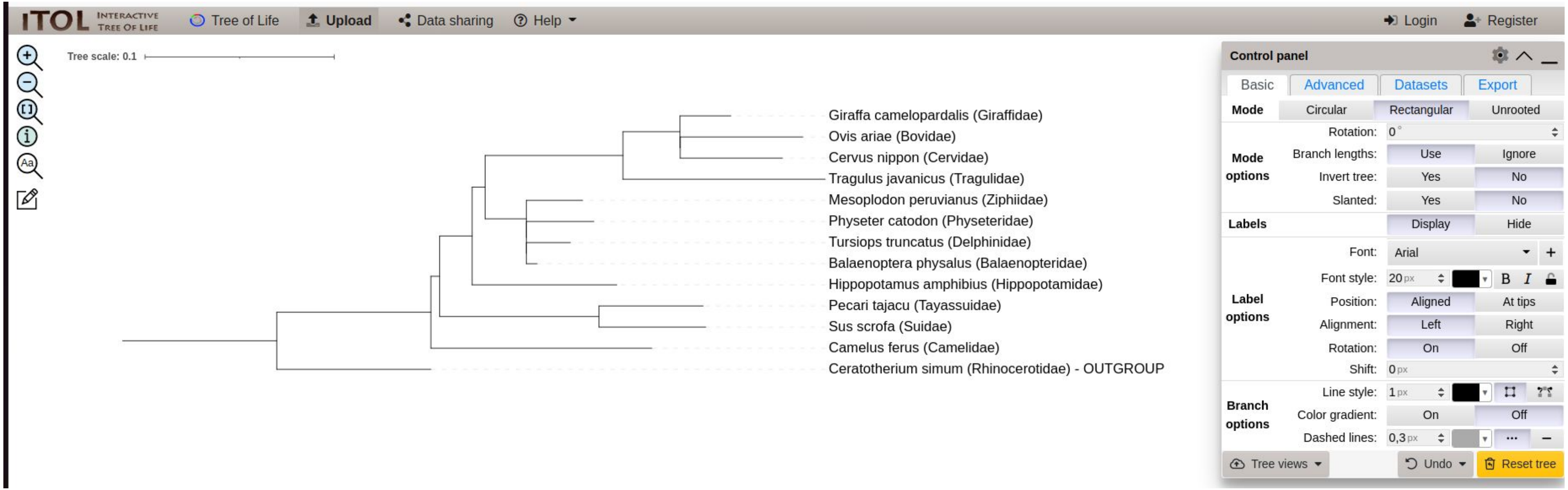
Branch options
Line style: 4 px
Color gradient: On [✓ Off]
Dashed lines: 0.3 px

Colored ranges
🟦 Basidiomycota
🟥 Ascomycota
⬜ Outgroup
Cover: [✓ Label] Clade Full Off
Border: 0 px

Datasets
🟦 GRAVY On
🟥 Signal Peptide On
🟧 #Tyrosine On
🟫 #Tryptophane Off
🟩 pl Off

4. how to build a phylogeny

https://itol.embl.de/


ITOL Interactive Tree Of Life
Tree of Life
Upload
Data sharing
Help
Login
Register
Tree scale: 0.1
Giraffa camelopardalis (Giraffidae)
Ovis ariae (Bovidae)
Cervus nippon (Cervidae)
Tragulus javanicus (Tragulidae)
Mesoplodon peruvianus (Ziphiidae)
Physeter catodon (Physeteridae)
Tursiops truncatus (Delphinidae)
Balaenoptera physalus (Balaenopteridae)
Hippopotamus amphibius (Hippopotamidae)
Pecari tajacu (Tayassuidae)
Sus scrofa (Suidae)
Camelus ferus (Camelidae)
Ceratotherium simum (Rhinocerotidae) - OUTGROUP
Control panel
Basic
Advanced
Datasets
Export
Tree views
Undo
Reset tree

Mode: Circular | Rectangular | Unrooted

Mode options
Rotation: 0°
Branch lengths: Use | Ignore
Invert tree: Yes | No
Slanted: Yes | No

Labels: Display | Hide

Label options
Font: Arial
Font style: 20 px B I
Position: Aligned | At tips
Alignment: Left | Right
Rotation: On | Off
Shift: 0 px

Branch options
Line style: 1 px
Color gradient: On | Off
Dashed lines: 0,3 px

5. Useful online and off-line resources

MEGA [https://www.megasoftware.net/] sequence editing, alignment, NJ & ML, tree editing

SeaView [https://doua.prabi.fr/software/seaview] alignment, alignment edit, NJ & ML, tree editing

PhyML [http://www.atgc-montpellier.fr/phyml/] model selection, ML

MrBayes [https://nbisweden.github.io/MrBayes/] bayesian phylogeny

RaxML [https://cme.h-its.org/exelixis/web/software/raxml/] model selection, ML multicore

FastTree [http://www.microbesonline.org/fasttree/] very fast and inaccurate ML

FigTree [http://tree.bio.ed.ac.uk/software/figtree/] desktop cool tree editing