Reverse Vaccinology for lactococcosis (Bioinformatics workflow)

PRIMA
PARTNERSHIP FOR RESEARCH AND INNOVATION
IN THE MEDITERRANEAN AREA

Vaccinology in the era of genomics

The concept of reverse vaccinology

- Disease prevention is the most effective approach for health
- Development of vaccines has proved a milestone in prevention of diseases

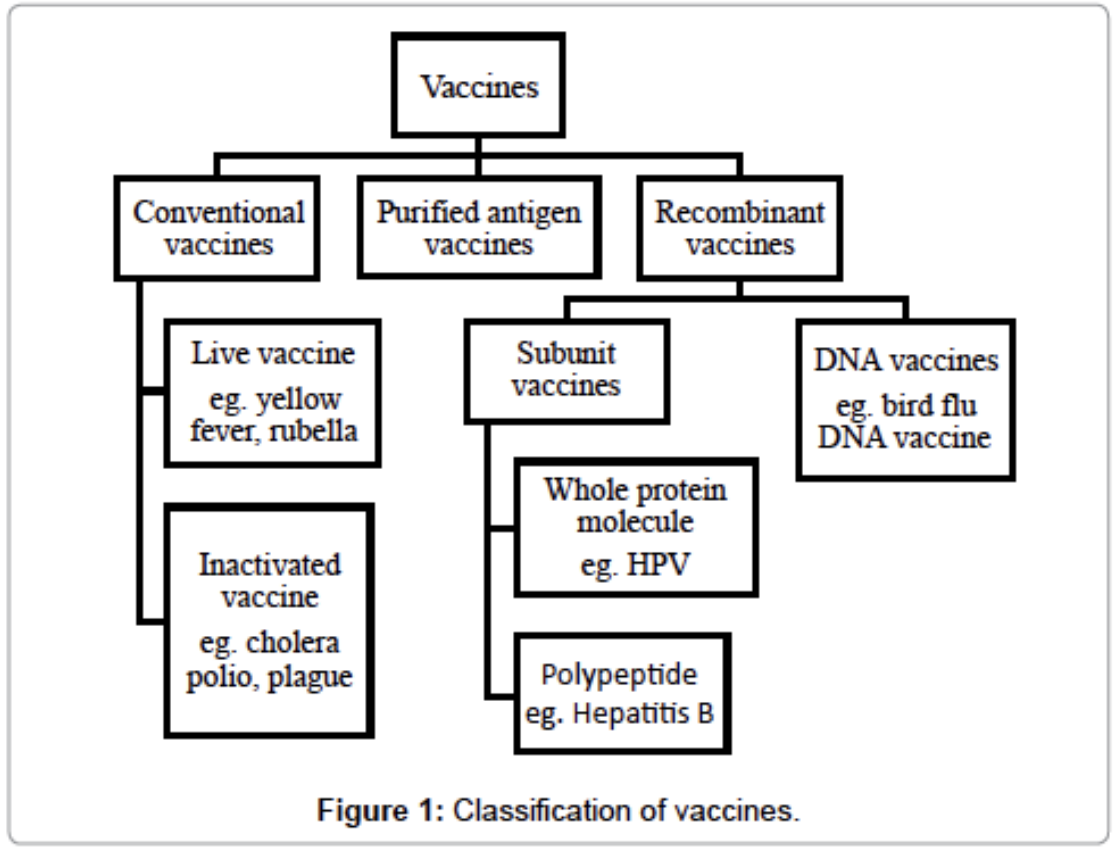
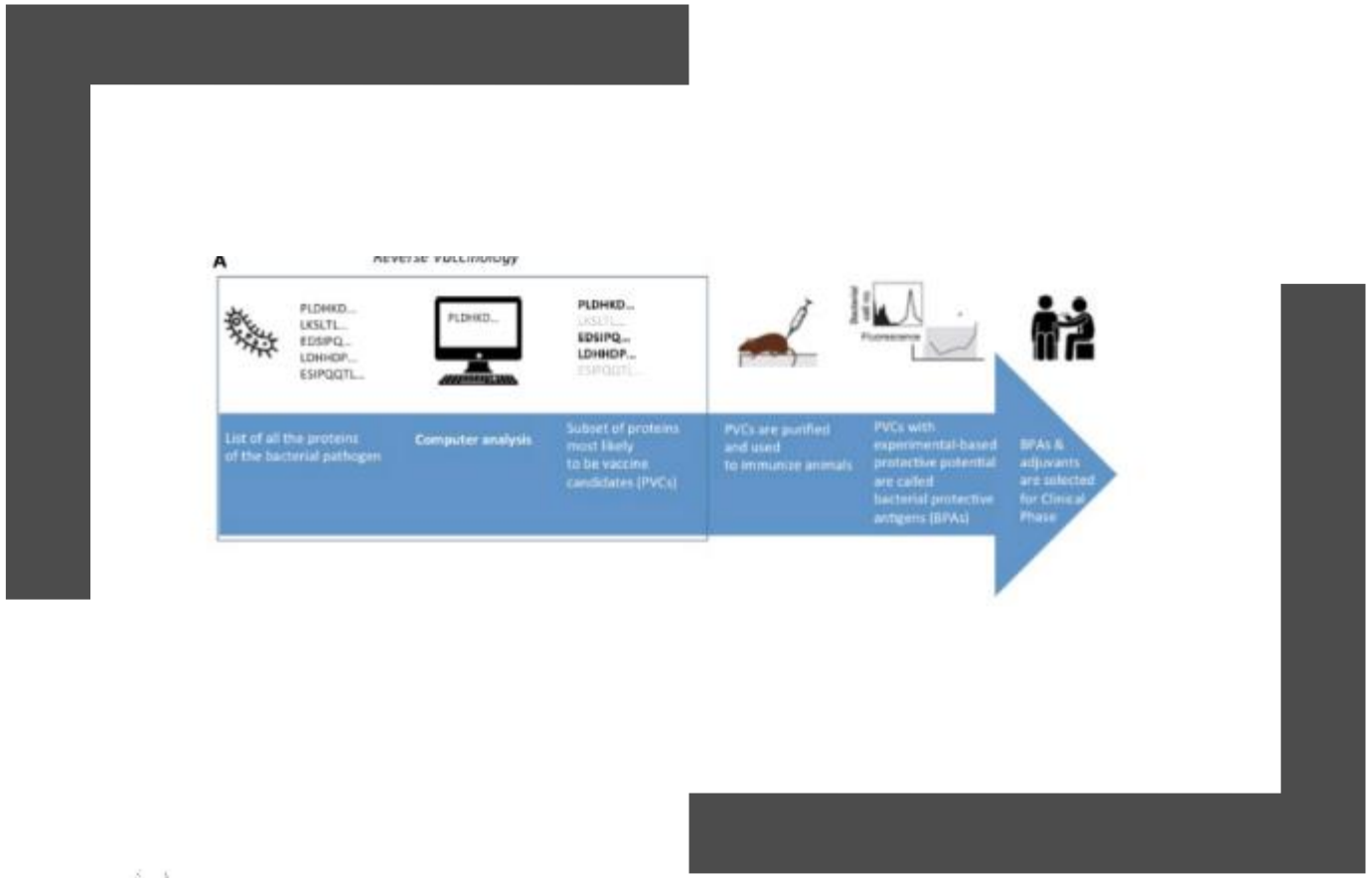


Figure 1: Classification of vaccines.



- Reverse Vaccinology (RV) is a genome-based approach developed for the first time in early 1990's by Rappuoli to identify protein vaccine candidates against meningococcus

- RV approach adopts computerized screening of protein sequences from the pathogen as the first step of the process, to select a subset of promising antigens, as potential vaccine candidates
- offers two main advantages compared to traditional vaccine development approaches: (i) identification of candidate antigens without the need to grow the pathogen (ii) identification of any antigen independently by its purified quantity to be suitable for vaccine testing.

The use of genomic information with aid of computer for the preparation of vaccines without culturing microorganism is known as reverse vaccinology (RV).

- RV explores the protein coding sequences that can be used as a potential target for vaccine preparation

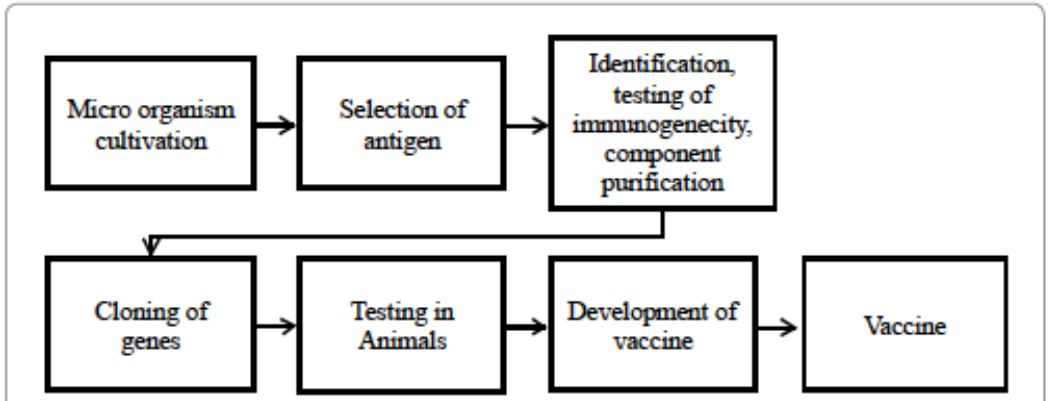


Figure 2: Pipeline for the development of the vaccines by conventional means.

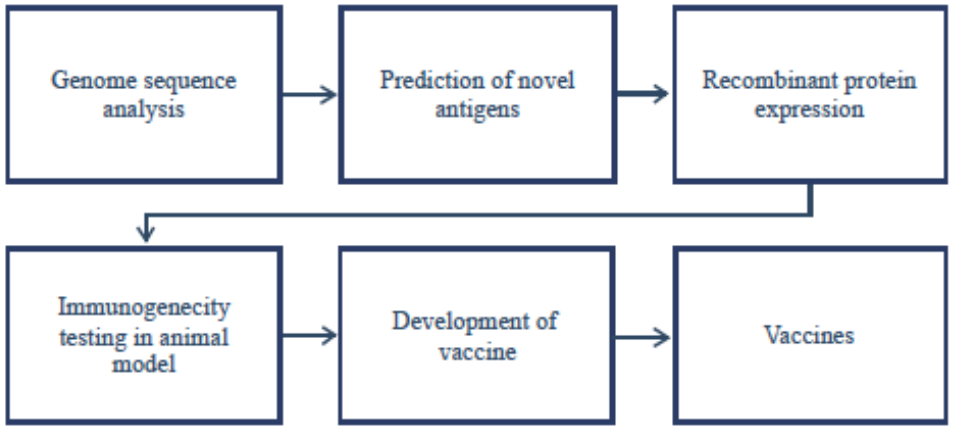


Figure 3: Pipeline for the development of the vaccines by reverse vaccinology.

- The genome sequences provide at once all protein antigens that the pathogen can express at any time. What follows:
- 1. Genome sequences
- 2. Computer analysis
- 3. Prediction of epitope/ antigen
- 4. construction of candidate vaccine

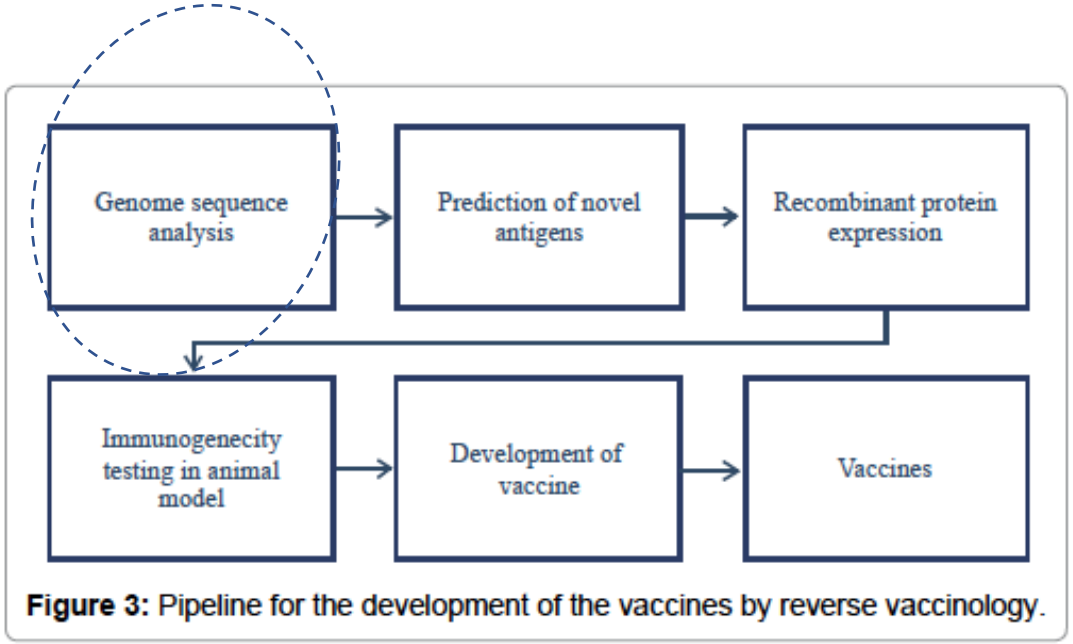


Figure 3: Pipeline for the development of the vaccines by reverse vaccinology.

Role of Epitope Prediction in Reverse Vaccinology

The original idea behind reverse vaccinology:

To **screen in silico** the entire genome of a pathogen **to identify epitopes** in proteins which are suitable vaccine targets, for example, proteins that are predicted to be surface exposed and are well conserved between strains **.**

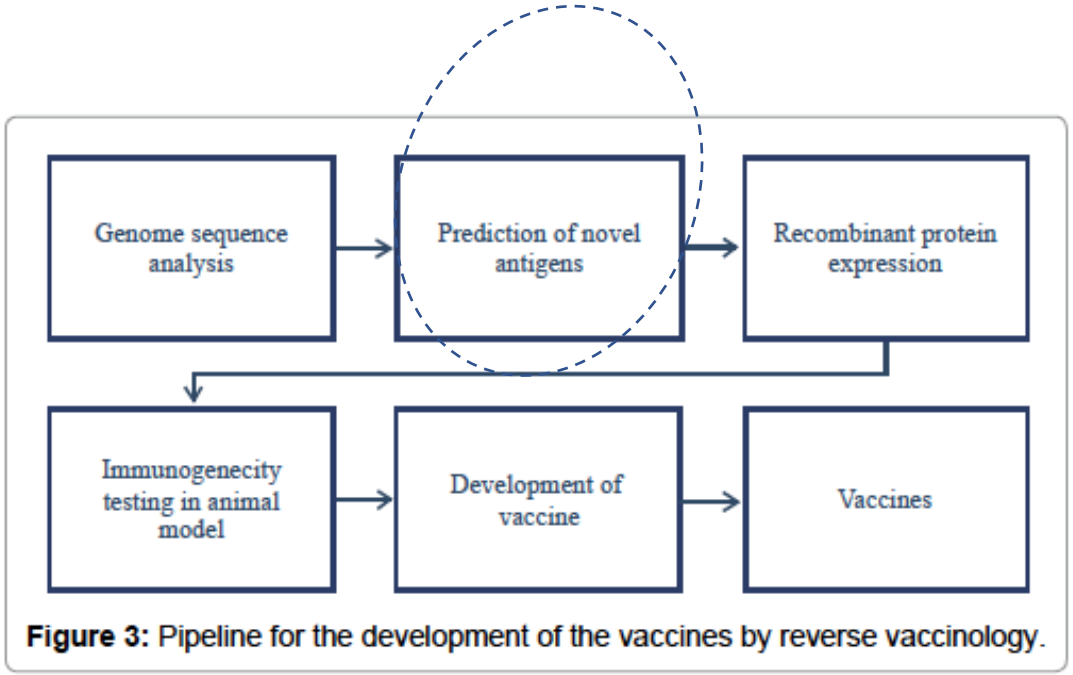


Figure 3: Pipeline for the development of the vaccines by reverse vaccinology.

- **Role of Epitope Prediction in Reverse Vaccinology**

How can we screen **in silico for new epitopes?**

- Improved sequencing technologies have greatly increased the number of genomes available and reverse vaccinology has evolved to include the analysis of several genomes within a species **(pan-genomic analysis)**
- Between closely related species-(**comparative-genomic analysis)**
- or between pathogenic and commensal members of the same species **(comparative/subtractive genomics)**

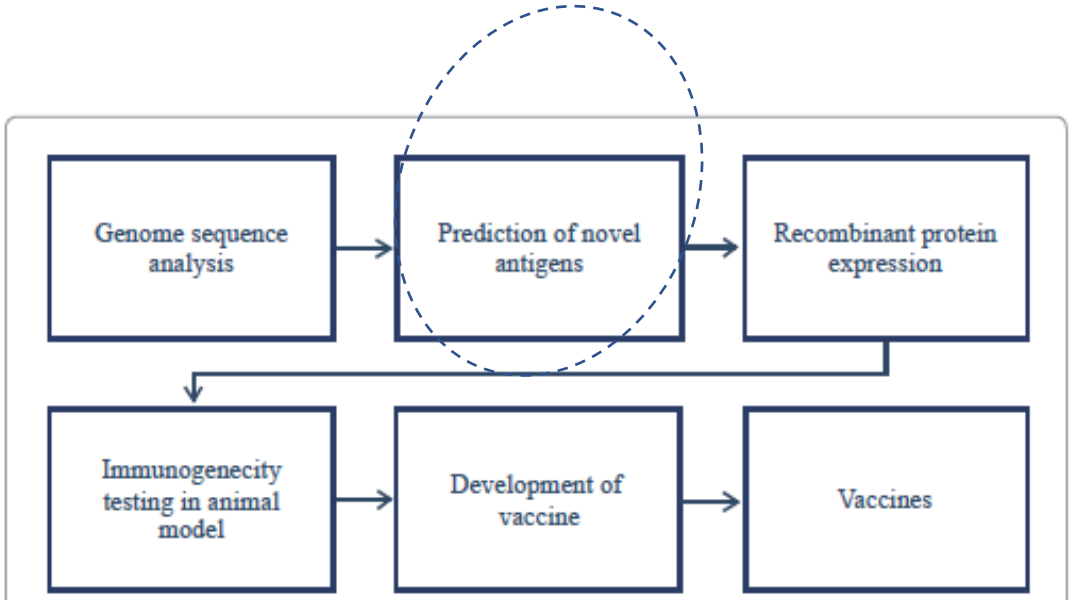


Figure 3: Pipeline for the development of the vaccines by reverse vaccinology.

Role of Epitope Prediction in Reverse Vaccinology

- The identification of antigens to induce immunogenic responses is crucial for the development of an effective vaccine. An epitope is an antigenic determinant that plays an important role in immunity of an organism.
- What kind of epitopes we try to identify?
- **T-Cell Epitope Mapping and Prediction.** T cell recognizes the antigenic peptides only when they are presented by MHC I or II, with the help of the CD4 and CD8 molecules. The target is to predict epitopes against a panel of MHC class I and/or class II alleles.
- **B-Cell Epitope Mapping and Prediction**

The B-cell epitopes are defined by a specific surface region of an antigenic protein and may be divided into two different types of epitopes: linear epitopes and conformational epitopes.

The linear epitopes are short peptides while conformational epitopes composed of amino acid folded in 3- dimensional protein structure

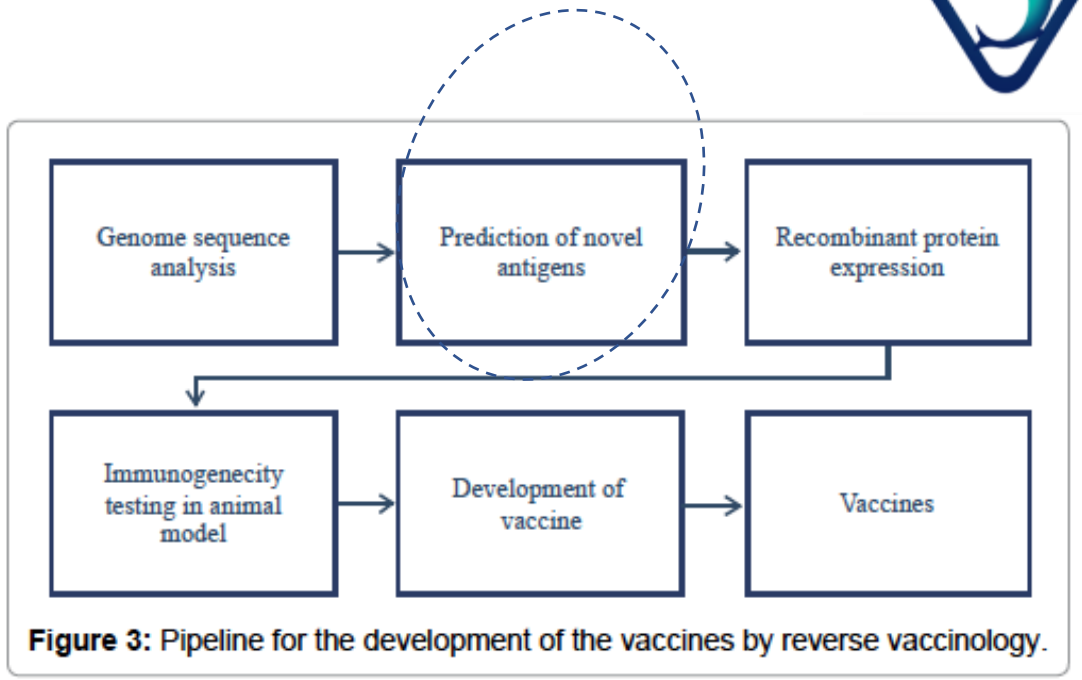


Figure 3: Pipeline for the development of the vaccines by reverse vaccinology.

Reverse Vaccinology for lactococcosis (Bioinformatics workflow)

A case study:

In this work, we are introducing an integrated framework that combines immuno-informatics approaches, bioinformatics tools, and supervised machine learning-based tools for vaccine construction against lactococcosis in supertrouts

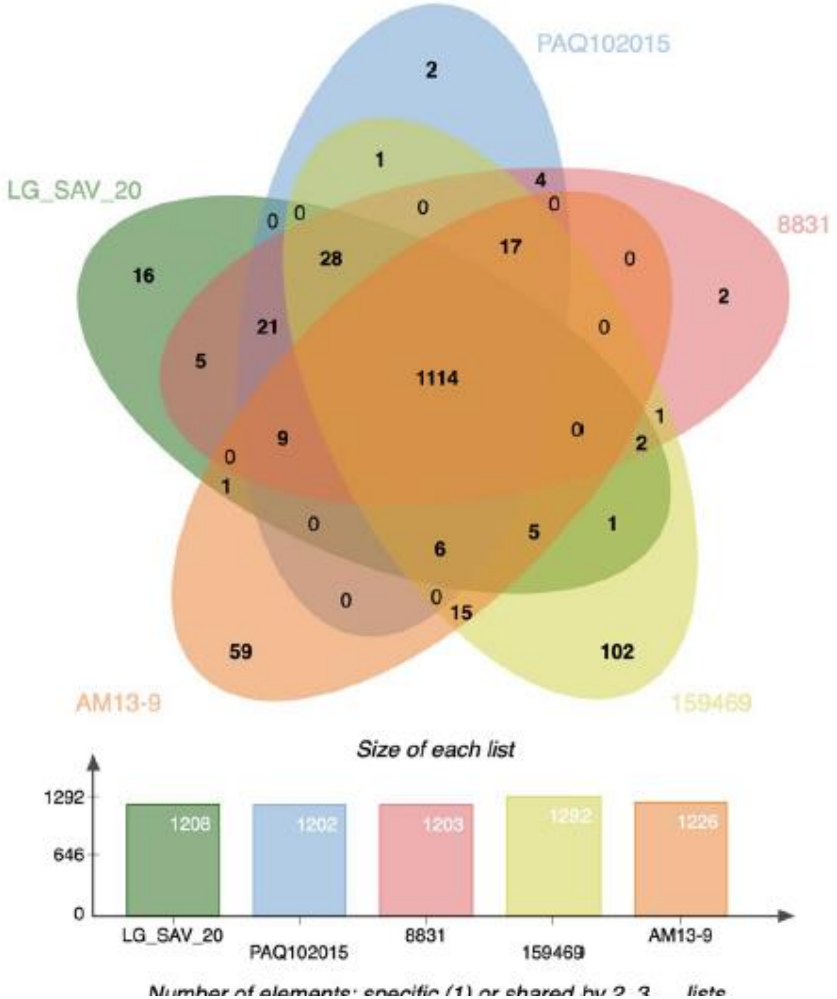
To develop a recombinant subunit-protein vaccine against lactococcosis to be administered by immersion.

Objectives

- Immunogenic protein identification
- recombinant-protein production
- experimental trial for vaccine efficacy

What we know for lactococcosis in rainbow trout

- *Lactococcus garvieae*, a highly diverse species, is considered to be the causative agent of the disease
- However, genomic information provides new insights into the role of novel species: *Lactococcus petauri* as an etiological agent of lactococcosis



First steps towards polypeptide vaccine construction

1. Protein selection and identification of highly conserved amino-acid loci
2. Prediction of Cytotoxic T Lymphocytes epitopes (CTL)
3. Prediction of Helper T Lymphocytes epitopes (HTL)
4. Prediction of B cell linear epitopes
5. Comparative analysis and selection of the best-fitted epitopes

Lactococcus petauri strains from Rainbow Trout

In an early approach, we retrieved the core genome data from 6 complete genome sequences of *L. petauri* strains isolated from diseased fish and employed the first steps of a reverse vaccinology analysis for the prediction of potential vaccine candidates. More sequences of *L. garvieae* and *L. petauri* isolated from lactococcosis cases from Greece, Italy, Spain, Turkey will be included in the study.

1. Five *L. petauri* strains isolated and analyzed at VRI in Greece (LG-SAV-20, LG1, LG3, LG5, LG6)
2. One *L. petauri* strain isolated at Faculty of Marine Sciences in Turkey (D375?)

we performed a detailed analysis of the genomic dataset of *L. garvieae* and *L. petauri* to shortlist 7 proteins as possible vaccine antigen candidates using properties such as mucus binding ability, surface-exposed nature, fibrogen-binding

Target Proteins:

1. PsaA, metal ABC transporter substrate-binding protein Cytoplasmic Membrane
2. Muc, mucus binding protein (MucBP domain)
3. Fibronectin/Fibrinogen-binding protein
4. LPXTG cell wall anchor domain-containing protein I (107 AA)
5. LPXTG cell wall anchor domain-containing protein II (225 AA)
6. LPXTG cell wall anchor domain-containing protein III (924 AA)
7. LPXTG cell wall anchor domain-containing protein IV (988 AA)

D:\LACTOCOCCUS\proteins from given strains\PsaA.fasta - Notepad++

File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?

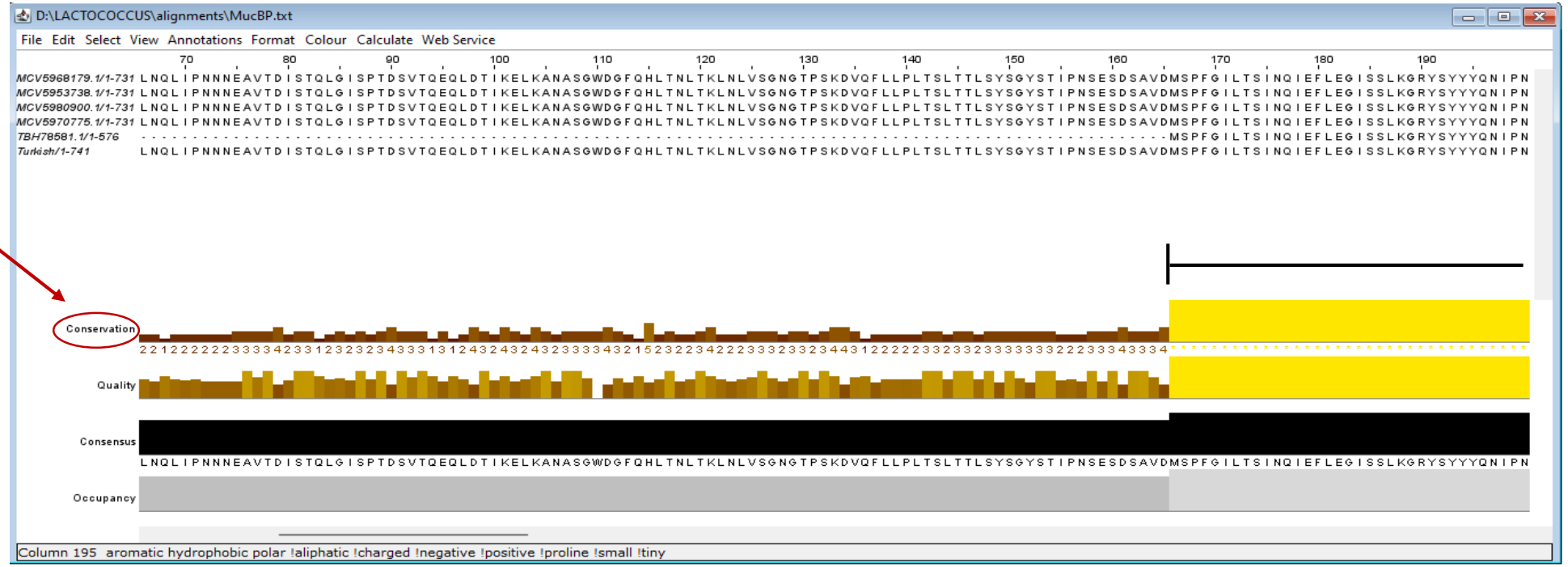
PsaA.fasta | LPXTG 924.fasta | LPXTG 976.fasta | JAOYNZ01.1.fsa_aa | new 1 | Fbp.fasta

```
>MCV5968045.1 metal ABC transporter substrate-binding protein [Lactococcus petauri]
MNRKQLSLITILASLLLFLTACSPKKEESKASDTLKVVTTYSIIADIAENIGKDHVDVYSMVPRGTDPHQ
YDPKPNDTQAVEKADLVFYNGLNLETGKGWFDKLIKNSRKEDSTFMVSQGVTPIHLSEKGKESEEDPHAW
LNIQNGIIYAQNIEKELSKKDPQNKEDYQKNLKVYTDKLQQLDTEAKAKIATIPQEDRILVTSEGAFKYF
SKQYGLTAEYIWEINTDNQGTPAQLNRINTIVKDKNVKALFVETSVSPKTMESVSRQTGVKIYSKIFTDS
LADEGQKGDTYYDMLQWNIEHITDGLSGK

>MCV5952953.1 metal ABC transporter substrate-binding protein [Lactococcus petauri]
MNRKQLSLITILASLLLFLTACSPKKEESKASDTLKVVTTYSIIADIAENIGKDHVDVYSMVPRGTDPHQ
YDPKPNDTQAVEKADLVFYNGLNLETGKGWFDKLIKNSRKEDSTFMVSQGVTPIHLSEKGKESEEDPHAW
LNIQNGIIYAQNIEKELSKKDPQNKEDYQKNLKVYTDKLQQLDTEAKAKIATIPQEDRILVTSEGAFKYF
SKQYGLTAEYIWEINTDNQGTPAQLNRINTIVKDKNVKALFVETSVSPKTMESVSRQTGVKIYSKIFTDS
LADEGQKGDTYYDMLQWNIEHITDGLSGK

>MCV5980937.1 metal ABC transporter substrate-binding protein [Lactococcus petauri]
MNRKQLSLITILASLLLFLTACSPKKEESKASDTLKVVTTYSIIADIAENIGKDHVDVYSMVPRGTDPHQ
YDPKPNDTQAVEKADLVFYNGLNLETGKGWFDKLIKNSRKEDSTFMVSQGVTPIHLSEKGKESEEDPHAW
LNIQNGIIYAQNIEKELSKKDPQNKEDYQKNLKVYTDKLQQLDTEAKAKIATIPQEDRILVTSEGAFKYF
SKQYGLTAEYIWEINTDNQGTPAQLNRINTIVKDKNVKALFVETSVSPKTMESVSRQTGVKIYSKIFTDS
LADEGQKGDTYYDMLQWNIEHITDGLSGK
```

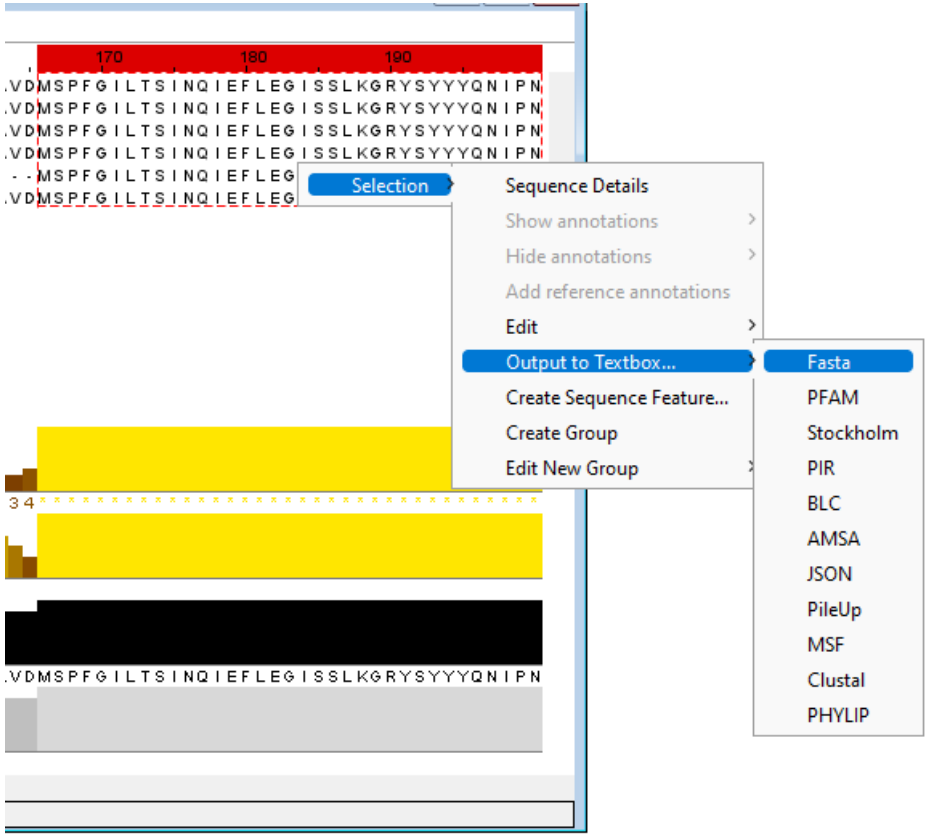
Jalview

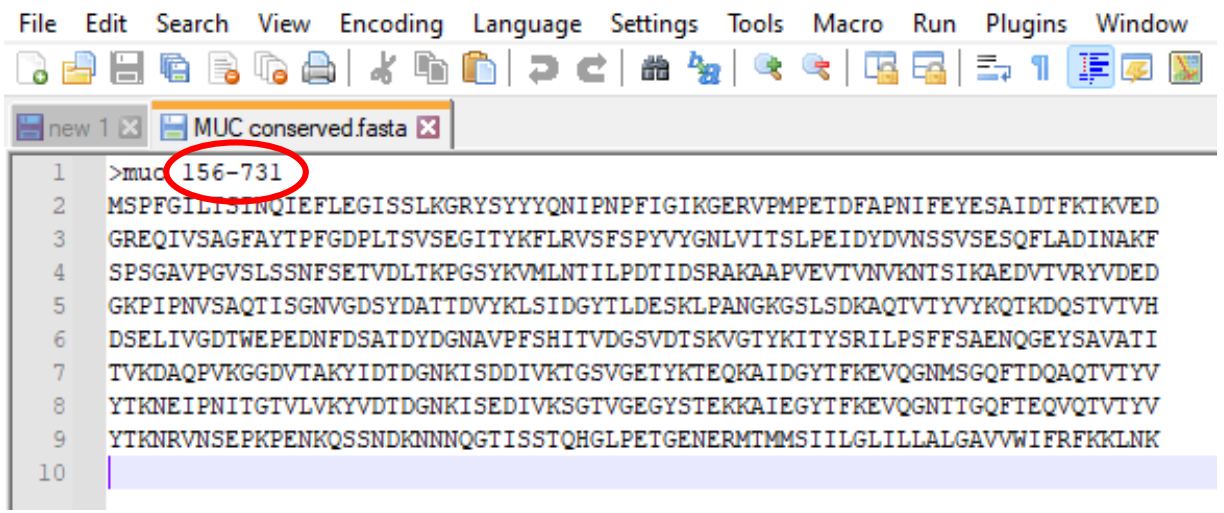
A workbench for multiple sequence alignment and analysis

Multiple allignment of proteins



Identification of conservative regions

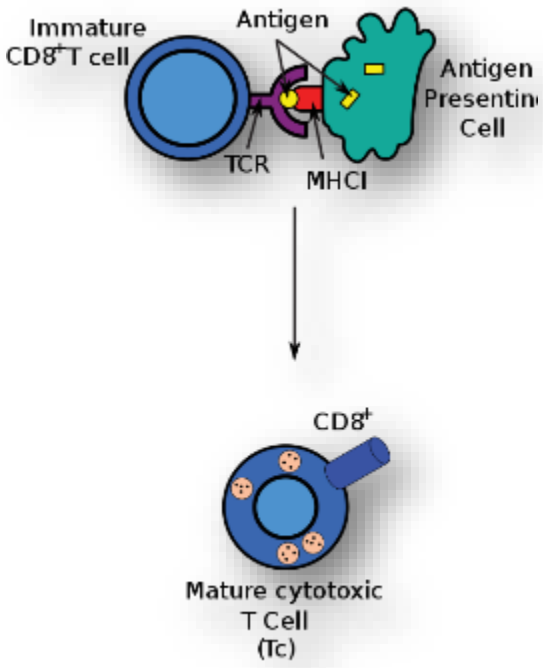




T-cell epitope prediction

The objective of T-cell epitope prediction is to identify short peptide sequences within an antigen that can act as a stimulant of CD4+ or CD8+ T-cells. More specifically:

1. Selection of cytotoxic T lymphocytes (CTL) epitopes. It predicts the MHC-class I binding peptide sequences
2. Selection of helper T cells (HTL) epitopes



1. Selection of cytotoxic T lymphocytes (CTL) epitopes, there is no selection of fish MHC I alleles to perform the analysis and **this is a major limitation**

But there is a solution :
NetMHCpan 4.1 & IPD-MHC Database, is a tool for predicting the binding capacity of protein regions that we submit. What interests us about this particular tool and differentiates it from the rest, is that there is an option to submit our own MHC I sequence and have the analysis done based on it.

PRIMA
PARTNERSHIP FOR RESEARCH AND INNOVATION IN THE MEDITERRANEAN AREA

IPD-MHC

1.

Salmonids

Welcome to the IPD-MHC FISH database, a specialist database for the Major Histo
database includes at present two salmonid species: Atlantic salmon (*Salmo salar*)
Atlantic salmon and rainbow trout possess a single classical MHC class I locus, UBA, a
encoded by closely linked DAA (IIA) and DAB (IIB) loci. Included in the database are
binding domain-coding sequences from published studies and from the EMBL and NCBI

For those readers who would like to know more about MHC sequences in salmonids, ple

Alleles
List all Salmonids alleles >

Species
Species belonging to the Salmonids group >

Align
Multi locus inter- and intra- species alignment >

2. ## *Salmonids* species

Below all organisms belonging to the *Salmonids* group are listed. The official designations are assigned by the Comparative MHC Nomenclature Committee as established by the International Society for Animal Genetics (ISAG) who are affiliated to the International Union of Immunological Societies (IUIS) - Veterinary Immunology Committee (VIC). Organism four-letters codes are written in parentheses.

If you are working on MHC nomenclature and your species is not represented, please contact the Comparitive MHC Nomenclature Committee.

Species	Common Name	Class I	Class II
Oncorhynchus mykiss (Onmy)	Rainbow Trout	*UBA*	*DAA, DAB*
Salmo salar (Sasa)	Atlantic Salmon	*UBA*	*DAA, DAB*

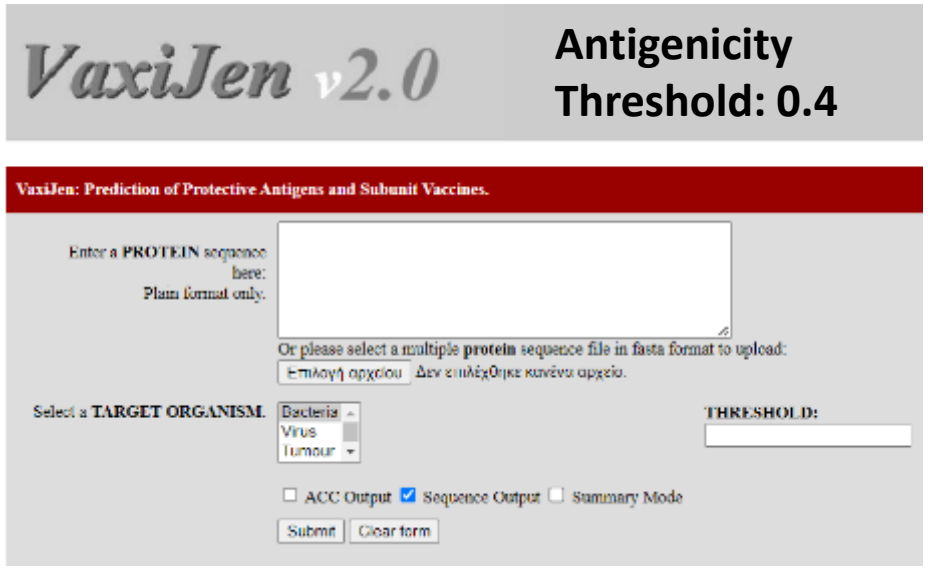
3.

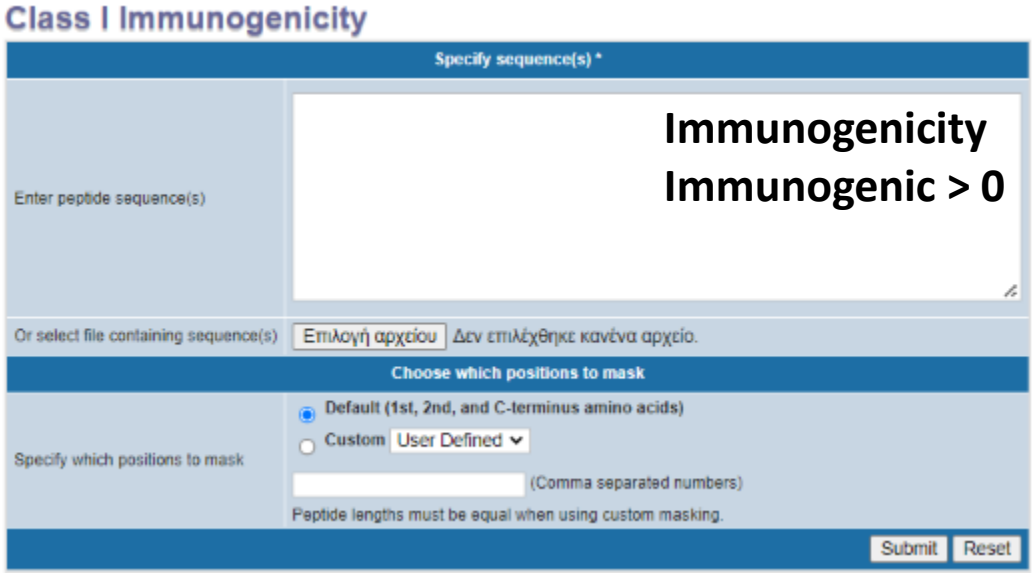
Submit | Clear | Download

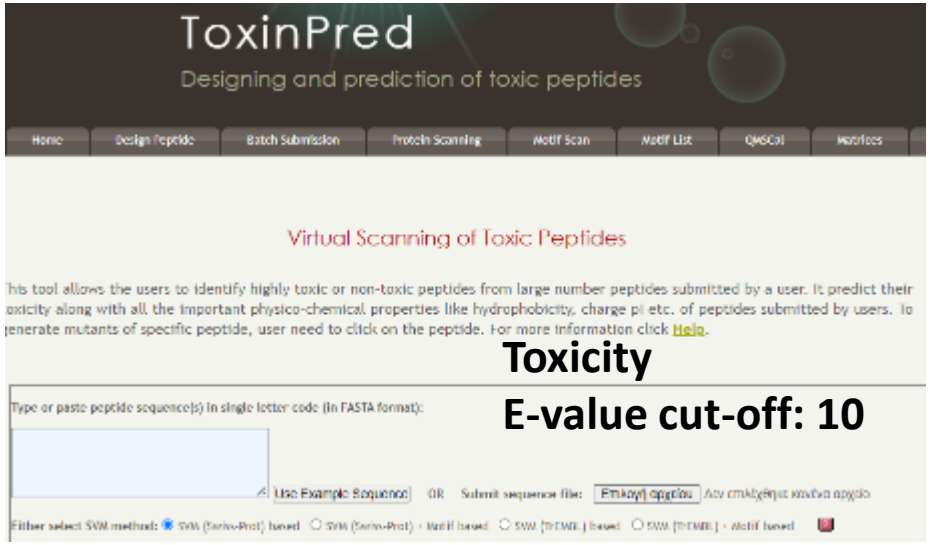
Result list
Nucleotide CDS FASTA
Genomic FASTA
Protein FASTA

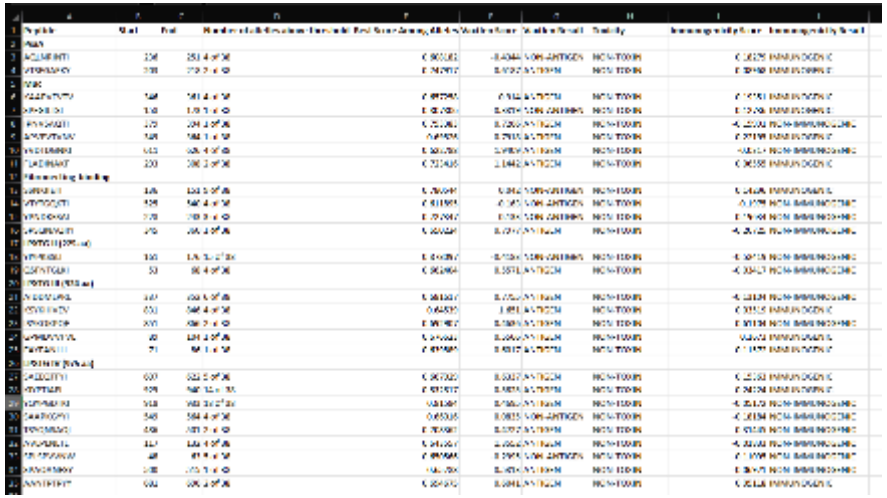
Accession		locus	species
FISH08144	*01:01:01	UBA	Onmy
FISH08145	*01:01:02	UBA	Onmy
FISH08146	Onmy-UBA*01:01:03	UBA	Onmy
FISH08147	Onmy-UBA*01:02	UBA	Onmy
FISH08148	Onmy-UBA*01:03	UBA	Onmy
FISH08149	Onmy-UBA*01:04	UBA	Onmy
FISH08150	Onmy-UBA*02:01	UBA	Onmy

Evaluation of Antigenic, Allergenic, Immunogenicity, and Toxicity of CTL epitopes





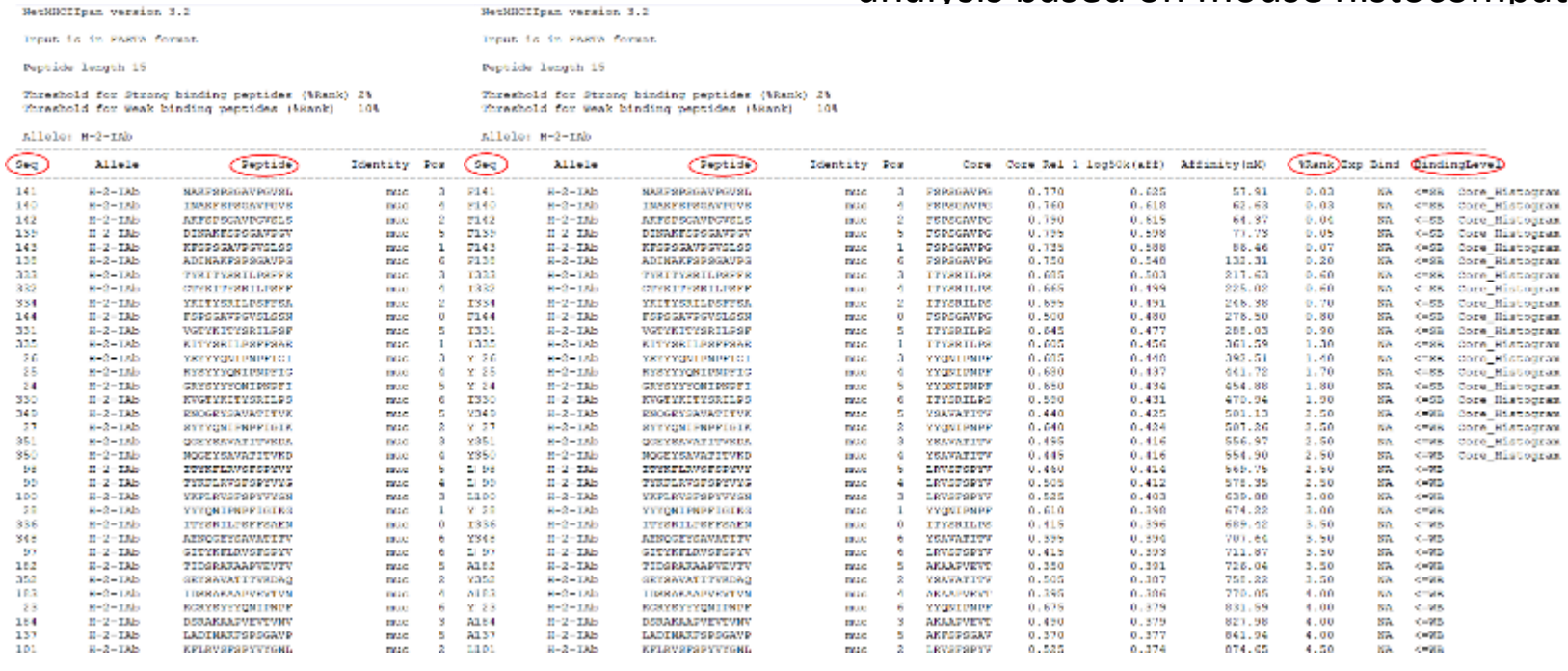




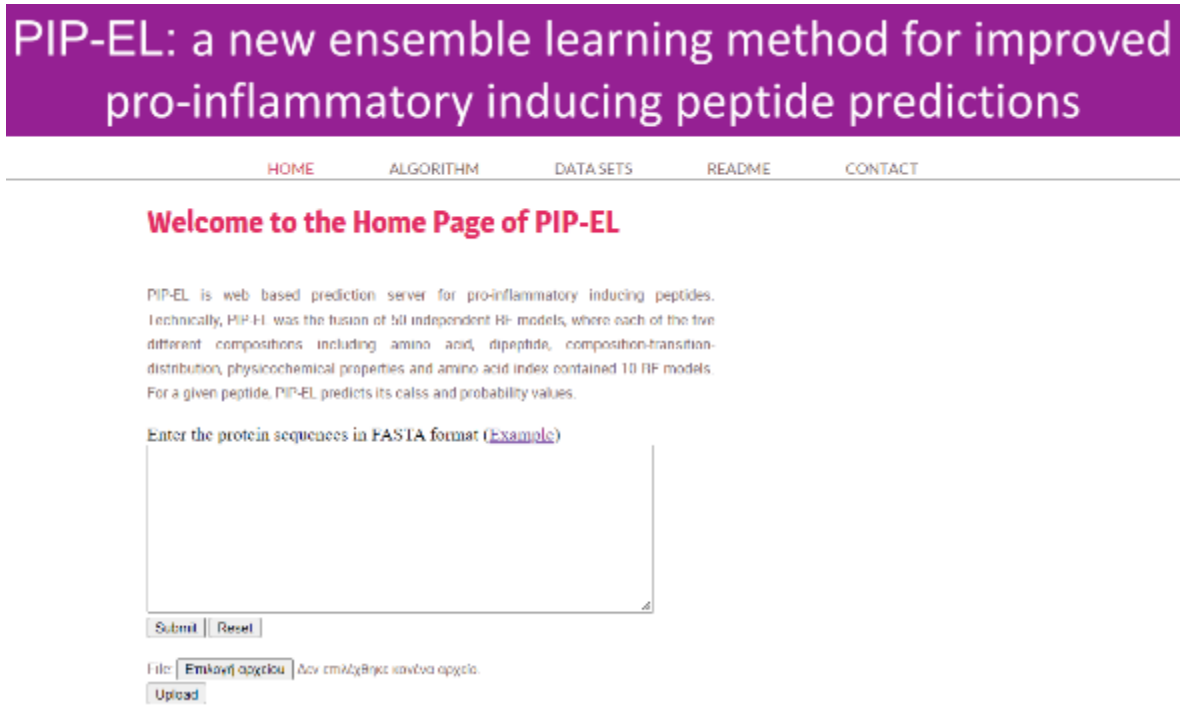
T-cell epitope prediction

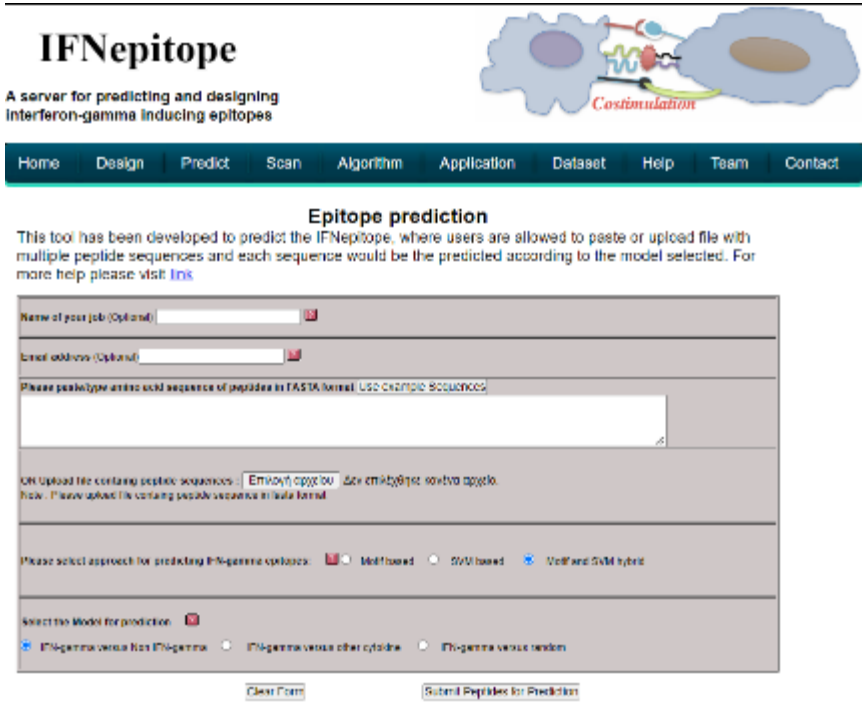
2. Selection of helper T cells (HTL) epitopes. As in the case of CTL epitopes, antigen presentation is required, but they take part in the MHC Class II process.

The problem: There is no tool that can do anything like what we did with MHC I. The only choices are human, bovine, or mouse available MHC II. Therefore, we chose to perform the analysis based on mouse histocompatibility molecules.



Additional properties of HTL epitopes (interferon-γ inducing ability, prediction of toxicity, prediction of pro-inflammatory cytokine production)

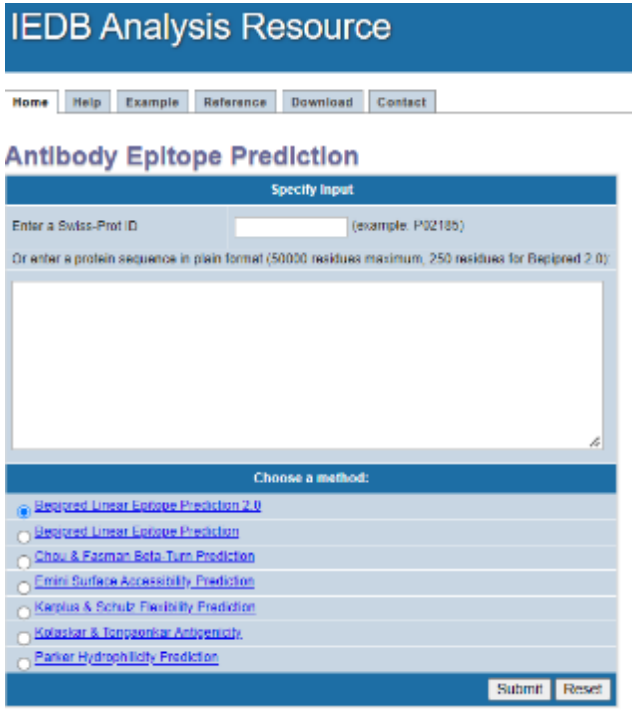




B-Cell epitope prediction

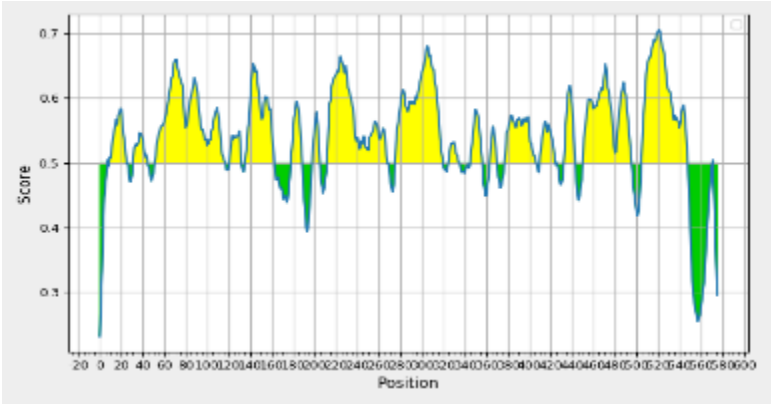
B-cell epitopes are two types, linear (continuous) and conformational (discontinuous).

In the case of vaccine design, only the linear epitopes are considered



Predicted peptides:

No.	Start	End	Peptide	Length
1	8	8	T	1
2	10	25	INQIEFLEGISSLKGR	16
3	32	45	NIPNPFIGIKGERV	14
4	53	117	APNIFEYESAIDTFKTKVEDGREQIVSAGFAYTPFGDPLTSVSEGITYKFLRVSFSPYVYGNLVI	65
5	122	132	EIDYDVNSSVS	11
6	137	163	LADINAKFSPSGAVPGVSLSSNFSETV	27
7	179	189	LPDTIDSRAKA	11
8	199	205	KNTSIKA	7
9	213	268	VDEDGKPIPNVSAQTISGNVGDSYDATTDVYKLSIDGYTLDESKLPANGKGSLSDK	56
10	277	320	KQTKDQSTVTVHDSELIVGDTWEPEDNFDSATDYDGNAVPFSHI	44
11	326	336	VDTSKVGTYKI	11
12	347	357	SAENQGEYSAV	11
13	365	370	AQPVKG	6
14	378	407	IDTDGNKISDDIVKTGSVGETYKTEQKAID	30
15	411	425	FKEVQGNMSGQFTDQ	15
16	427	427	Q	1
17	434	443	TKNEIPNITG	10
18	451	496	DTDGNKISEDIVKSGTVGEGYSTEKKAIEGYTFKEVQGNTTGQFTE	46
19	506	548	TKNRVNSEPKPENKQSSNDKNNNQGTISSTQHGLPETGENERM	43
20	572	572	K	1



Initial results and further developments

1. Analysis completed for MHC-I alleles of rainbow trout
2. CTL epitopes were examined for their antigenicity, immunogenicity, allergenicity and toxicity.
3. For HTL epitopes there is no available reference MHC II alleles for rainbow trout.
4. Analysis completed for B-cell epitopes

Further steps are needed:

Analyses of the physicochemical properties of the peptide, 3D modelling, molecular docking, immuno-response simulations, etc.

More *L. petauri* strains and *L. garvieae* isolates are needed…

The workflow ends in the construction of a multi-epitope vaccine

The vaccine construct will be designed with the inclusion of high scored T-cell and B-cell epitopes predicted from various epitope prediction tools and exhibiting high immunogenicity, non-toxicity, non-allergenic and strong binding affinity to a maximum number of HLA alleles.

Reverse Vaccinology for lactococcosis (Bioinformatics workflow)

In conclusion:

Despite significant advancements in vaccinology, computational proteomics, machine learning, and reverse vaccinology, finding vaccine candidates, producing them in the laboratory, and confirming their efficacy in animal models remains a complicated undertaking.

.

Objectives

- Immunogenic protein identification
- recombinant-protein production
- experimental trial for vaccine efficacy