

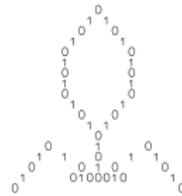
R-emo

Estudio de la relación de divergencia en el uso de codones sinónimos entre virus-huésped y presencia de microRNA

Riberi, Franco Gaspar



UNIVERSIDAD NACIONAL DE RÍO CUARTO
FAC. DE CS. EXACTAS, FCO-QCAS Y NATURALES
DEPARTAMENTO DE COMPUTACIÓN



FUDEPAN
FUNDACIÓN PARA EL DESARROLLO DE LA PROGRAMACIÓN
EN ÁCIDOS NUCLEICOS

8 de Noviembre de 2013

Temario:

1 Introducción

- FuDePAN
- Motivación

2 Marco Teórico

- Conceptos de Biología

3 Descripción del Problema

4 R-emo

- Metodología de Desarrollo
- Elicitación y Análisis
- Diseño
- Implementación

5 Conclusión

6 Bibliografía

FuDePAN

La **Fundación para el Desarrollo de la Programación en Ácidos Nucleicos** es una ONG (organización no gubernamental) concebida para fomentar el desarrollo de técnicas y tecnologías para la Programación en Ácidos Nucleicos, al servicio de la salud humana.

- Fundada en 2006.
- Bioinformática ingenieril.
- I+D.
- Colaboran
 - Estudiantes: Voluntarios y tesistas
 - Profesionales: Ámbito industrial y académico.
- Todo lo desarrollado por la fundación se encuentra bajo GPLv3.

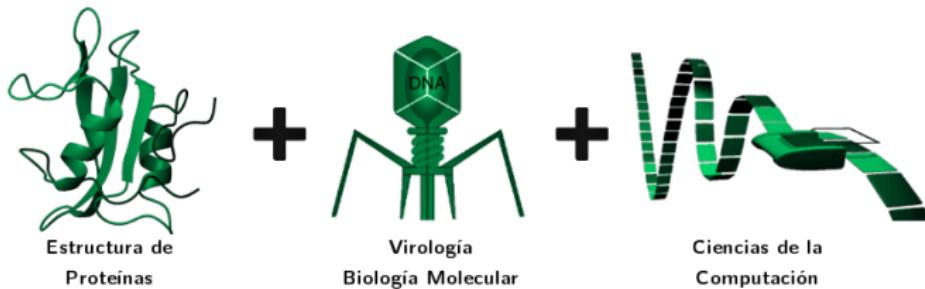
FuDePAN

La **Fundación para el Desarrollo de la Programación en Ácidos Nucleicos** es una ONG (organización no gubernamental) concebida para fomentar el desarrollo de técnicas y tecnologías para la Programación en Ácidos Nucleicos, al servicio de la salud humana.

- Fundada en 2006.
- Bioinformática ingenieril.
- I+D.
- Colaboran
 - Estudiantes: Voluntarios y tesistas
 - Profesionales: Ámbito industrial y académico.
- Todo lo desarrollado por la fundación se encuentra bajo GPLv3.

FuDePAN (cont.)

Pilares

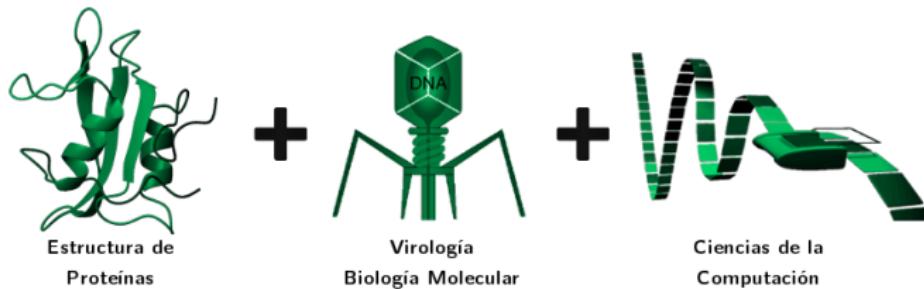


Naturaleza de los problemas

- Corren mucho tiempo
- Alta performance
- Resultados impactan en nuevos problemas y personas.

FuDePAN (cont.)

Pilares



Naturaleza de los problemas

- Corren mucho tiempo
- Alta performance
- Resultados impactan en nuevos problemas y personas.

Motivación

- **Dado a que:** FuDePAN tenía un trabajo disponible.
- **Surgió la necesidad de:** Investigar y desarrollar un software.
- **Objetivo Principal:** contrastar una teoría (una hipótesis).

Dr. Daniel Rabinovich



Perspectiva Personal

- Tarea desafiante.
- Integrarse en un grupo de desarrollo formado.
- Adaptarse a nuevos entornos.
- Contribuir a la salud.
- Retribuir con trabajo.

Motivación

- **Dado a que:** FuDePAN tenía un trabajo disponible.
- **Surgió la necesidad de:** Investigar y desarrollar un software.
- **Objetivo Principal:** contrastar una teoría (una hipótesis).

Dr. Daniel Rabinovich

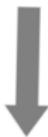


Perspectiva Personal

- Tarea desafiante.
- Integrarse en un grupo de desarrollo formado.
- Adaptarse a nuevos entornos.
- Contribuir a la salud.
- Retribuir con trabajo.

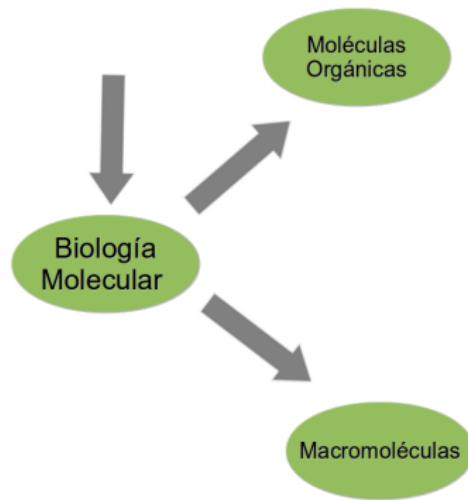
Conceptos de Biología

Conceptos de Biología

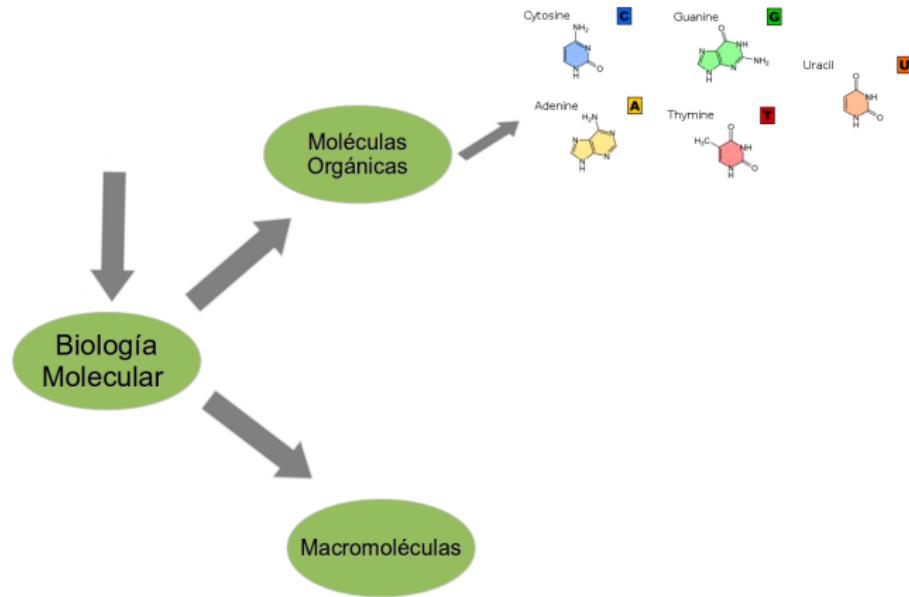


Biología
Molecular

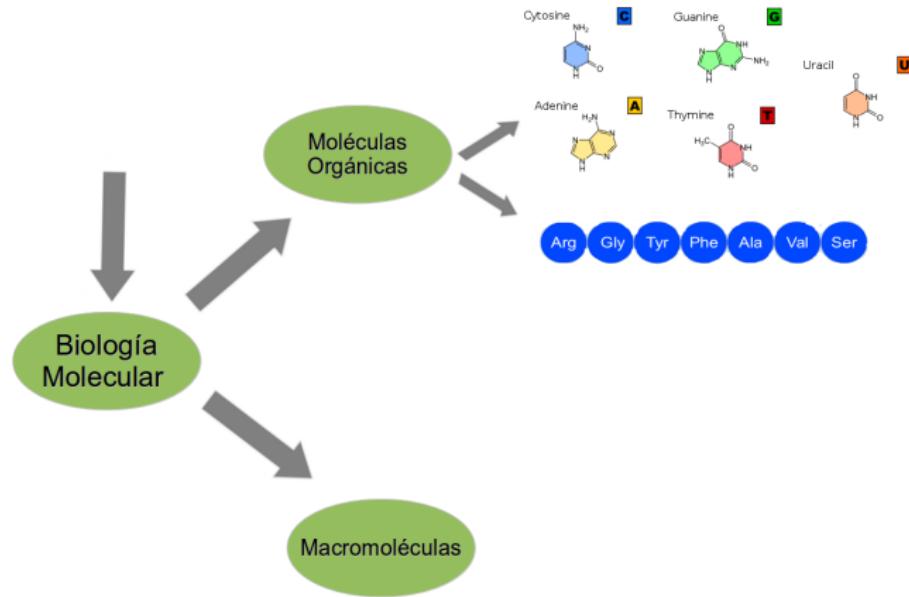
Conceptos de Biología



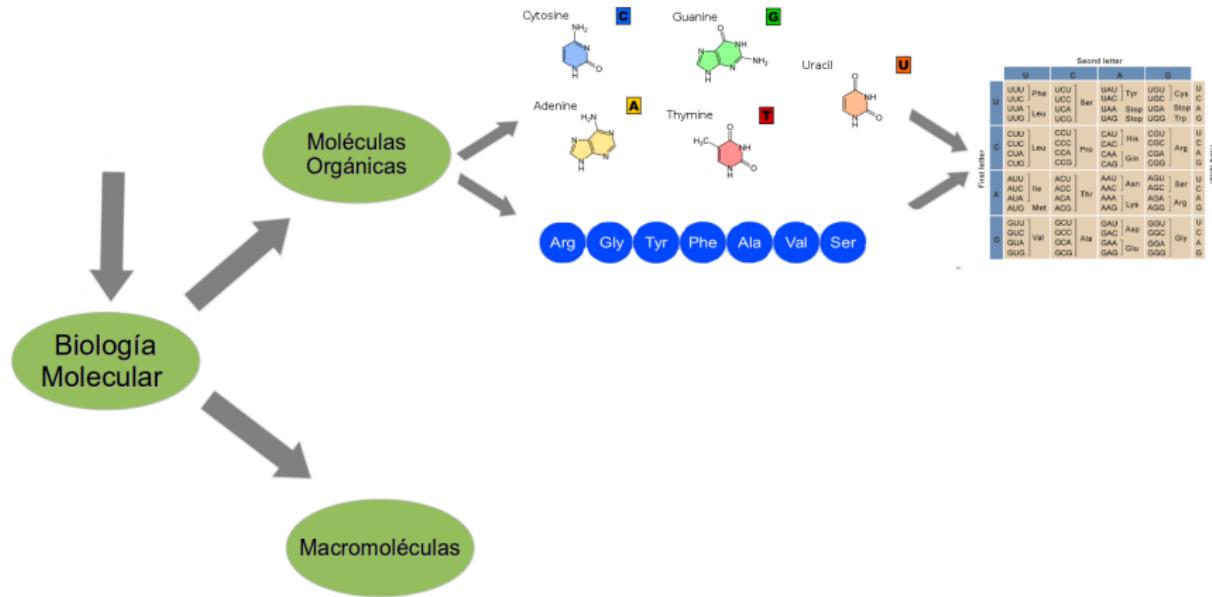
Conceptos de Biología



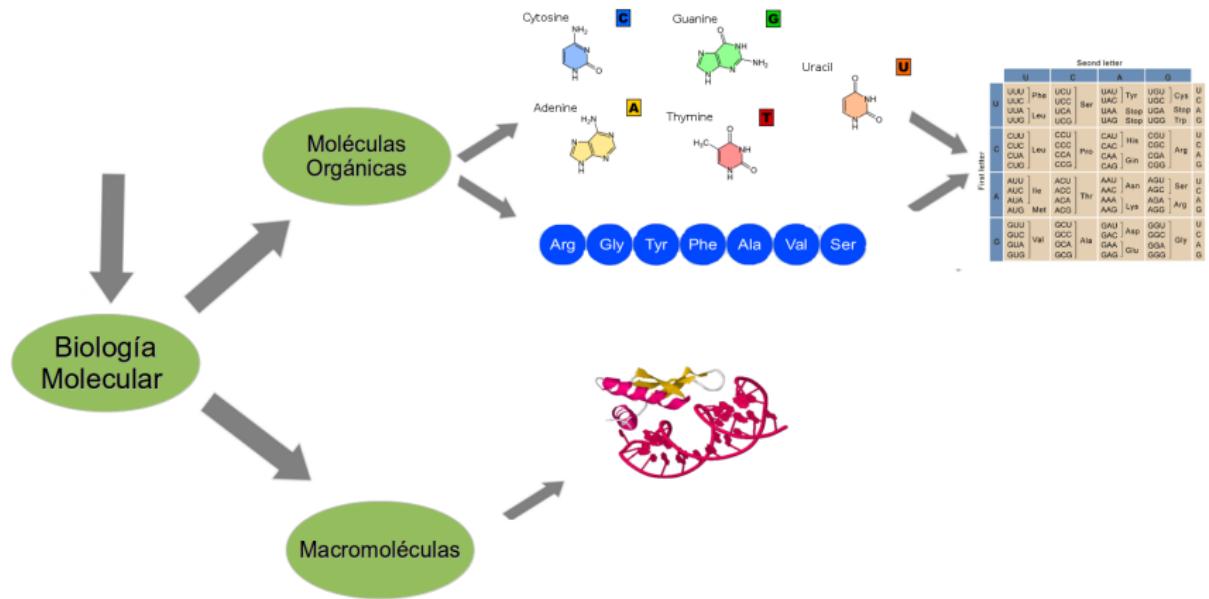
Conceptos de Biología



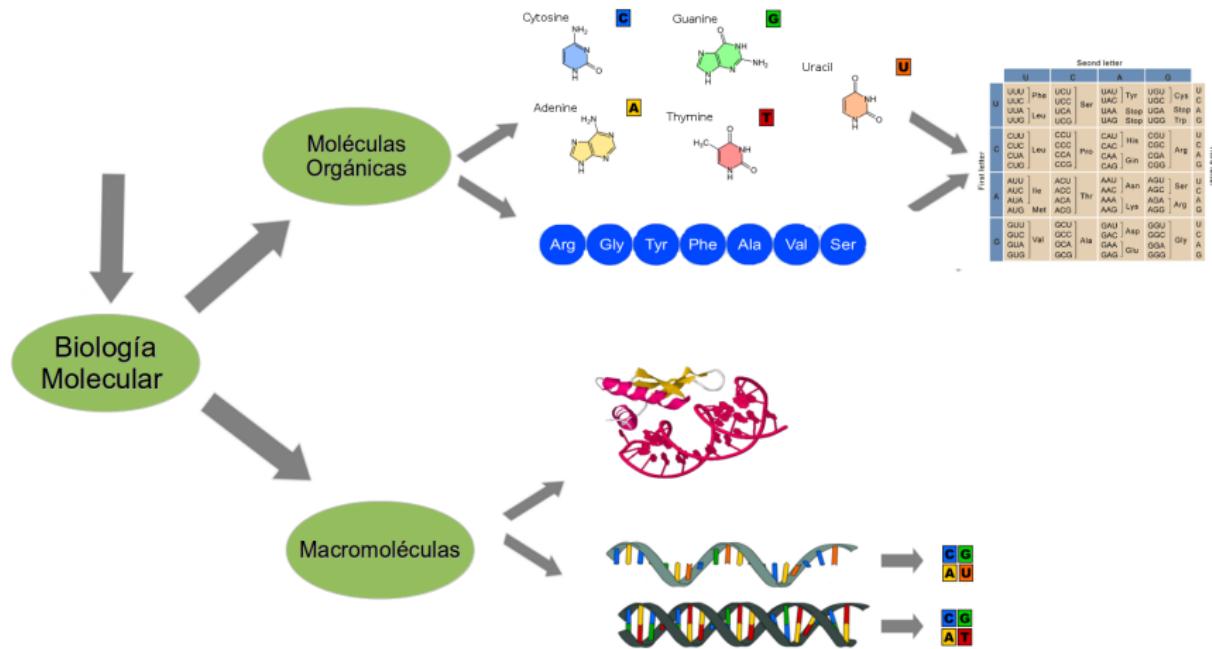
Conceptos de Biología



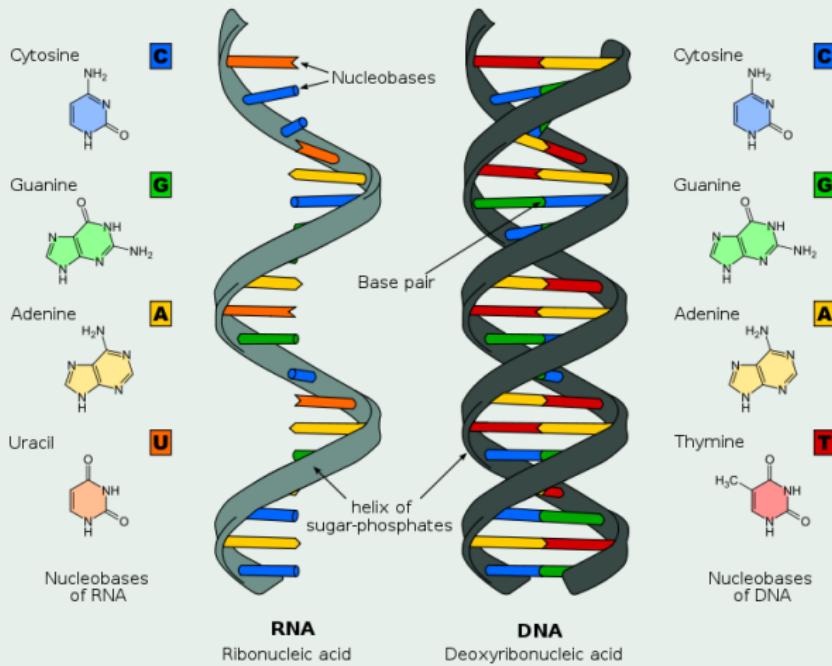
Conceptos de Biología



Conceptos de Biología



Ácidos Nucleicos



Ácidos Nucleicos (cont.)

RNA

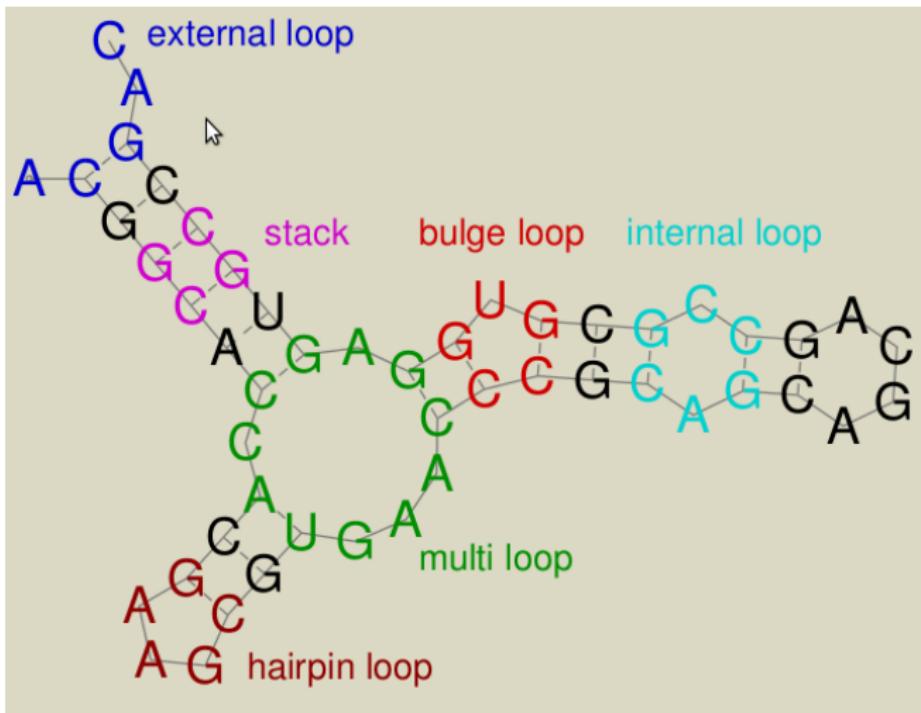
Generalmente, consiste en una hebra de cadena simple, la cual usualmente es trascrita a partir de una porción de ADN y se utiliza posteriormente en la célula para la síntesis de proteínas. Algunos virus poseen ARN como único material genético cuya monohebra puede plegarse, dando lugar a lo que se conoce como estructura secundaria.



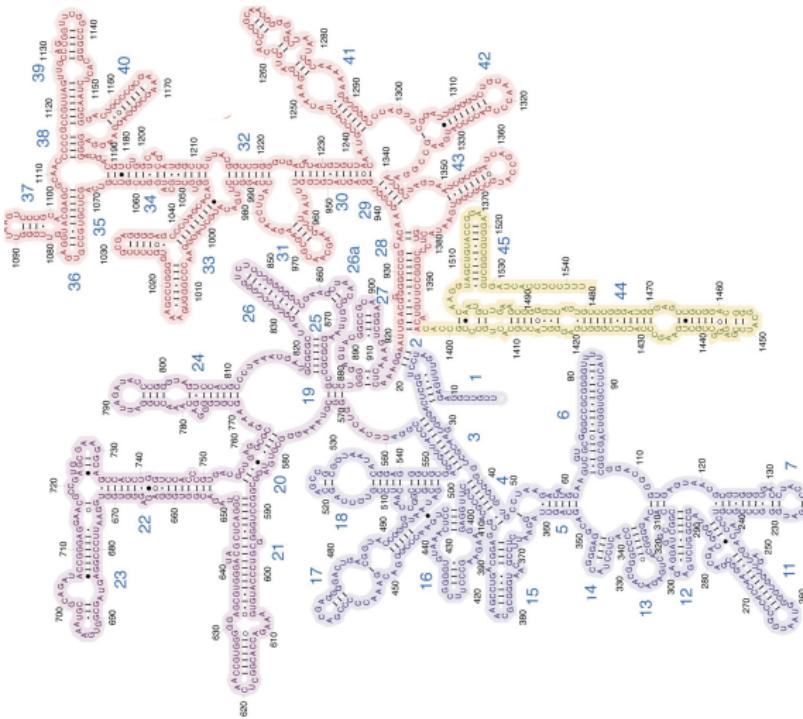
La Estructura Primaria y Secundaria del RNA

La estructura primaria del ARN, es una secuencia de nucleótidos de longitud n , $A = a_1 a_2 a_3 \dots a_n$ con $a_i \in \{A, U, G, C\}$. El plegamiento de una secuencia de ARN entre sus bases complementarias determina lo que se denomina estructura secundaria de ARN.

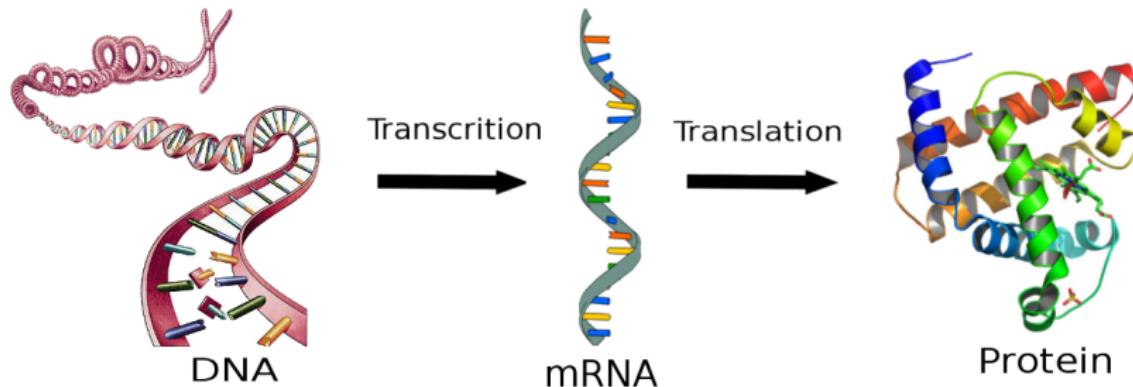
Estructura Secundaria del RNA (motifs)



Estructura Secundaria del RNA (cont.)



Dogma Central de la Biología



Descripción del Problema

Hechos importantes:

- Código genético degenerado.
- Codones Sinónimos.
- Uso de codones universal.
- Divergencia de codones no esclarificada.

CCU
CCC
CCA
CCG } Pro

Hipótesis

“La divergencia en el uso de codones sinónimos entre virus y huésped contribuye a disminuir la interferencia de los *miRNA* en la expresión de los RNA mensajeros de origen viral.”

Descripción del Problema

Hechos importantes:

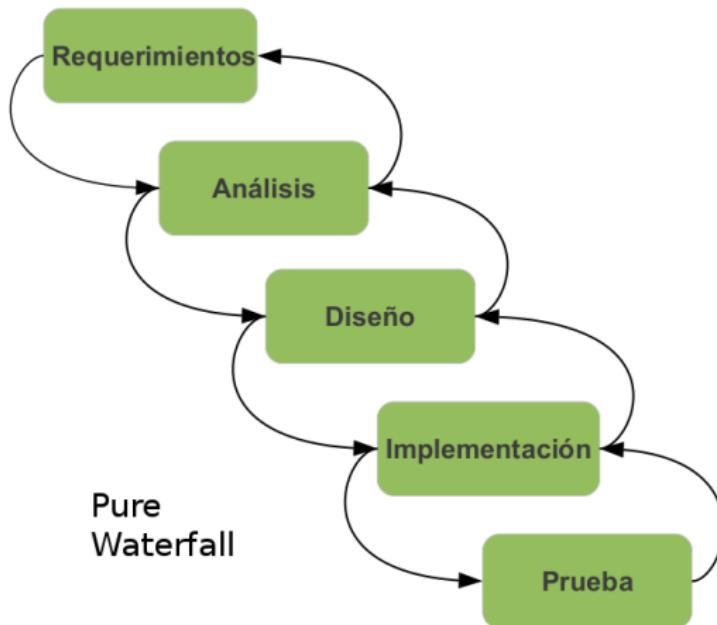
- Código genético degenerado.
- Codones Sinónimos.
- Uso de codones universal.
- Divergencia de codones no esclarificada.

CCU
CCC
CCA
CCG } Pro

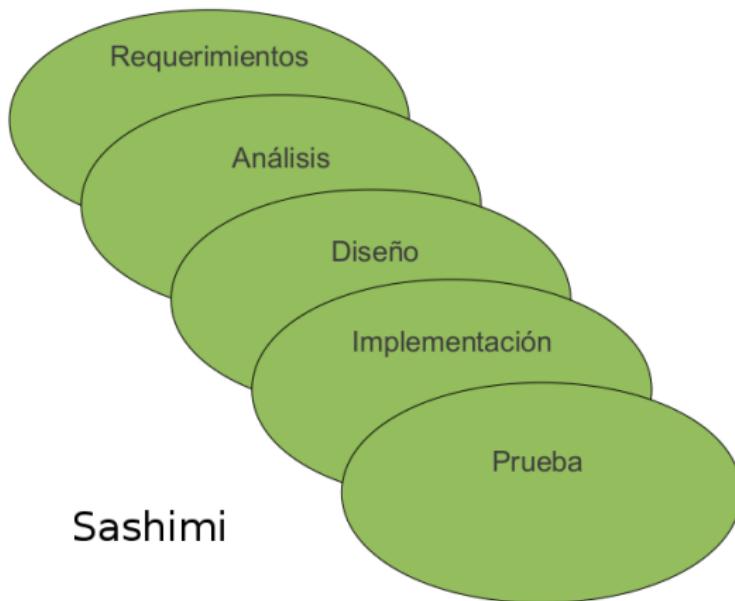
Hipótesis

“La divergencia en el uso de codones sinónimos entre virus y huésped contribuye a disminuir la interferencia de los *mi*RNA en la expresión de los RNA mensajeros de origen viral.”

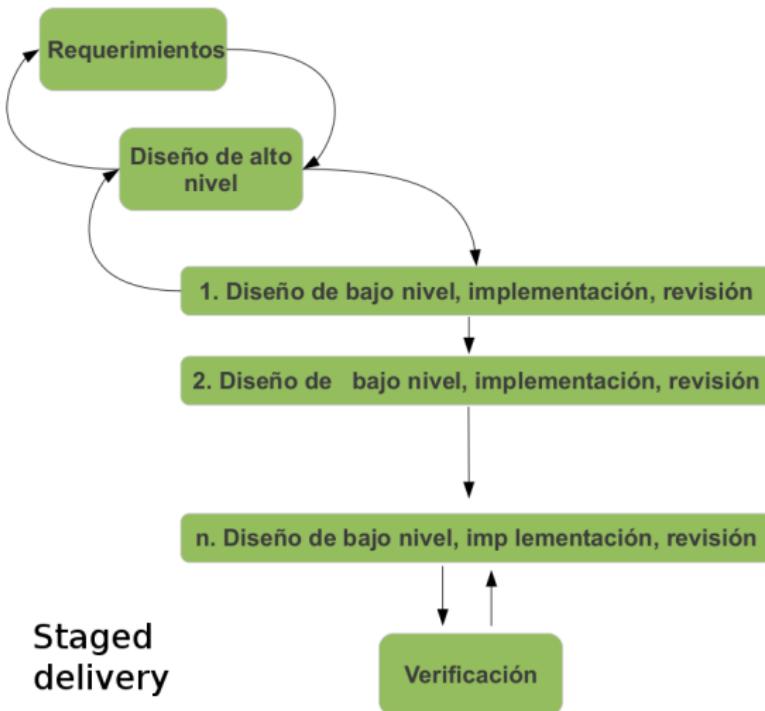
Metodología de Desarrollo



Consideraciones sobre la Metodología



Consideraciones sobre la Metodología (cont.)



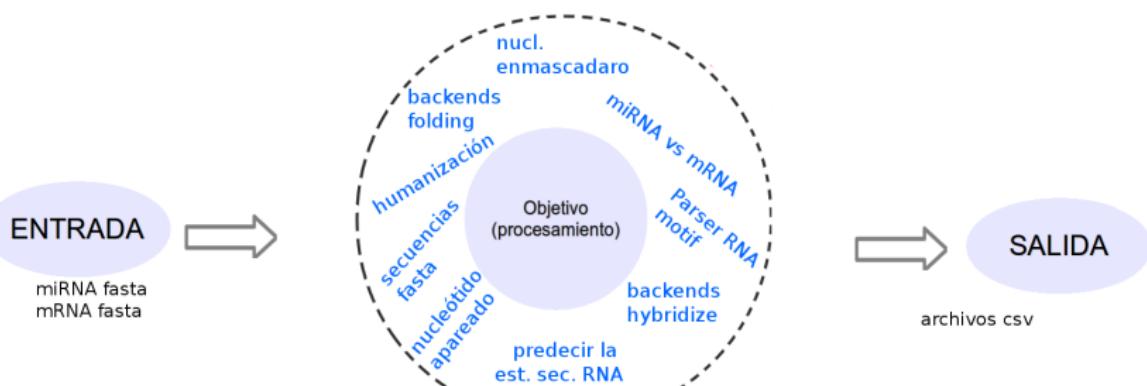
Especificación

>hsa-mir-140-5p MIMAT0000431
AAGCUUGGUAAAUCACCUUAGUAG
>hsa-mir-2682-5p MIMAT0013517
CACGAGCUAGCUCUGACAGUC
>hsa-mir-3761 MIMAT0002172
AUCAUAAGGGAAAACAUAGGUU
>hsa-mir-5197-3p MIMAT0021131
AACAGAGACAGUCAGUCAGAU
>hsa-mir-522-5p MIMAT005451
CUCUAGGAGGAAGCCUUUCUC
>hsa-mir-557 MIMAT0003221
GUUUGCACGGGGGCCGUUCUG
>hsa-mir-548ac-3p MIMAT0021030
AAAGACCGUC4aaac-3p UUUGCA
>hsa-mir-548av-5p MIMAT0022303
AAAGUACGUCGGGCUA

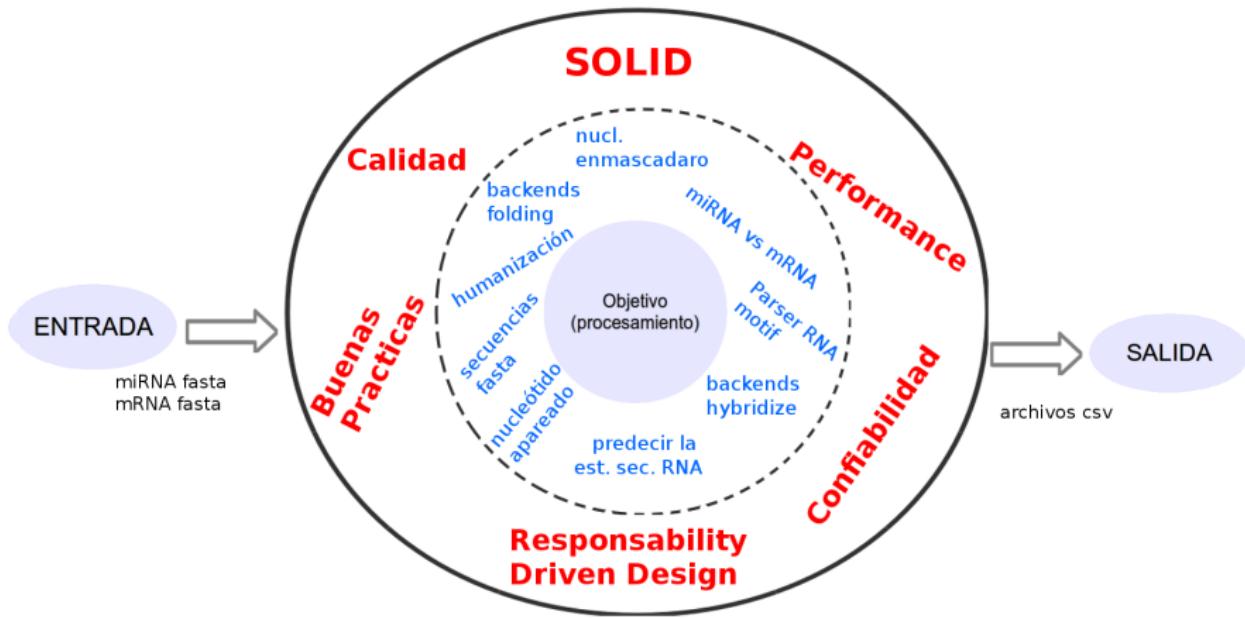
miRNA fasta



Especificación (cont.)



Especificación (cont.)



Creación de librerías

Surgen:

- **fideo** : Folding Interface Dynamic Exchange Operations



- **acuoso** : Abstract Codon Usage Optimization Software for Organisms



- **etilico** : External Tools Invocation Library and COmponent



Principios de Diseño

Responsability-Driven Design



- Que responsabilidades debe cubrir el sistema.
- Cuales serán los objetos responsables de llevarlas a cabo.

SOLID

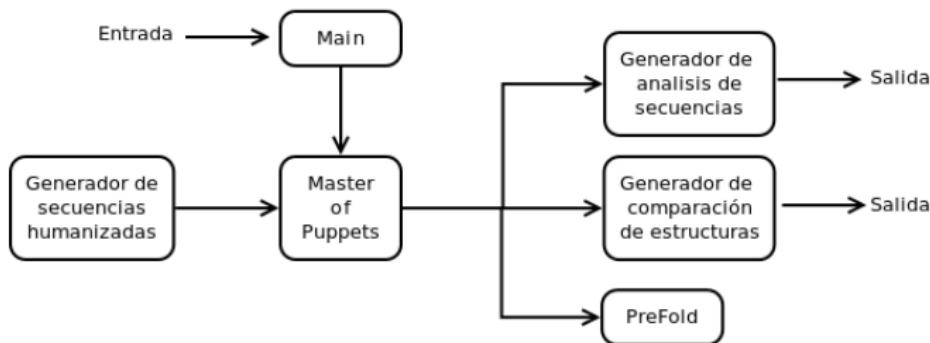
- Single Responsibility Principle (SRP)
- Open-Closed Principle (OCP)
- Liskov Substitution Principle (LSP)
- Interface Segregation Principle (ISP)
- Dependency Inversion Principle (DIP)



SOLID

Software Development is not a Jenga game

Arquitectura



Arquitectura (refinando)

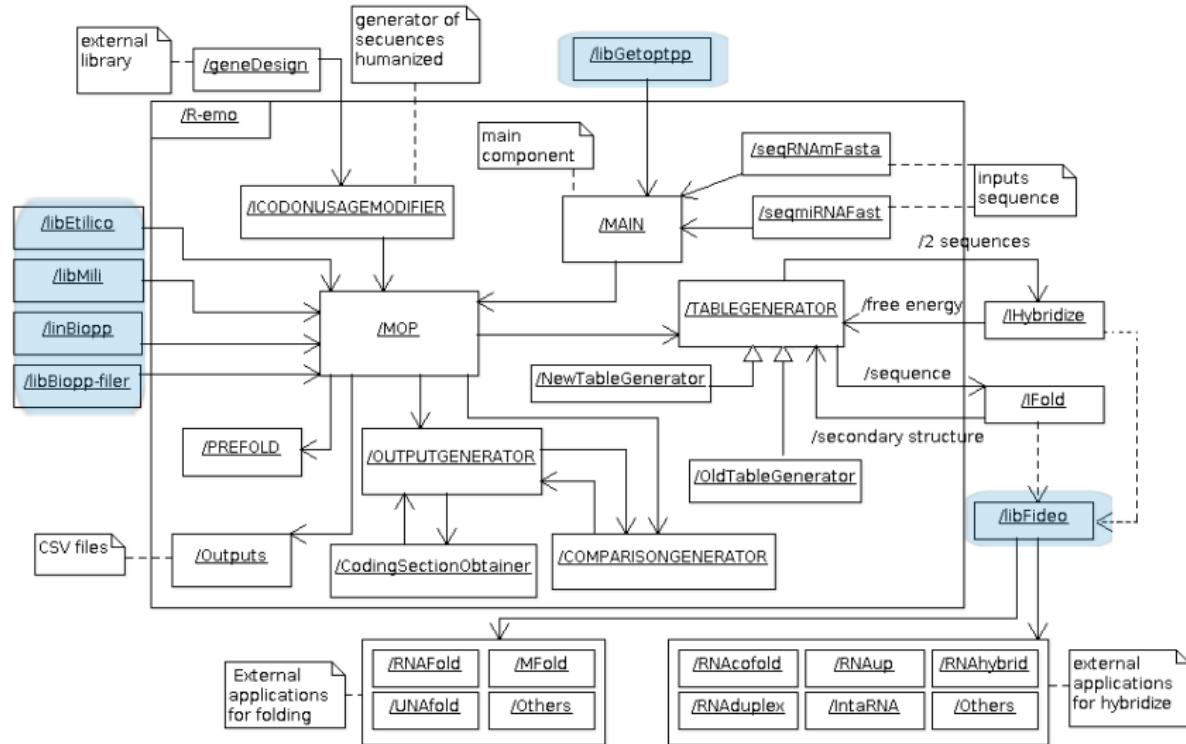


Diagrama de Clases

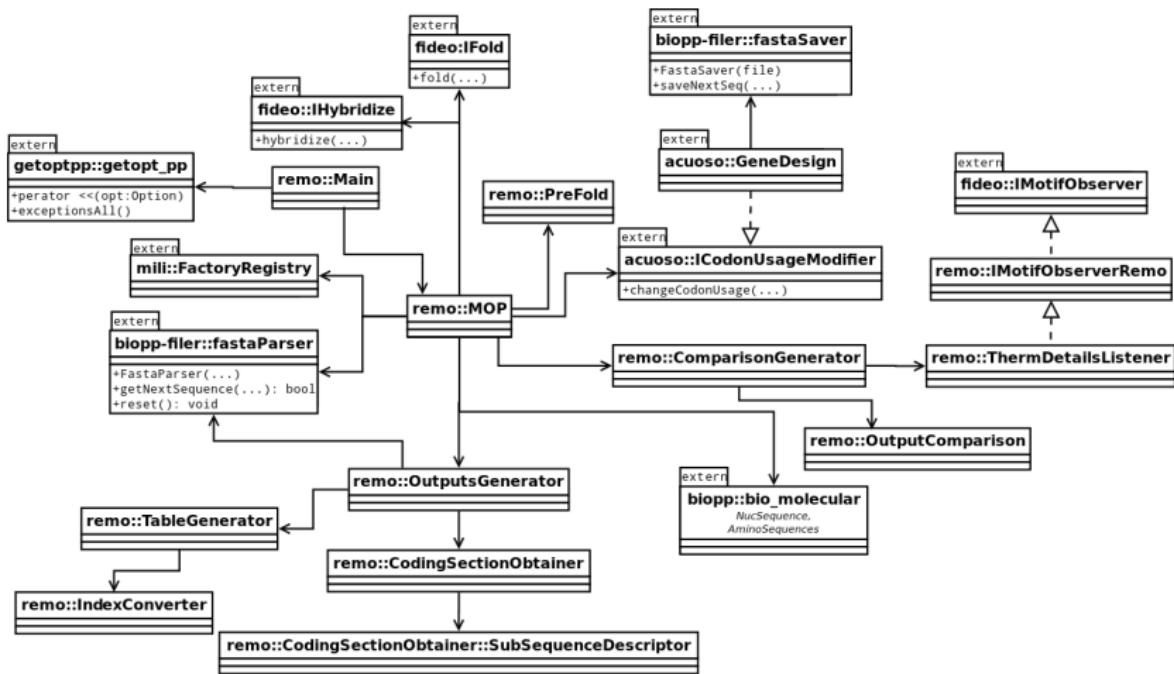


Diagrama de Clases (Cont.)

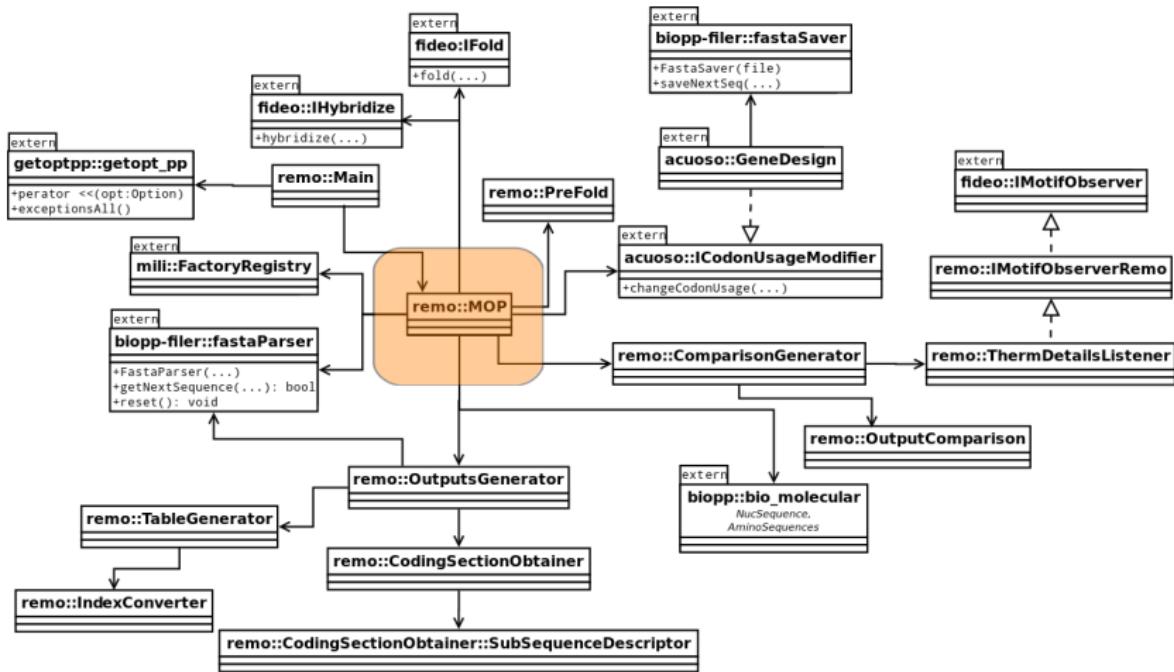


Diagrama de Clases (Cont.)

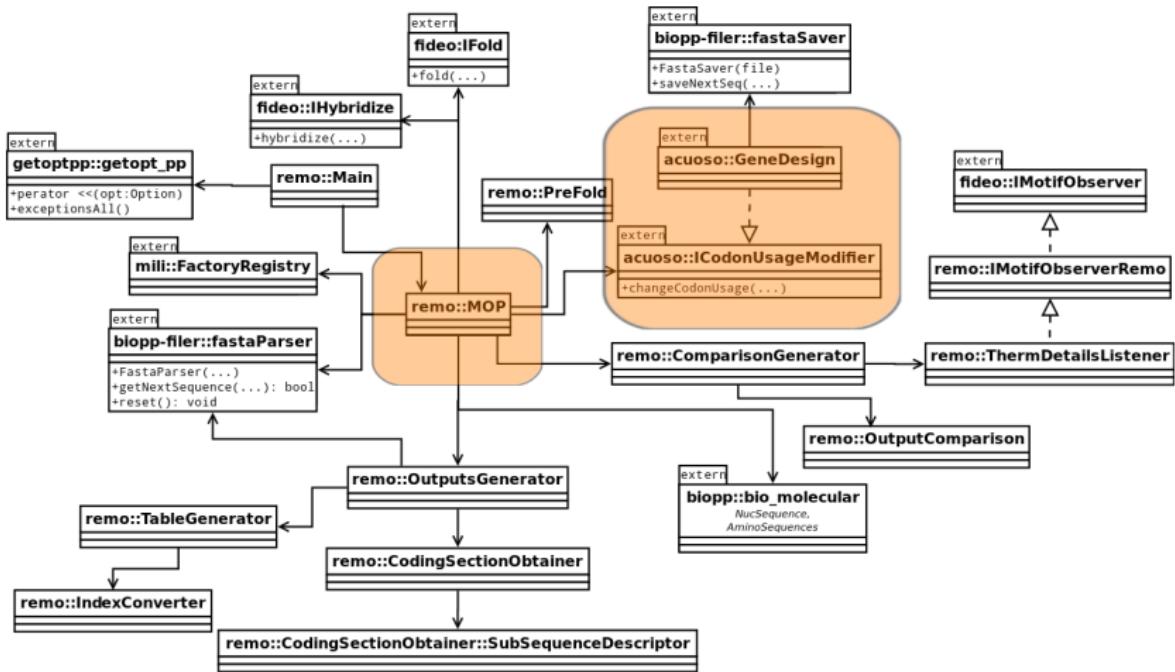
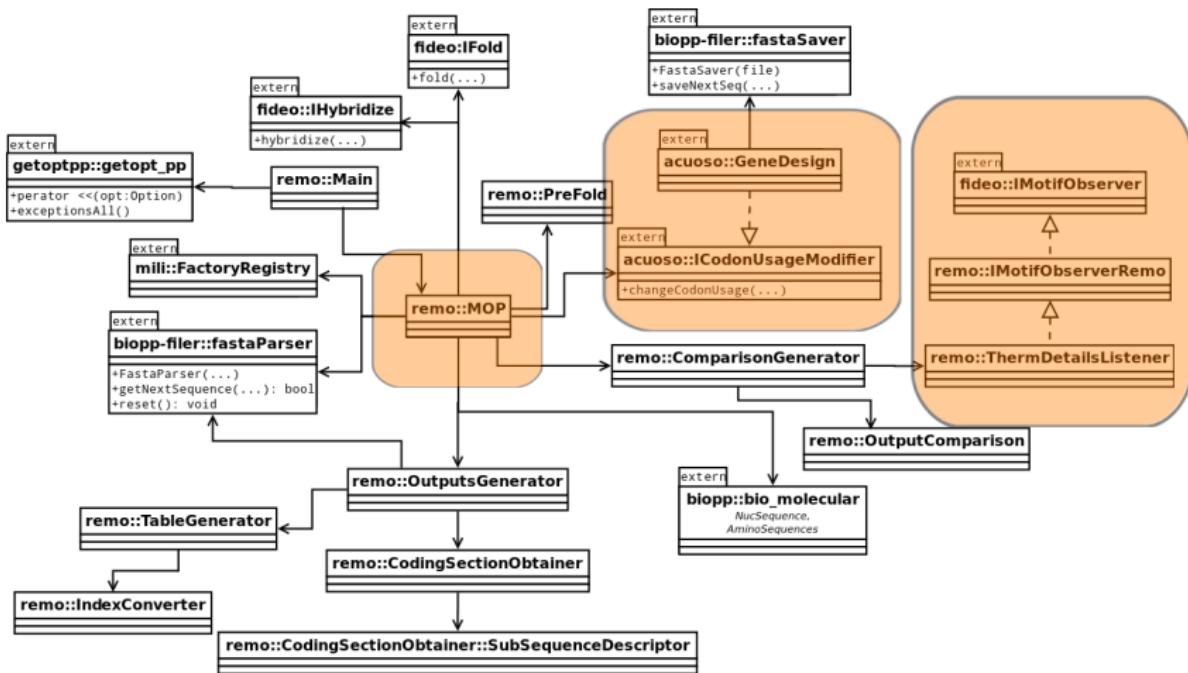
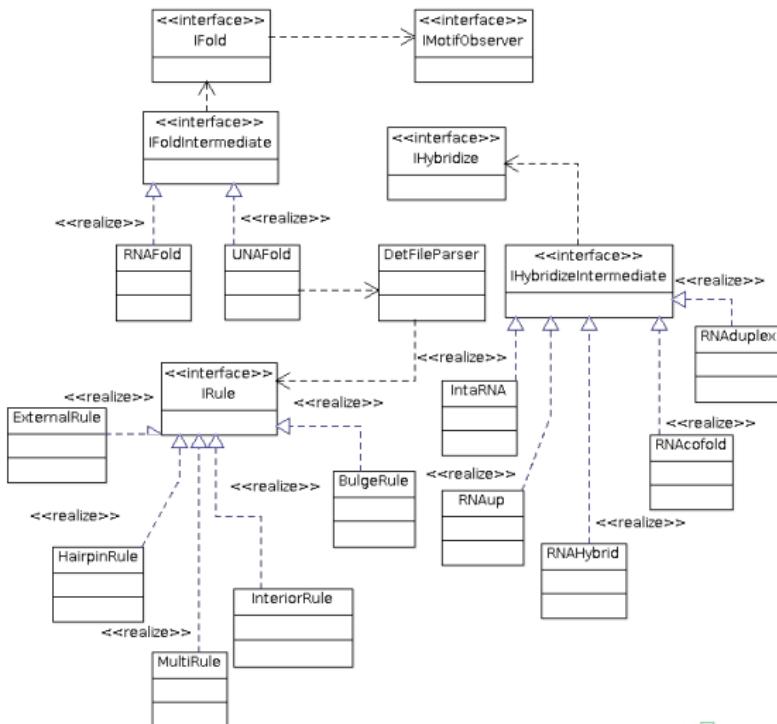


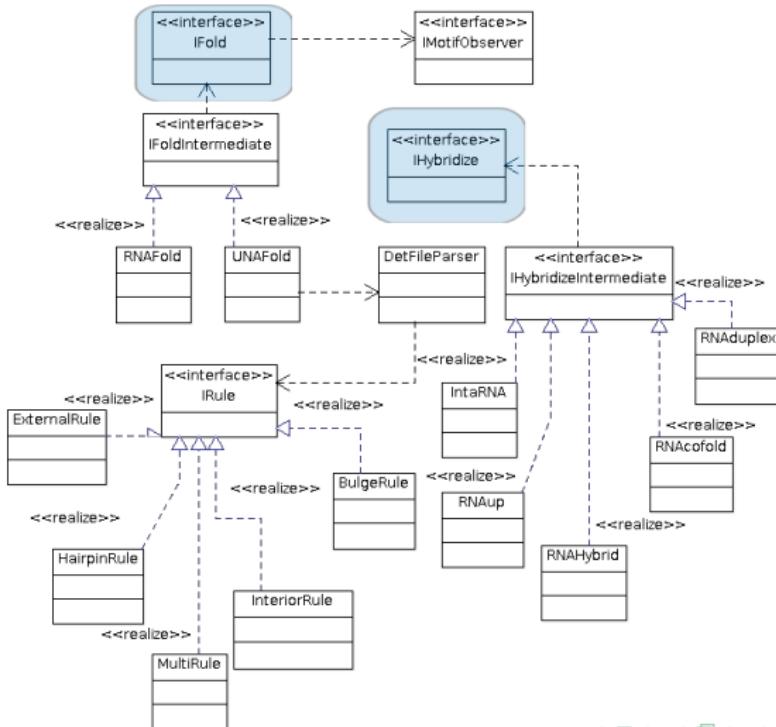
Diagrama de Clases (Cont.)



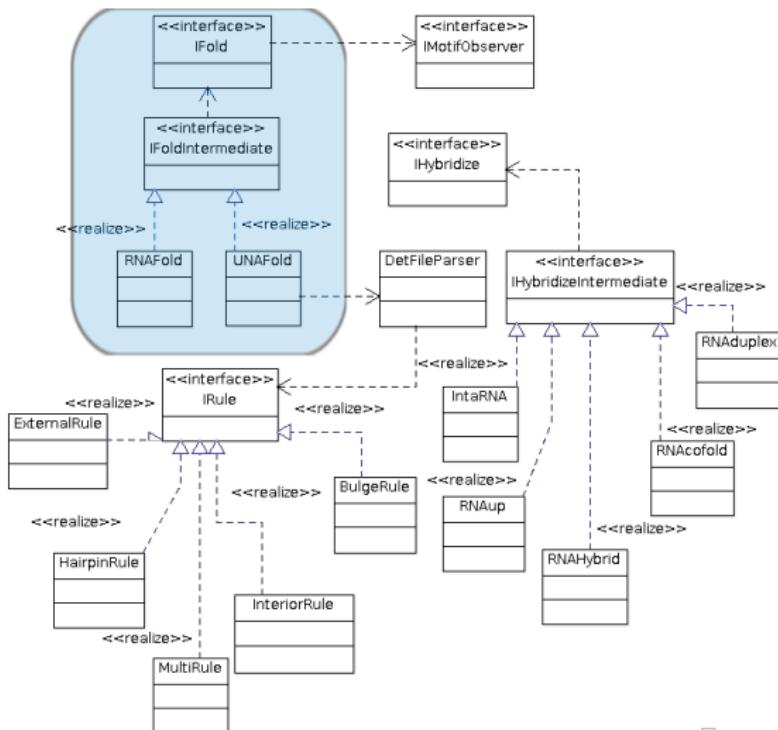
Librerías Implementadas



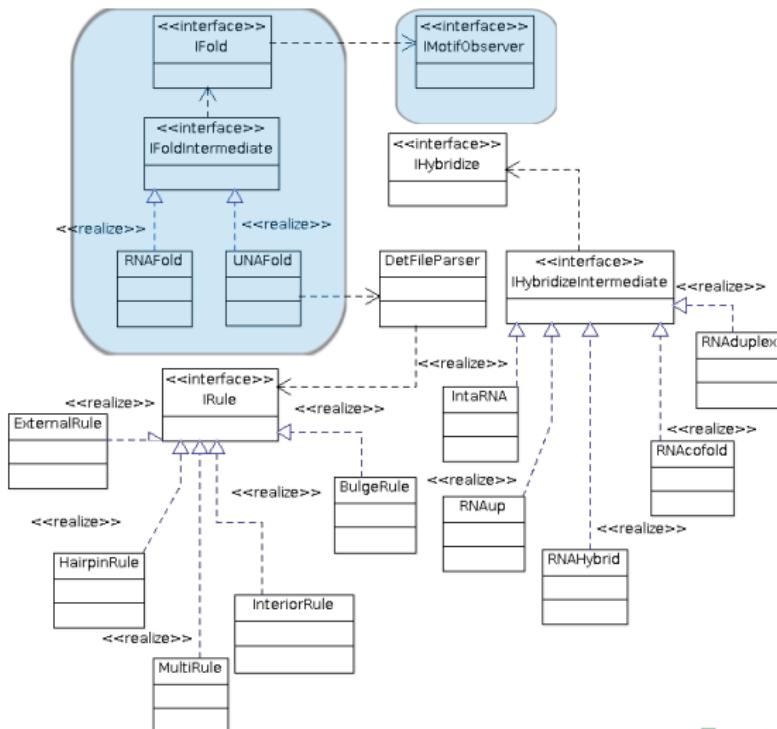
Librerías Implementadas



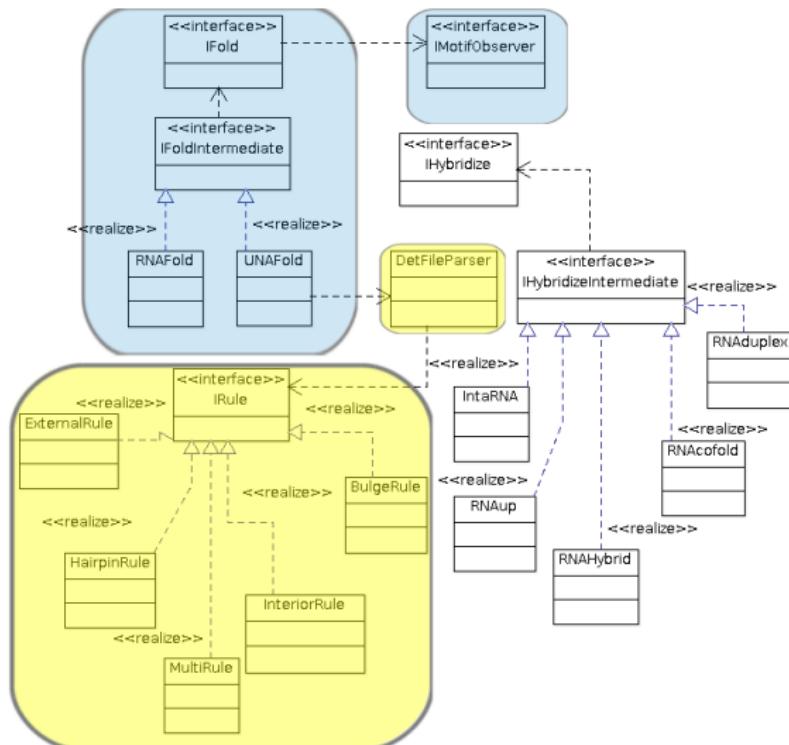
Librerías Implementadas



Librerías Implementadas



Librerías Implementadas



Dependencias

Algunas de ellas son:

- mili
- biopp
- biopp-filer
- fud

```
template <class Hierarchy>
class GenericException : public std::exception
{
protected:
    const std::string description;

public:
    explicit GenericException(const std::string& description) :
        description(description)
    {}

    ~GenericException() throw() {};

    //! what redefinition
    virtual const char* what() const throw()
    {
        return description.c_str();
    }
};
```

Métricas de Código

Files		Line Types			%
Type	Count	Blank	Comment	Source	#Comms./Tot.
C++ source	10	161	553	1108	33.2
C++ header	11	150	757	273	73.4
Total	21	311	1310	1381	48.6

Métricas CLOC para **Remo**.

Files		Line Types			%
Type	Count	Blank	Comment	Source	#Comms./Tot.
C++ source	10	161	344	967	26.2
C++ header	20	243	1222	603	66.9
Total	30	404	1566	1570	49.9

Métricas CLOC para **fideo**.

Métricas de Código (cont)

Component	Count	Blank	Comment	Source	#Comms./Tot.
r-emo	21	311	1310	1381	48.6
fideo	30	404	1566	1570	49.9
acuoso	7	72	319	272	53
etilico	8	82	327	297	52.4
r-emo client	4	45	152	133	0.46
r-emo server	8	104	272	531	0.33
Total	78	1018	3946	4184	0.48

Resumen Código producido.

Comentarios Doxygen

```
/*
 * @file      CodingSectionObtainer.h
 * @brief     CodingSectionObtainer provides the interface ...
 *
 * @author   Franco Riberi
 * @email    fgribieri AT gmail.com
 *
 * Contents: Header file for remo providing ...
 *
 * System:   remo: RNAemo — RNA research project
 * Language: C++
 *
 * @date     October 2012
 *
 * This file is part of remo.
 *
 * Copyright (C) 2012 — Franco Riberi , FuDePAN.
 *
 * Remo is free software: you can redistribute it and/or
 * modify it under the terms of the GNU General Public
 * License as published by the Free Software Foundation ,
 * either version 3 of the License , or (at your option)
 * any later version .
 *
```

Resultados

Termodinámicos

- *Estructura Secundaria:* mayor cantidad de uniones GC en el humanizado que en el original. Mayor cantidad de stacks y más largos en el humanizado que en el original.
- *miRNA vs RNA_m:* miRNA tienen una mayor capacidad de hibridar con los RNA humanizados que con los originales.

Biológicos

En los virus estudiados se pudo observar que existe un uso de codones divergente respecto de los del huésped humano. En principio significaría una menor velocidad de síntesis proteica. Es decir, al virus le convendría usar los codones humanizados, pero sin embargo disminuye la cantidad de nucleótidos disponible para aparearse, los cuales estimulan la producción de interferón.

Conclusión

Profesional

- Nuevos conceptos (SOLID, TMP)
- Profundización de conceptos
- Exigencia en calidad y eficiencia
- Trabajo interdisciplinario como forma de aplicar los conocimientos científicos en la resolución concreta de problemas que afectan la calidad de vida de las personas.

Personal

- Trabajo grupal
- Adaptación a un nuevo entorno de desarrollo
- Fuertes lazos humanos

Aportes

- A FuDePAN:

- R-emo: r-emo.googlecode.com.
- Fideo: fideo.googlecode.com.
- Acuoso: acuoso.googlecode.com.
- Etilico: etilico.googlecode.com.
- Otros: biopp.googlecode.com.

- A la comunidad científica:

- Análisis y la formalización del problema biológico.
- Solución computacional al problema.
- Publicaciones (SAV 2012 - 3CAB2C).

Trabajos Futuros

Futuros Desarrollos

- Agregar nuevos backends para folding e hibridización en *fideo*.
- Agregar nuevos backends para la humanización en *acuoso*.
- Implementar un módulo de control de marco de lectura de las secuencias.
- *Realizar el mismo estudio aplicado a proteínas.*

Trabajos relacionados indirectamente

- Pruebas de laboratorio.
- Refutación de las nuevas hipótesis.

Bibliografía I



“Biología.”

Curtis H., Sue Barnes N., Schnek A. & Flores G. 2006



“UNAFold: Software for Nucleic Acid Folding and Hybridization.”

N.R. Markham & M.Zuker. 2008



“Thermodynamics of RNA-RNA Binding.”

Muckstein et al. 2005



“siRNA, miRNA and HIV: promises and challenges.”

Man Lung YEUNG. 2005



“RNA structural motifs : building blocks of a modular biomolecule.”

D. K. Hendrix at el. 2006

Bibliografía II

-  "Thermodynamics of RNA-RNA Interaction."
U. Muckstein & H. Tafer. 2006
-  "The extent of codon usage bias in human RNA viruses and its evolutionary origin."
Gareth M. Jenkins and Edward C. Holmes. 2003
-  "Computational Methods for RNA Secondary Structure."
Zuker, Michael. 2006
-  "Fast folding and comparison of RNA secondary structures."
Hofacker and at el. 1994

Bibliografía III

-  “The C++ Programming Language, Third Edition.”
Bjarne Stroustrup. Addison-Wesley. 1997
-  “Object Design Roles, Responsibilities and Collaborations.”
Rebeca Wirfs-Brock and Alan McKean. Addison-Wesley. 2002
-  “Design Principles and Design Patterns.”
Robert C. Martin. 2000
-  “An Introduction to Software Architecture.”
Garlan Shaw. Carnegie Mellon University. 1994
-  A Framework for Small Distributed Projects and a Protein Clusterer Application.”
Guillermo Biset. UNRC. 2009

Preguntas



Agradecimientos

“La vida sin pruebas y desafíos no sería provechosa, sino.
¿Como aprenderíamos?”

Josue Alvarez. Guatemala

Muchas Gracias
por su atención...!

Agradecimientos

“La vida sin pruebas y desafíos no sería provechosa, sino.
¿Como aprenderíamos?”

Josue Alvarez. Guatemala

**Muchas Gracias
por su atención...!**