



# Desarrollo de una aplicación web para la caracterización genética de virus aviaries

*Tesina de grado, Licenciatura en Ciencias Biológicas. Sección Genética  
Evolutiva, Facultad de Ciencias, Universidad de la República*

*Estudiante: Andrés Milano*

*Orientador: Gonzalo Tomás*

*Co-orientadora: Ana Marandino*

*2017*

## Anexo I

### Desarrollo de la aplicación web

La infraestructura de la aplicación se basa en un modelo Cliente - Servidor, el cual consta de 3 partes fundamentales:

- Un servidor web
- Clientes que acceden a la información alojada en el servidor (a través de navegadores, o *browsers*)
- Conexión entre el cliente y el servidor

En este modelo de tres capas, los navegadores solicitan información al servidor vía el protocolo HTTP; el servidor provee los archivos solicitados, los cuales son desplegados por el navegador (cliente) y permiten la interacción del usuario final con los datos.

Actualmente, una de las arquitecturas más utilizadas para el desarrollo de aplicaciones es un modelo basado en 3 capas:

1. Capa de presentación
2. Capa de lógica
3. Capa de datos

Estrictamente, el sitio web desarrollado consta de dos aplicaciones:

#### ***Cliente***

Esta es la aplicación de *front-end* que renderiza las vistas en HTML (capa de presentación) y es la encargada de la interacción con el usuario. Está implementada en AngularJS (“AngularJS,” 2016), toma librerías de Bootstrap (“Bootstrap,” n.d.) y utiliza el servidor ExpressJS (“ExpressJS,” n.d.) de NodeJS (“NodeJS,” n.d.). Para el diseño y desarrollo de esta aplicación se utilizó el IDE Atom (“Atom Editor,” n.d.)

AngularJS es una plataforma de uso libre y gratuito basado puramente en Javascript, desarrollado en 2009 y mantenido actualmente por Google. Propone una estructura fuertemente orientada al patrón de diseño conocido como Model-View-Controller; su principal cometido es la separación de la presentación, la lógica de la aplicación, y el manejo del modelo de datos en capas bien definidas y poco acopladas. Este patrón de diseño permite trabajar de forma más limpia, y poder mantener y modificar un componente particular del sistema sin afectar el resto, o con un impacto mínimo. En este sentido, mientras se mantenga una estructura acorde a esta filosofía, se puede añadir el

análisis de nuevos tipos de virus sin demasiados cambios. Una característica importante de AngularJS es que su diseño permite la vinculación de los datos de la aplicación con la presentación de una manera muy sencilla, a través de directivas propias de AngularJS que extienden las etiquetas convencionales de HTML. AngularJS es compatible con los navegadores más utilizados en la industria; el manejo del código Javascript está diseñado para adaptarse automáticamente a los navegadores actuales, ya sea Mozilla Firefox, Google Chrome, Edge, etc.

La estructura resultante se concentra dentro de un directorio *app*, junto con otros directorios accesorios que alojan librerías necesarias para el funcionamiento del servidor de *NodeJS*, más archivos de configuración con especificaciones sobre el arranque de la aplicación, puertos, variables, etcétera. Está desarrollada utilizando los lenguajes de programación Javascript, HTML y CSS.

Dentro de *app* tenemos una estructura bastante simple, tomada de lo que se conoce como un *seed* de AngularJS, con los siguientes directorios:

- *css: cascade style sheets*; archivos de descripción de la presentación de las vistas (archivos HTML). Permiten darles estética a las vistas, y determinan cómo se renderizan estos archivos
- *imgs*: imágenes que utiliza el sitio (fondos, logos, filogenias estáticas, etcétera)
- *lib*: librerías utilizadas por AngularJS
- *recursos*: archivos *json* con información estática, utilizados para cargar información de manera sencilla
- *views*: vistas (páginas en HTML)
- *js*: tal vez el directorio más importante. Contiene la estructura propiamente dicha de la aplicación cliente, siguiendo la filosofía de AngularJS. Dentro de esta carpeta tenemos los *controllers*, *factories*, *services*, y *directives*, más el archivo de configuración *app.js*. Los *controllers* son archivos que se encargan de manejar la lógica del modelo. Típicamente, tenemos un archivo para cada vista. El modelo de datos es llevado al cliente por medio de las *factories*, con lo que tenemos un patrón de diseño conocido como Model-View-Controller. En *app.js* se declara el módulo principal de toda la aplicación, y se determina qué controlador maneja qué vista, y las rutas de cada vista. Se importan algunas de

las librerías utilizadas, como ser la encargada de manejar los archivos con las secuencias subidas mediante el formulario principal.

### ***Servidor***

La capa de lógica del sistema corre del lado del servidor, y está desarrollada enteramente en python, mediante el uso del paquete bioinformático Biopython (Cock et al., 2009). Python es un lenguaje con las características apropiadas para el tipo de aplicación que nos ocupa. Es un lenguaje conciso, de fácil interpretación, multiparadigma (permite la programación orientada a objetos, así como la programación procedural) y multiplataforma. Tornado es un mini-framework liviano desarrollado en python, con la capacidad de funcionar asincrónicamente, compatible con la herramienta de *front-end* propuesta. Dado que el principal cometido de la aplicación involucra el manejo de información biológica, una de las opciones que aparecen como posibles soluciones es el paquete de aplicaciones Biopython, de libre uso y con una amplia gama de módulos para la implementación de programas bioinformáticos. Biopython es un paquete de desarrollo *open-source*, con una licencia muy poco restrictiva, y cuenta con una gran comunidad de desarrolladores que aportan nuevas implementaciones y conforman un recurso valioso en cuanto a la transferencia de conocimiento. Este paquete cuenta con módulos que permiten el manejo de secuencias de ácidos nucleicos, el alineamiento de las mismas, la inferencia y la visualización de árboles filogenéticos, la conexión con bases de datos, y la manipulación de información de secuencias proteicas, entre muchas otras herramientas.

El proyecto está conformado por paquetes (similares a directorios), los cuales agrupan módulos (archivos .py con diversas funcionalidades). Así, tenemos paquetes encargados del manejo de secuencias, alineamientos, filogenias, y análisis de marcadores. Por otro lado, se agrupan módulos con funcionalidades utilitarias: manejo de constantes, errores, validaciones, limpieza de archivos, y funciones utilizadas por el resto de los paquetes. Por último, se tiene un paquete con módulos encargados de recibir y derivar las peticiones del cliente.

El servidor y *framework* donde corren estos programas es Tornado (“Tornado,” 2016). El desarrollo de esta aplicación fue hecho sobre el IDE Eclipse (“Eclipse IDE,” n.d.).

## ***Hosting***

El sistema en su conjunto está levantado y accesible en internet de forma permanente (las 24 horas) mediante dos servicios: *supervisor* mantiene AngularJS corriendo y se inicia automáticamente ante eventuales apagones de la máquina, y *pm2*, encargado de levantar y monitorear Tornado. Se accede al sitio desde internet mediante un servidor HTTP denominado nginx (“Nginx,” n.d.).

Se configuraron repositorios donde se aloja el código fuente ([www.bitbucket.org](http://www.bitbucket.org), (“Bitbucket,” n.d.) y se utiliza el sistema de versionado git (<https://git-scm.com/>, (“Github,” n.d.)) para trabajar de forma limpia y ordenada en ambientes de desarrollo y producción.

Todos los componentes del sistema se encuentran completamente alojados en un servidor virtual Linux de la Facultad de Ciencias, Universidad de la República.

## **Estructura bioinformática**

La aplicación servidor gira en torno a Tornado, por lo tanto, esta estructura contiene y se asocia con los paquetes y módulos encargados de procesar los análisis bioinformáticos. A continuación, una breve explicación de cómo están empaquetadas las diferentes funcionalidades del sistema.

### **Módulo de secuencias**

Se encarga de tomar los datos recibidos desde el cliente, y transformarlos en objetos SeqSecord de Biopython. Realiza algunas validaciones propias de una secuencia de nucleótidos, y ejecuta ciertas restricciones con el objetivo de estandarizar el método de análisis y la interpretación de los resultados: deben ser secuencias de ADN; se aceptan solamente secuencias en formato FASTA; se restringe la longitud de las secuencias ingresadas. De estas restricciones, probablemente la más importante sea la de controlar el largo mínimo que deben tener las secuencias ingresadas para ser consideradas válidas. En el caso de utilizar secuencias significativamente más largas que nuestro alineamiento de referencia, los programas de MSA y de inferencia filogenética se encargan de quitar información sobrante. Por otro lado, se apoya en el módulo utilitario para validaciones más generales.

## Módulo de alineamientos

Este módulo tiene funciones para trabajar con objetos de tipo `AlignIO.Alignment`, y maneja los *wrappers* para los programas de alineamiento múltiple, como `ClustalOmegaCommandLine`, o `MuscleCommandLine`.

## Módulo de inferencias filogenéticas

Aquí tenemos todo el manejo de los algoritmos de inferencia filogenética; funciones que reciben alineamientos y devuelven estructuras en formato Newick que representan las relaciones filogenéticas, accesibles por programas de visualización de árboles. Se relaciona con otros módulos utilitarios, necesarios para realizar conversiones de formato, por ejemplo, de FASTA a PHYLIP (único formato considerado válido para el programa de inferencia filogenética PhyML).

## Módulo de visualización de árboles

Se encarga de procesar árboles en formato Newick, y devuelve imágenes con determinado formato, resaltando la secuencia problema. Utiliza la librería *ete2* de la plataforma ETE ToolKit.

## Módulo de análisis de marcadores moleculares

En este módulo se ubica todo el manejo correspondiente a la obtención de los aminoácidos presentes en la secuencia problema, y su posterior análisis. Consiste en pequeños *scripts* que se ocupan de recorrer secuencias de ADN, traducirlas, ubicarlas en los marcos de lectura correctos, y calcular similitudes con respecto a la configuración de marcadores presentes en cada linaje de las cepas de Gumboro.

## Algoritmos utilizados para el análisis filogenético de una secuencia problema

Para **IBDV**, utilizamos PhyML. Se debe convertir el alineamiento en formato FASTA, a un formato aceptado por PhyML (PHYLIP). Se utiliza el *wrapper* de biopython, con los siguientes parámetros:

```
datatype = "nt"  
modelo_sust = "GTR"  
bootstrap = "-4"  
params_opt = "lr"  
alpha = "e"  
prop_invar = "e"
```

Para **IBV**, se utiliza el *wrapper* para FastTree, con los siguientes parámetros:

```
nt = "True"  
gtr = "True"  
gamma = "True"  
cat = 4
```

## Bibliografía

---

---

AngularJS. (2016). Retrieved from <https://angularjs.org/>

Atom Editor. (n.d.). Retrieved from <https://atom.io/>

Bitbucket. (n.d.).

Bootstrap. (n.d.). Retrieved from <http://getbootstrap.com>

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., ... de Hoon, M. J. L. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), 1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>

Eclipse IDE. (n.d.). Retrieved from <https://www.eclipse.org/>

ExpressJS. (n.d.).

Github. (n.d.). Retrieved from <https://git-scm.com/>

Nginx. (n.d.). Retrieved from <https://nginx.org/en/>

NodeJS. (n.d.). Retrieved from <https://nodejs.org/es/>

Tornado. (2016). Retrieved from <http://www.tornadoweb.org/en/stable/>



## Contenido

Contenido .....	2
Resumen .....	4
Introducción y antecedentes .....	5
Industria avícola .....	5
Virus de Gumboro.....	6
Enfermedad .....	6
Estructura y genoma.....	6
Clasificación.....	7
Herramientas de control .....	8
Virus de la Bronquitis Infecciosa .....	9
Enfermedad .....	9
Estructura y genoma.....	9
Clasificación.....	11
Herramientas de control .....	12
Importancia y alcance de la genotipificación en IBV e IBDV .....	12
Genotipificación correcta de IBV e IBDV .....	13
Hipótesis.....	14
Objetivos .....	14
Objetivo general .....	14
Objetivos específicos .....	14
Metodología .....	15
Estrategia.....	15
Insumos previos al análisis de una secuencia problema.....	16
Obtención de las secuencias de referencia .....	16
Generación de alineamientos de referencia y reconstrucciones filogenéticas.....	16
Relevamiento y análisis de marcadores moleculares aminoacídicos en IBDV.....	16
Análisis de una secuencia problema.....	17
Alineamiento frente al set de referencia.....	18
Clasificación mediante análisis filogenéticos.....	18
Visualización de árboles filogenéticos .....	19
Análisis de marcadores moleculares aminoacídicos .....	19
Desarrollo de la aplicación web .....	19
Resultados .....	20

Insumos previos al análisis de una secuencia problema.....	20
Alineamientos de referencia.....	20
Reconstrucciones filogenéticas .....	20
Análisis de marcadores moleculares .....	23
Estructura de la aplicación web y caso de uso principal .....	24
Discusión.....	35
Importancia de la genotipificación.....	35
Secuencias de referencia.....	35
Alineamientos e inferencias filogenéticas .....	36
Marcadores moleculares aminoacídicos.....	37
Características y ventajas de la aplicación VirusAviars .....	37
Perspectivas.....	38
Agradecimientos .....	39
Bibliografía .....	40

## Resumen

La enfermedad de Gumboro y la Bronquitis Infecciosa son dos patologías virales altamente contagiosas que afectan la industria avícola, generando grandes pérdidas económicas a nivel mundial. Ambos virus contienen genomas de ARN, lo cual ha facilitado su diversificación genética, existiendo hoy en día variantes virales que difieren en su manifestación clínica y composición antigénica. La correcta identificación de la variante viral circulante en un determinado lugar es fundamental para el control de estos virus. La aproximación más comúnmente utilizada para lograr esa identificación es a través de la caracterización genética, ya sea mediante análisis filogenéticos y/o la identificación de marcadores moleculares aminoacídicos. Sin embargo, aún hoy existen ciertas discrepancias en cuanto a la caracterización genética de algunas cepas, principalmente por la falta de uniformidad en cuanto a variables como la región genómica utilizada, las cepas de referencia utilizadas e incluso el método empleado para obtener las relaciones filogenéticas. Teniendo esta problemática en consideración, en el presente trabajo se diseñó una aplicación web que permite realizar de forma rápida, sencilla y de fácil interpretación, la caracterización genética de una muestra viral problema. La aplicación incorpora todo el conocimiento previo generado en cuanto a la caracterización genética de ambos virus y propone e implementa herramientas de clasificación que consisten en análisis filogenéticos y/o la identificación de marcadores moleculares aminoacídicos. Para el desarrollo de esta aplicación, en una primera instancia se realizó un detallado análisis de las secuencias disponibles de ambos virus, procurando incluir variantes virales de todos los linajes/genotipos existentes en cada caso. Una vez obtenida una adecuada base de datos de secuencias virales y habiendo optimizado los análisis filogenéticos y de marcadores moleculares aminoacídicos, se procedió con la adaptación de esa información para su utilización en una aplicación web a través de herramientas bioinformáticas de libre acceso. Como resultado, se presenta la aplicación web <http://gevolutiva.fcien.edu.uy/>. Además de brindar información acerca de la clasificación de las variantes virales analizadas, esta aplicación web pretende nuclear y difundir el conocimiento adquirido en torno a los virus causantes de estas enfermedades, ofreciendo links de interés, publicaciones recientes, proyectos de investigación en marcha e incluso un formulario de contacto donde presentar dudas y/o sugerencias.

## **Introducción y antecedentes**

### **Industria avícola**

El consumo de carne de ave ha ido aumentando progresivamente en los últimos años, local y mundialmente. En Uruguay, existe un marcado incremento en la producción y consumo de carne de ave. En el período 2010-2015 el consumo de carne por habitante por año se incrementó un 49,9 %, pasando de 20,4 a 30,5 kg/hab/año (DIEA, 2016). Esta misma tendencia ocurre a nivel global desde hace años. De acuerdo a proyecciones publicadas recientemente, la producción mundial de carne de ave crecerá en el período 2016-2025 un 19,01 %, transformándose en el tipo de carne más producida y consumida en el mundo (OECD/FAO, 2016).

El liderazgo en el comercio mundial alcanzado por la industria avícola se debe esencialmente a la creciente demanda de sus productos y a su impulso productivo. El desarrollo y la transferencia de tecnologías de producción, faena y elaboración, han mejorado la inocuidad y eficiencia de los productos avícolas, satisfaciendo las exigencias de la demanda, y en consecuencia, estimulando la formación de unidades productivas a gran escala.

El aspecto sanitario es uno de los factores de mayor impacto en la avicultura ya que incide en los parámetros zootécnicos y puede convertirse en un fundamento de restricciones comerciales en el mercado internacional. La Enfermedad Infecciosa de la Bursa, o Gumboro, y la Bronquitis Infecciosa, se destacan como dos de las principales y más relevantes patologías virales que afectan hoy en día a esta industria. Para controlar a estos agentes patógenos es imprescindible conocer qué variantes virales están presentes en cada región en particular, para así aplicar estrategias de control específicas, y por ende, más eficientes.

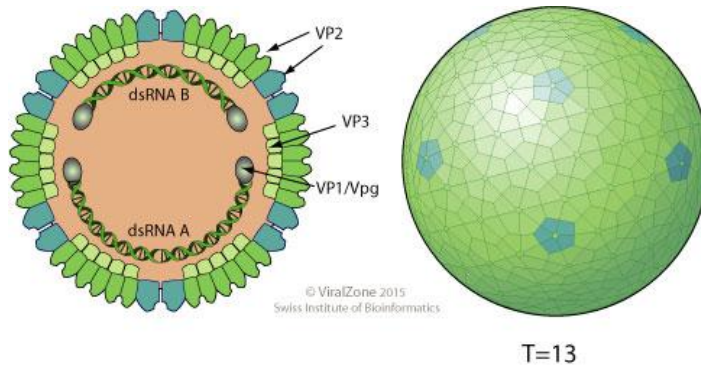
## Virus de Gumboro

### Enfermedad

La enfermedad de Gumboro es una patología que afecta el sistema inmune de las aves (Hirai, Funakoshi, Nakai, & Shimakura, 1981). Se encuentra distribuida mundialmente y está fuertemente asociada con la cría comercial de aves, tanto en líneas de producción de carne como de huevos. Tiene la particularidad de restringirse únicamente a aves jóvenes, de entre 2 y 6 semanas de edad, debido a que el principal órgano afectado (bursa de Fabricius) se atrofia de forma natural en aves de mayor edad (Etteradossi & Saif NM, 2008). Las aves afectadas pueden cursar desde una enfermedad subclínica leve hasta una enfermedad grave, con altas tasas de mortalidad. Esto depende de varios factores, entre los que se destaca la cepa viral involucrada y el estado inmune de las aves.

### Estructura y genoma

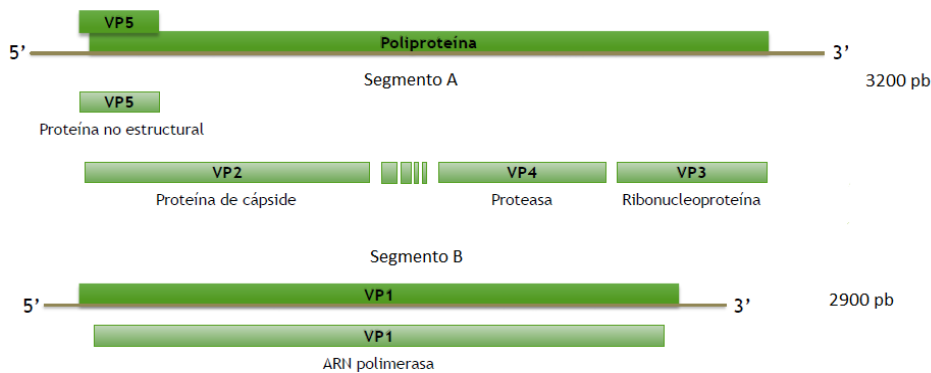
El virus de Gumboro o de la Enfermedad Infecciosa de la Bursa (IBDV, por sus siglas en inglés) pertenece a la familia Birnaviridae, género Avibirnavirus. Es un virus no envuelto, con una cápside icosaédrica de aproximadamente 60 nm de diámetro (Figura 1).



**Figura 1** – Representación esquemática de una partícula viral de IBDV.

Como el nombre de su familia lo indica, es un virus de ARN doble hebra, bisegmentado; el segmento A (3,3 Kb) contiene dos marcos de lectura abiertos: uno codifica la proteína no estructural VP5, y el otro, solapado con el anterior, codifica una poliproteína que se procesa para generar la proteína de cápside VP2, la nucleoproteína VP3 y la proteasa viral VP4 (Figura 2). El segmento B (2,9 Kb) codifica la ARN polimerasa dependiente de ARN, denominada VP1.

VP2 es el principal componente de la cápside viral. Es la proteína encargada de interactuar con los receptores celulares para mediar el ingreso del virus a la célula, y es también importante en la respuesta inmune del hospedero ya que contiene los principales determinantes antigénicos de la respuesta inmune humoral (Delgui, Gonzalez, & Rodriguez, 2009; Fahey, Erny, & Crooks, 1989).



**Figura 2** - Esquema representando los dos segmentos del genoma de IBDV y sus productos proteicos.

### Clasificación

En cuanto a la clasificación viral, se han caracterizado dos serotipos, 1 y 2, estando solamente el 1 asociado a virus patogénicos (Ismail, Saif, & Moorhead, 1988; McFerran et al., 1980). Aunque existen criterios patogénicos, antigénicos y genéticos para la clasificación, hay una clara tendencia hacia el uso de estos últimos, por ser técnicas más rápidas, sensibles y fáciles de estandarizar. El método más robusto y empleado para la caracterización genética de IBDV es el análisis filogenético de la región hipervariable del gen de la proteína VP2. A través de este tipo de análisis es posible agrupar todas las

variantes virales existentes en cinco cepas o linajes: cepas clásicas virulentas (cvIBDV), cepas variantes (vaIBDV), cepas hipervirulentas (vvIBDV), cepas vacunales atenuadas, también llamadas clásicas atenuadas (atIBDV) y cepas divergentes o *distinct* (dIBDV) (Hernández et al., 2015). Prácticamente todas las cepas tienen distribución mundial, a excepción de las vaIBDV que se encuentran casi exclusivamente en algunos países de América. Paralelamente, existen marcadores moleculares nucleotídicos y aminoacídicos que permiten clasificar las cepas. Los marcadores moleculares nucleotídicos se han usado para desarrollar métodos de clasificación basados en RT-PCR/RFLP y PCR en tiempo real (Hernández et al., 2011; Tomas et al., 2012). Estos métodos tienen la ventaja de ser rápidos, pero no dejan de ser métodos de clasificación indirectos, los cuales presentan limitaciones con respecto a la cantidad de información que se puede obtener a partir de ellos. Los marcadores moleculares aminoacídicos, por otro lado, permiten obtener información adicional luego de la clasificación filogenética. Por ejemplo, hay evidencias fuertes que asocian a determinados marcadores con la atenuación y capacidad de infectar cultivos celulares, y con la disminución de patogenicidad *in vivo* (Noor et al., 2014; Van Loon, De Haas, & Mundt, 2002). Otros marcadores se asocian con cambios a nivel antigénico, a través de ensayos que miden la capacidad de un virus para unirse a distintos paneles de anticuerpos monoclonales (Etteradossi et al., 1999; Letzel et al., 2007). Por último, algunos marcadores se han asociado con un tipo específico de cepa y suelen ser analizados en conjunto con la filogenia para la clasificación viral (Hernández et al., 2015; Rudd, Heine, Sapats, Parede, & Ignjatovic, 2002; Xia, Wang, Huang, & Zhang, 2008).

### Herramientas de control

El control de este virus se realiza a través de la aplicación de estrictos manejos biosanitarios y de la vacunación. Hay diversos tipos de vacunas, siendo las más utilizadas aquellas desarrolladas a partir de virus vivos atenuados y vacunas inactivadas, pero existiendo también vacunas recombinantes y vacunas de complejo antígeno-anticuerpo (Muller, Mundt, Etteradossi, & Islam, 2012). La estrategia más utilizada para generar una protección continua en las aves implica la hiperinmunización de las aves reproductoras, las cuales transfieren al huevo y al pollito una determinada cantidad de anticuerpos maternos (inmunidad pasiva), y la posterior aplicación de vacunas en campo para perpetuar la protección una vez que los anticuerpos maternos son metabolizados por los pollitos (inmunidad activa). Las vacunas a virus vivos atenuados son

desarrolladas a partir de cepas cvIBDV, atIBDV, vaIBDV, y en menor medida, vvIBDV. No existen hasta el día de hoy vacunas desarrolladas a partir de cepas dIBDV, probablemente debido a lo reciente de su descripción (Hernández et al., 2015). Si bien hay vacunas desarrolladas a partir de cuatro cepas diferentes, desde un punto de vista antigénico existen dos grandes grupos o protectotipos: por un lado, las cepas vaIBDV, y por el otro lado, las cepas cvIBDV, atIBDV, y vvIBDV. El término protectotipo, originalmente empleado en el virus de la Bronquitis Infecciosa pero aplicable también a IBDV, define grupos de cepas o variantes que muestran un nivel adecuado de protección cruzada (de Wit et al., 2011). Aún no se ha determinado a qué protectotipo pertenecen las cepas dIBDV. A partir del diagnóstico y clasificación de IBDV se pueden realizar planes de control efectivos, aplicando las vacunas más adecuadas y tomando las medidas biosanitarias acorde.

## **Virus de la Bronquitis Infecciosa**

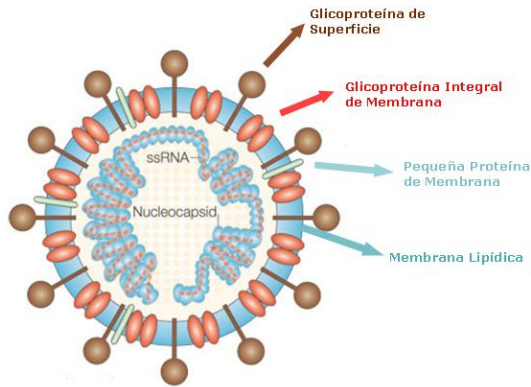
### **Enfermedad**

La Bronquitis Infecciosa es una enfermedad respiratoria que afecta aves de corral de todas las edades, aunque se ha visto que es más grave en aves de hasta dos semanas de edad (Crinion, Ball, & Hofstad, 1971; Muneer, Newman, Halvorson, Sivanandan, & Coon, 1987). Las aves afectadas presentan signos clínicos respiratorios tales como tos, estornudos, exudados nasales y oculares, estridor y disnea (Schalk & Hawn, 1931). En aves ponedoras produce problemas de baja de postura, además de provocar el descenso en la calidad de los huevos, tanto de la cáscara como de la yema y albúmina (Crinion et al., 1971). También se han reportado casos de aves enfermas con problemas renales y aumento de mortalidad, principalmente en aves jóvenes (Winterfield & Hitchner, 1962). Al igual que en IBDV, la gravedad de esta enfermedad depende principalmente de la cepa viral involucrada así como del estado inmunitario de las aves afectadas.

### **Estructura y genoma**

El virus de la Bronquitis Infecciosa (IBV) pertenece al género Gammacoronavirus de la familia Coronaviridae. Es un virus pleomórfico (80-120 nm), envuelto por una bicapa





**Figura 3-** Representación esquemática de una partícula viral de IBV.

lipídica con proyecciones superficiales denominadas espículas (Berry, Cruickshank, Chu, & Wells, 1964) (Figura 3). Posee un genoma de ARN simple hebra, de polaridad positiva, no segmentado, con aproximadamente 27,6 Kb de longitud (Figura 4). Este genoma codifica cuatro proteínas estructurales: una glicoproteína de superficie (S), una glicoproteína de membrana (M), una proteína de envoltura (E) y una proteína de nucleocápside (N) (Stern & Sefton, 1982). También codifica para al menos 19 proteínas no estructurales, entre las que se encuentra la ARN polimerasa, involucradas en la replicación y la transcripción de los genes virales. La proteína S es particularmente importante, ya que es la mayor inductora de anticuerpos neutralizantes y la responsable de la unión del virus y de su entrada a las células huésped (Lai & Cavanagh, 1997). Esta proteína, ubicada en la superficie de la envoltura viral, es clivada post-traduccionalmente formando las subunidades S1 y S2.



**Figura 4 -** Representación esquemática del genoma de IBV. El gen 2 codifica para la glicoproteína de superficie S.

## Clasificación

De forma análoga a lo que ocurre con IBDV, existen varios criterios para clasificar IBV, siendo la caracterización genética la más utilizada actualmente. La porción S1 es especialmente divergente, ya que contiene tres regiones hipervariables (Liu et al., 2009). Por lo tanto, el análisis de la región del gen S1 (parcial o completa) suele utilizarse para clasificar a IBV. Recientemente, en un intento por lograr una clasificación robusta y estándar para las cepas circulantes de IBV, Valastro *et al.* (2016) hicieron una búsqueda exhaustiva de secuencias completas del gen de S1 y realizaron un análisis profundo de las relaciones filogenéticas entre estas secuencias, resultando en una clasificación armoniosa y estable, con una nomenclatura unificada y simple de utilizar. Dicha clasificación propone la existencia de 32 linajes agrupados en 6 genotipos (Tabla 1), y fue rápidamente aceptada por la comunidad científica especializada en la temática (Brandao et al., 2016; Marandino et al., 2017; Zanaty et al., 2016).

**Tabla 1** – Cepas de IBV prototipo de cada linaje, con su período de circulación.

Linaje	Período de circulación	Cepa prototipo			
		Nombre	País de origen	Fecha de colecta	N° de acceso en GenBank
GI-1	1937–2013	Beaudette	EEUU	1937	M95169
GI-2	1954–2006	Holte	EEUU	1954	GU393336
GI-3	1960–2006	Gray	EEUU	1960	L14069
GI-4	1962–1998	Holte	EEUU	1962	L18988
GI-5	1962–2012	N1/62	Australia	1962	U29522
GI-6	1962–2010	VicS	Australia	1962	U29519
GI-7	1964–2012	TP/64	Taiwan	1964	AY606320
GI-8	1965–1967	L165	EEUU	1965	JQ964061
GI-9	1973–2011	ARK99	EEUU	1973	M99482
GI-10	1970s–2000s	B	Nueva Zelanda	1970s	AF151954
GI-11	1975–2009	UFMG/G	Brasil	1975	JX182775
GI-12	1978–2006	D3896	Países Bajos	1978	X52084
GI-13	1983–2013	Moroccan-G/83	Marruecos	1983	EU914938
GI-14	1984–2006	B1648	Bélgica	1984	X87238
GI-15	1986–2008	B4	Corea	1986	FJ807932
GI-16	1986–2011	IZO 28/26	Italia	1986	KJ941019
GI-17	1988–1999	CA/Machado/88	EEUU	1988	AF419315
GI-18	1993–1999	JP8127	Japón	1993	AY296744
GI-19	1993–2012	58HeN-93II	China	1993	KC577395

GI-20	1996–1999	Qu_mv	Canada	1996	AF349621
GI-21	1997–2005	Spain/97/314	España	1997	DQ064806
GI-22	1997–2011	40GDGZ-97I	China	1997	KC577382
GI-23	1998–2012	Variant 2	Israel	1998	AF093796
GI-24	1998–2013	V13	India	1998	KF757447
GI-25	2004–2013	CA/1737/04	EEUU	2004	EU925393
GI-26	2006–2007	NGA/B401/2006	Nigeria	2006	FN182243
GI-27	2008–2013	GA08	EEUU	2008	GU301925
GII-1	1979–1984	D1466	Países Bajos	1979	M21971
GIII-1	1988–2008	N1/88	Australia	1988	U29450
GIV-1	1992–2003	DE/072/92	EEUU	1992	U77298
GV-1	2002–2008	N4/02	Australia	2002	DQ059618
GVI-1	2007–2012	TC07-2	China	2007	GQ265948

### Herramientas de control

Al igual que ocurre con IBDV, el control de este virus se sustenta en la implementación de buenos manejos biosanitarios y la aplicación de vacunas, tanto a virus vivo atenuado como inactivadas. Existen decenas de vacunas de serotipos diferentes a nivel mundial. Esto se debe a la existencia de múltiples variantes antigénicas circulando en todo el mundo, las cuales no tienen buen nivel de protección cruzada, y por lo tanto es necesario el desarrollo de nuevas vacunas cada vez que surge una variante antigénica. En la mayoría de los países sudamericanos, las únicas vacunas autorizadas oficialmente provienen de cepas del serotipo Massachusetts, como es el caso de Uruguay. A pesar de que la vacunación es una práctica que está muy extendida y se realiza prácticamente sin excepciones, siguen reportándose casos de la enfermedad (Cavanagh, 2003).

### Importancia y alcance de la genotipificación en IBV e IBDV

La rápida detección y clasificación de estos virus frente a un eventual brote resulta fundamental, debido a que la aplicación de tratamientos exitosos depende enormemente de la correcta clasificación de las cepas. En el caso del virus de Gumboro, las cepas variantes deben ser controladas con vacunas hechas a partir de cepas homólogas, ya que la inmunización con cepas clásicas virulentas o clásicas atenuadas (vacunas más comunes) no otorga buena protección a las aves. Por otro lado, las cepas hipervirulentas son capaces de romper tempranamente con la protección generada por la inmunidad pasiva, por lo cual para su control se recomienda adelantar la aplicación de vacunas en

campo, eligiendo vacunas más agresivas que las normalmente utilizadas.

En cuanto al virus de la Bronquitis Infecciosa, existen decenas de serotipos reportados, los cuales muestran un grado variable de protección cruzada (de Wit et al., 2011). A su vez, los serotipos de IBV tienen una alta correlación con los genotipos formados tras análisis genéticos de la proteína de superficie S1, y por lo tanto se asume que dos genotipos cercanamente emparentados probablemente tengan un grado de protección cruzada mayor comparado con dos serotipos menos emparentados (Ladman, Loupos, & Gelb, 2006).

Este nivel de correlación entre genotipos/linajes y la protección vacunal observada en ambos virus hace que la genotipificación sea una herramienta de mucho valor no sólo para implementar estudios evolutivos y epidemiológicos, sino también para obtener información útil a la hora de tomar medidas de control (Berg, 2000; de Wit et al., 2011).

#### **Genotipificación correcta de IBV e IBDV**

Del alcance que tiene la genotipificación de estos virus se desprende la importancia de contar con herramientas prácticas y eficientes que permitan realizarla de manera sencilla y mediante métodos estandarizados. Actualmente, la metodología más utilizada para genotipificar a los virus de Gumboro y Bronquitis Infecciosa es la amplificación por PCR, secuenciación de regiones genómicas informativas y su posterior análisis filogenético (Bing et al., 2007; de Wit et al., 2011). Esta metodología brinda la mayor información posible en cuanto a la composición nucleotídica de las secuencias estudiadas, e implica la utilización de tecnologías de secuenciación genómica y el uso de herramientas bioinformáticas de edición y análisis de secuencias. Estos últimos análisis deben ser realizados siguiendo ciertos lineamientos y utilizando los programas y algoritmos adecuados, ya que de lo contrario se corre el riesgo de clasificar erróneamente una cepa. Consideraciones tales como qué cantidad y cuáles secuencias utilizar como referencia en los análisis, qué tamaño y qué región genómica es más adecuado analizar o qué algoritmo implementar para alinear y posteriormente determinar las relaciones filogenéticas, son algunos de los principales puntos que deben ser tenidos en cuenta.

## **Hipótesis**

Las cepas circulantes de los virus de Gumboro y Bronquitis Infecciosa pueden ser clasificadas de forma concisa a través de la caracterización genética utilizando herramientas bioinformáticas implementadas mediante una aplicación web de acceso público.

## **Objetivos**

### **Objetivo general**

Desarrollar una aplicación web de acceso público que permita a los usuarios, a partir del ingreso de una secuencia de ADN, clasificar genéticamente a los virus de Gumboro y de la Bronquitis Infecciosa.

### **Objetivos específicos**

- Generar, para cada virus, una base de datos con secuencias de referencia, a partir de las cuales se puedan obtener filogenias que reconstruyan los linajes y/o genotipos descritos hasta ahora.
- Obtener información detallada acerca de los marcadores aminoacídicos descritos hasta el momento, útiles para clasificar y conocer características patogénicas y antigénicas del virus de Gumboro.
- Programar y evaluar el funcionamiento de una serie de herramientas bioinformáticas que permitan, a partir del ingreso de una secuencia problema del virus de Gumboro o de la Bronquitis Infecciosa, clasificar a dicha secuencia a través de una filogenia y brindar información acerca de los marcadores aminoacídicos en el caso del virus de Gumboro.
- Generar una aplicación web que trabaje en base a los programas bioinformáticos diseñados previamente, que sea de fácil interpretación y utilización, y que permita a los usuarios obtener la clasificación de un virus problema e información adicional acerca del virus analizado.

## Metodología

### Estrategia

Para lograr los objetivos específicos planteados nos propusimos seguir la siguiente serie de pasos:

- Obtener las secuencias de referencia para cada virus.
- Generar alineamientos de referencia que permitan reconstruir las relaciones filogenéticas esperadas.
- Relevar información sobre los marcadores moleculares de IBDV.

Estos pasos son los primeros que debieron realizarse, ya que son el insumo necesario para el desarrollo y funcionamiento de la aplicación. Pueden considerarse como requerimientos imprescindibles y estáticos; sus resultados permanecen incambiables a lo largo del uso de la aplicación, y una vez almacenados no vuelven a modificarse.

El análisis de una secuencia problema implica los siguientes pasos:

- Alinear una secuencia problema frente al alineamiento de referencia correspondiente.
- Inferir las relaciones filogenéticas entre la secuencia problema y las secuencias de referencia.
- Visualizar en pantalla el árbol filogenético resultante, resaltando la secuencia problema.
- Analizar la presencia o ausencia de determinados marcadores moleculares aminoacídicos en la secuencia problema (solamente en el caso de IBDV), y presentar información que acompañe la clasificación de la cepa problema en base a estos resultados.

**Comentario [L1]:** Esto en ambos virus

## **Insumos previos al análisis de una secuencia problema**

### **Obtención de las secuencias de referencia**

El set de secuencias utilizado para la clasificación de las cepas de IBDV fue obtenido de la base de datos del GenBank, específicamente aquellas correspondientes a la región hipervariable del gen VP2. Se buscaron aquellas secuencias correspondientes a cepas de referencia de cada uno de los linajes de IBDV. Las cepas de referencia fueron elegidas en base al nivel de caracterización genética, antigénica y patogénica, considerando más adecuadas aquellas que hubieran sido analizadas en mayor profundidad en los tres niveles.

En el caso de IBV, el alineamiento de referencia está basado en las secuencias utilizadas por Valastro et al. (Valastro et al., 2016). La región genómica utilizada en este caso es la del gen de la proteína S1, cuyas secuencias fueron descargadas de la base de datos del GenBank.

### **Generación de alineamientos de referencia y reconstrucciones filogenéticas**

A partir de las secuencias de referencia de cada virus se generaron alineamientos de referencia utilizando el programa de alineamiento múltiple MAFFT (Katoh, Misawa, Kuma, & Miyata, 2002) con los parámetros por defecto. Luego se determinó el modelo evolutivo adecuado para cada alineamiento utilizando el programa jModelTest2 (Darriba, Taboada, Doallo, & Posada, 2012). Finalmente se realizaron las inferencias filogenéticas utilizando el programa PhyML (Guindon et al., 2010), a través del sitio web <http://www.phylogeny.fr/> (Dereeper et al., 2008).

### **Relevamiento y análisis de marcadores moleculares aminoacídicos en IBDV**

El análisis de marcadores aminoacídicos en IBDV permite obtener información adicional acerca de la secuencia analizada. Como ya se dijo más arriba, existe evidencia que asocia determinados marcadores aminoacídicos con cepas atenuadas o adaptadas a cultivos celulares, o también con cepas que tienen cambios antigénicos. Además, algunos marcadores fueron propuestos como típicamente presentes en determinadas cepas o linajes virales. Para investigar este tema, se realizó una búsqueda bibliográfica exhaustiva de trabajos que propongan marcadores moleculares aminoacídicos en IBDV. Luego se descargaron de la base de datos del GenBank la totalidad de las secuencias de IBDV que contenían al menos la región hipervariable del gen VP2. Las secuencias fueron alineadas utilizando el programa MAFFT (Katoh et al., 2002), para luego

realizar un análisis filogenético de máxima verosimilitud aproximada con el programa FastTree (Price, Dehal, & Arkin, 2009). Todas las secuencias pertenecientes a un mismo linaje según el análisis filogenético fueron agrupadas en un mismo alineamiento, obteniendo así cinco sets de secuencias, cada uno correspondiente a un linaje de IBDV. Por último, se analizó la frecuencia de cada marcador aminoacídico propuesto para cada uno de los linajes (Figura 5). De esta forma, se identificaron y seleccionaron aquellos marcadores más robustos para ser incluidos en el análisis de una secuencia problema.

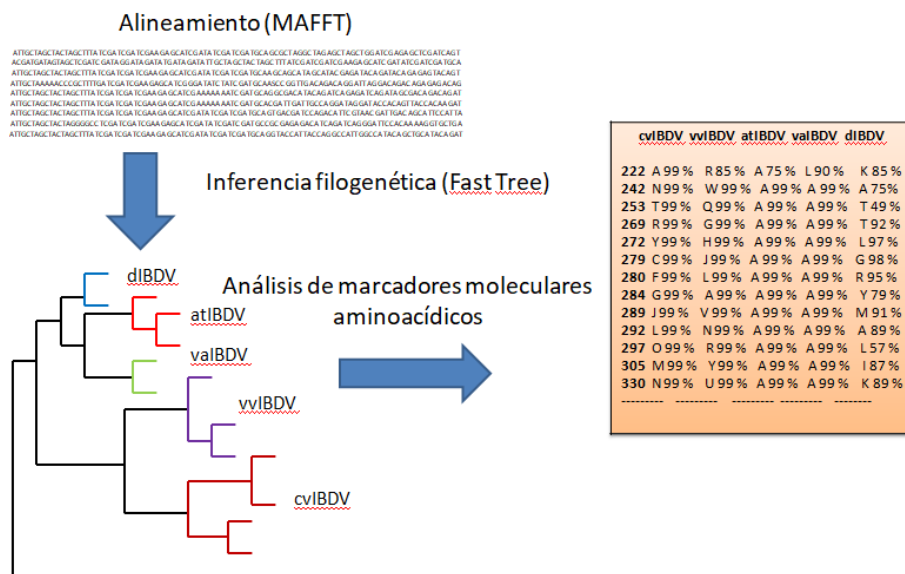


Figura 5 - Esquema representando las etapas del análisis de marcadores moleculares aminoacídicos de IBDV. Los datos son ilustrativos.

### Análisis de una secuencia problema

Una vez obtenidos los insumos necesarios descritos previamente, quedan establecidas las condiciones para analizar secuencias problema. El flujo del análisis puede verse esquematizado en la figura 6.



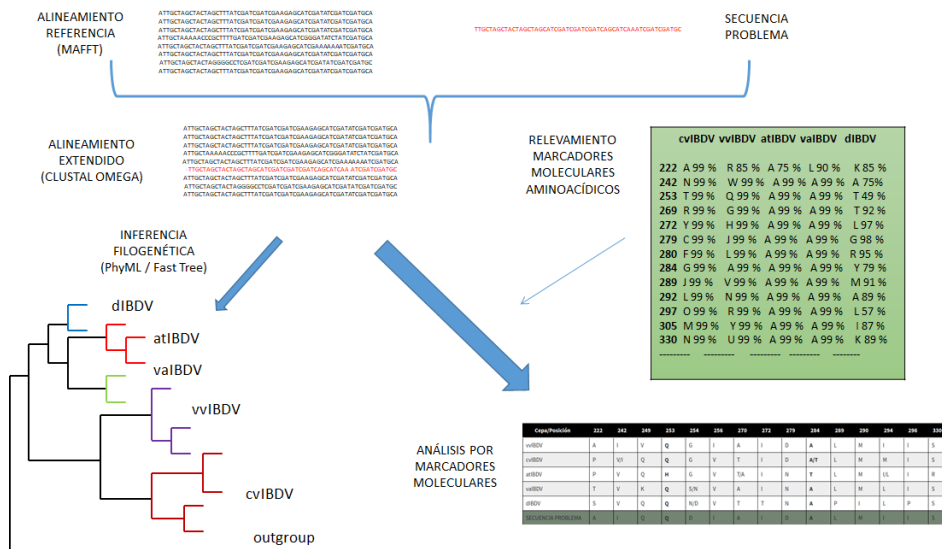


Figura 6 - Esquema representando el flujo del análisis de una secuencia problema.

**Alineamiento frente al set de referencia**

Los dos tipos de análisis disponibles en la aplicación (análisis filogenético y análisis de marcadores moleculares aminoacídicos en el caso de IBDV), requieren que la secuencia problema ingresada por el usuario sea alineada con el set de secuencias de referencia correspondiente, previamente alineado. Para esto, se utilizó el programa de alineamiento múltiple Clustal Omega (Sievers et al., 2011), el cual permite considerar el alineamiento como un perfil, manteniendo intactas las columnas del alineamiento original.

**Clasificación mediante análisis filogenéticos**

En el caso de IBDV, para reconstruir las relaciones filogenéticas entre una secuencia problema y las secuencias del alineamiento de referencia se utilizó el programa PhyML (Guindon et al., 2010), el cual emplea un algoritmo de máxima verosimilitud. El modelo evolutivo para este set es el determinado previamente para el alineamiento de referencia.

Por otro lado, en el caso de IBV, y con el objetivo de reducir costos computacionales, la reconstrucción filogenética se realizó utilizando el programa FastTree (Price et al., 2009). El modelo evolutivo utilizado fue el determinado previamente para el alineamiento de referencia.

Para computar los valores de soporte estadístico, se usó el test de Shimodira-Hasegawa en ambos casos (Anisimova & Gascuel, 2006).

### **Visualización de árboles filogenéticos**

Luego de realizada la inferencia filogenética, se utilizó una plataforma de análisis y visualización de árboles (filogenéticos en su aplicación más conocida, pero de propósito general) basada en el lenguaje de programación Python. Esta plataforma se denomina ETE ToolKit (Huerta-Cepas, Serra, & Bork, 2016) y es también de libre acceso.

### **Análisis de marcadores moleculares aminoacídicos**

En cuanto a los pasos a seguir para el análisis de marcadores en una secuencia problema, se considera el alineamiento de referencia extendido (esto es, con la adición de la nueva secuencia) y se procede a leer la secuencia problema en el marco de lectura correcto. Luego, se traduce la secuencia, anotando los aminoácidos presentes en las posiciones correspondientes a los marcadores moleculares aminoacídicos, a partir de los cuales se brinda información de interés con respecto a la secuencia analizada. Adicionalmente, se desarrolló un algoritmo que pondera la presencia o ausencia de los marcadores moleculares más frecuentes, a partir del cual se puede establecer a qué cepa o linaje es más similar la secuencia analizada teniendo en cuenta únicamente los marcadores aminoacídicos que contiene. Para cada posición analizada, se considera el porcentaje de secuencias de cada linaje que presenta el aminoácido de la secuencia problema; los linajes con alto porcentaje son más valorados, y terminarán estableciendo mayor semejanza con la secuencia analizada.

### **Desarrollo de la aplicación web**

Finalmente, se realizaron diagramas de flujo de la aplicación a construir. Esta tarea consistió en la investigación y posterior implementación de programas que reunieran las herramientas previamente utilizadas, y su traslado a un ambiente web. De esta forma, se desarrolló el cliente web (aplicación que corre sobre los navegadores de internet) y el servidor web donde se ejecutan las herramientas bioinformáticas, testeadas previamente en ambientes de escritorio. Las plataformas de diseño de aplicaciones, las herramientas bioinformáticas, y los lenguajes de programación utilizados, se detallan en el Anexo I.

## Resultados

### Insumos previos al análisis de una secuencia problema

#### Alineamientos de referencia

En el caso de IBDV, se generó un alineamiento formado por 82 secuencias de 369 nucleótidos de longitud, correspondientes a la región hipervariable del gen VP2, entre la posición 794 y 1163 (numeración en función de la secuencia AF240686). En este alineamiento se incluyeron entre 9 y 20 secuencias representantes de cada linaje, según se detalla a continuación: 9 cvIBDV, 13 atIBDV, 18 vaIBDV, 20 vvIBDV y 20 dIBDV (Tabla 1, Anexo II). Las 82 secuencias se completaron incluyendo 2 secuencias correspondientes a IBDVs del serotipo 2, las cuales funcionan como grupo externo.

En el caso de IBV, se generó un alineamiento con 99 secuencias completas del gen de la proteína S1, con una longitud de 1758 nucleótidos. Cada genotipo está representado por entre 2 y 5 secuencias de referencia (Tabla 2, Anexo II).

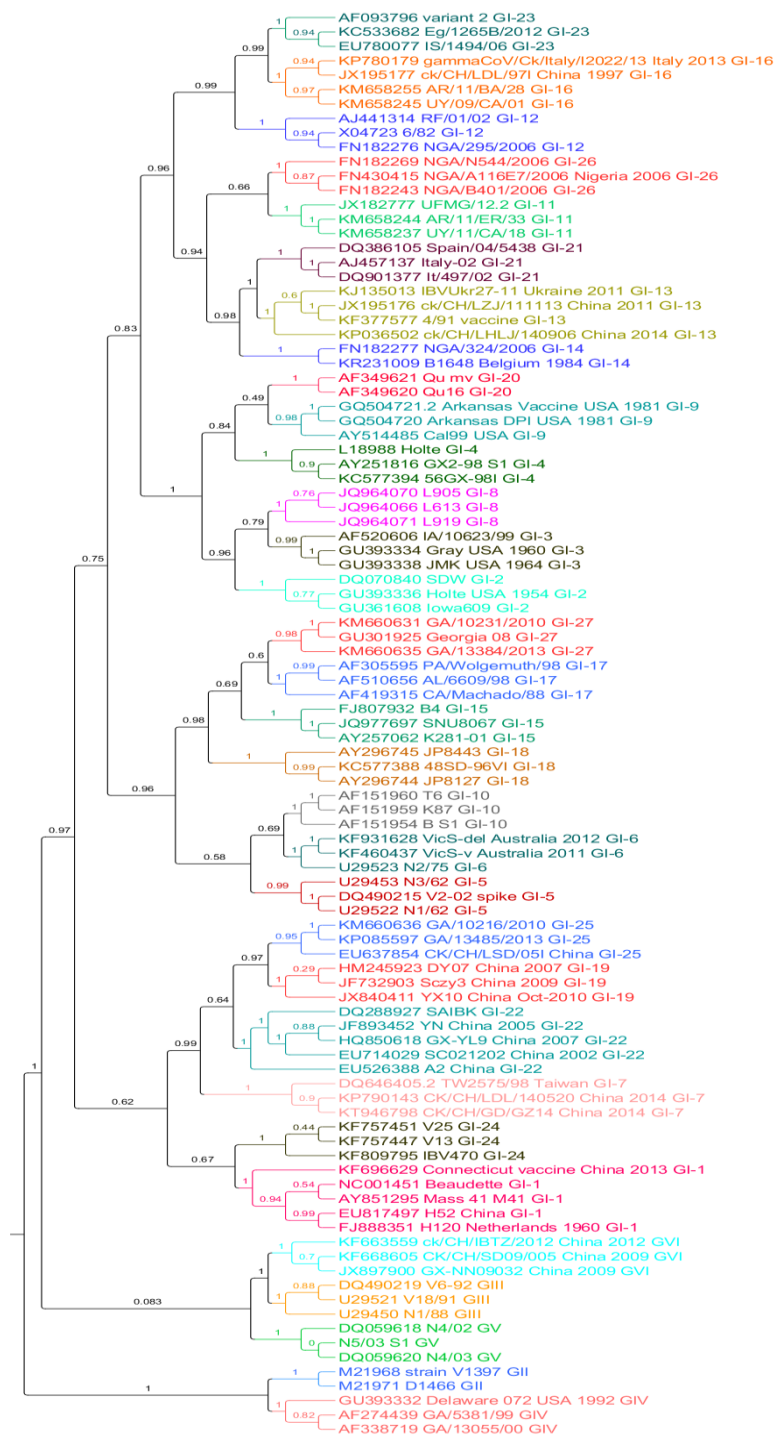
Ambos alineamientos de referencia se almacenaron en el sistema en formato FASTA, como información de solo lectura; cada vez que en la aplicación se solicitan análisis de una nueva secuencia, se genera una copia de este archivo, y se le añade la secuencia del usuario.

#### Reconstrucciones filogenéticas

Para IBDV se lograron obtener los 5 linajes previamente establecidos, con altos valores de apoyo estadístico en los nodos de cada linaje (Figura 7). Con respecto a IBV, también se pudo recuperar la estructura filogenética propuesta por Valastro et al. , en la que se tienen 32 linajes, agrupados en 6 genotipos, con valores de apoyo estadístico igual o muy cercano a 1 en todos los nodos comunes a cada genotipo (Figura 8).



**Figura 7** - Árbol filogenético de máxima verosimilitud de IBDV, implementado en el programa PhyML. Se incluyen 82 secuencias de 369 nucleótidos de longitud, correspondientes a la región hipervariable del gen VP2.



**Figura 8** – Árbol filogenético de máxima verosimilitud aproximada de IBV, implementado en el programa FastTree. Se incluyen 99 secuencias de IBV correspondientes al gen completo de la proteína S1.

### Análisis de marcadores moleculares

El análisis de los marcadores moleculares de los linajes de IBDV involucró la recolección de 2472 secuencias de la región hipervariable del gen VP2. Los análisis filogenéticos realizados permitieron determinar que del total de las secuencias, 1381 correspondieron a vvIBDV, 169 a cvIBDV, 192 a atIBDV, 604 a vaIBDV, y 126 a dIBDV. En la Tabla 2 se muestra el análisis de los aminoácidos encontrados en cada una de las 14 posiciones evaluadas según el linaje.

La posición 222 juega un papel importante en la antigenicidad, ya que cambios de aminoácidos alteran la capacidad de unión de diferentes anticuerpos monoclonales (Etteradossi, Arnauld, Toquin, & Rivallan, 1998; Letzel et al., 2007). Una Alanina en esta posición es característica de las cepas vvIBDV, una Treonina es característica de las cepas vaIBDV, una Prolina es compartida por las cepas cvIBDV y atIBDV, y una Serina está presente en las cepas dIBDV (Hernández et al., 2015; Xia et al., 2008). La presencia de los marcadores I242, I256 e I294 es característica de las cepas vvIBDV, mientras que V242, V256 y L294 se encuentran en el resto de las cepas, con la excepción de las cvIBDV, quienes pueden tener tanto I como V en la posición 242 (Rudd et al., 2002; Xia et al., 2008). Por otro lado, la presencia de T272, P289, I/V290 y F296 es característica de las cepas dIBDV, mientras que en el resto de las cepas se encuentra I272, L289, M290 e I296 (Hernández et al., 2015; Tomás et al., 2017). Las cepas atIBDV presentan a H253, N279, T284 y R330 como marcadores fuertes, mientras que las cepas vaIBDV se caracterizan por tener los marcadores K249 y S254 (Xia et al., 2008). La presencia de R249, H253, V256, N279 y T284 está asociada con la capacidad de un virus para replicar en cultivos celulares, atenuación y reducción de patogenicidad *in vivo* (Ben Abdeljelil et al., 2014; Noor et al., 2014; Van Loon et al., 2002). En general, se cumple que para cualquier linaje, el aminoácido mayormente representado en determinada posición, aparece en al menos un 95 % de las secuencias de dicho linaje; en menor grado se ven secuencias con aminoácidos que aparecen entre un 65 % y un 95 % de las secuencias analizadas.

**Tabla 2 - Resultado del análisis de marcadores moleculares aminoacídicos de IBDV. Se detalla para cada posición el tipo y cantidad de aminoácidos encontrados en función del linaje.**

	Very Virulent			Classic Virulent			Classic Attenuated			Variant			Distinct		
	AA	N°	%	AA	N°	%	AA	N°	%	AA	N°	%	AA	N°	%
222	A	1327	96.1	P	122	79.74	P	179	100	T	471	77.98	S	108	99.08
	S	28	2.02	L	16	10.46				A	63	10.43	T	1	0.92
	P	11	0.8	S	11	7.19				Q	31	5.13			
	T	10	0.72	T	4	2.61				P	23	3.81			
	Q	4	0.29							S	16	2.65			
	V	1	0.07												
242	I	1373	99.42	V	125	73.96	V	190	98.96	V	582	96.36	V	126	100
	V	7	0.51	I	44	26.04	G	1	0.52	I	20	3.31			
	M	1	0.07				A	1	0.52	G	2	0.33			
249	Q	1357	98.26	Q	136	80.47	Q	166	86.46	K	595	98.51	Q	126	100
	S	21	1.52	H	28	16.57	R	21	10.94	R	3	0.49			
	T	2	0.15	N	4	2.37	K	3	1.56	Q	2	0.33			
	K	1	0.07	Y	1	0.59	L	2	1.04	H	2	0.33			
										N	1	0.17			
253	Q	1375	99.56	Q	159	94.08	H	137	71.35	Q	573	94.87	Q	124	98.41
	S	4	0.29	H	10	5.92	Q	36	18.75	E	19	3.15	L	2	1.59
	L	2	0.15				N	13	6.77	H	10	1.66			
							L	4	2.08	R	2	0.33			
							D	1	0.52						
254	G	1095	79.29	G	167	98.82	G	186	96.88	S	398	65.89	N	72	57.14
	S	271	19.62	D	2	1.18	D	6	3.12	N	196	32.45	D	32	25.4
	D	14	1.01							K	5	0.83	S	21	16.67
	R	1	0.07							G	3	0.49	G	1	0.79
										T	1	0.17			
256	I	1371	99.28	V	130	76.92	V	186	96.88	V	603	99.83	V	125	99.21
	V	6	0.43	A	33	19.53	L	4	2.08	A	1	0.17	T	1	0.79
	T	4	0.29	I	5	2.96	I	1	0.52						
				T	1	0.59	A	1	0.52						
272	I	1348	97.61	I	169	100	I	192	100	I	593	98.18	T	124	98.41
	T	31	2.24							V	10	1.66	I	2	1.59
	V	2	0.15							T	1	0.17			
279	D	1335	96.67	D	147	86.98	N	184	95.83	N	595	98.51	N	126	100
	N	46	3.33	N	20	11.83	D	4	2.08	D	7	1.16			
				Y	1	0.59	H	2	1.04	S	2	0.33			
				H	1	0.59	S	1	0.52						
							Y	1	0.52						
284	A	1370	99.2	A	126	74.56	T	188	97.92	A	589	97.52	A	126	100
	T	10	0.72	T	42	24.85	A	4	2.08	T	12	1.99			
	S	1	0.07	P	1	0.59				V	2	0.33			
										G	1	0.17			
289	L	1380	99.93	L	169	100	L	192	100	L	602	99.67	P	126	100
	P	1	0.07							P	2	0.33			
290	M	1372	99.35	M	168	99.41	M	147	76.56	M	602	99.67	I	120	95.24
	V	4	0.29	V	1	0.59	L	34	17.71	V	1	0.17	V	6	4.76
	I	3	0.22				S	9	4.68	T	1	0.17			
	L	2	0.15				R	2	1.04						
294	I	1337	96.81	I	89	52.66	L	191	99.48	L	578	95.7	L	126	100
	V	33	2.39	L	80	47.34	I	1	0.52	I	22	3.64			
	L	10	0.72							P	3	0.49			
	T	1	0.07							F	1	0.17			
296	I	1378	99.78	I	167	98.82	I	191	99.48	I	602	99.67	P	125	99.21
	T	1	0.07	F	1	0.59	V	1	0.52	V	2	0.33	I	1	0.79
	F	1	0.07	V	1	0.59									
	V	1	0.07												
330	S	1375	99.57	S	155	91.72	R	183	95.31	S	603	99.83	S	77	100
	R	5	0.36	K	2	1.18	S	3	1.56	G	1	0.17			
	K	1	0.07	M	1	0.59	K	1	0.52						

### Estructura de la aplicación web y caso de uso principal

La diagramación del flujo de trabajo de la aplicación determinó que se empleara una serie de programas conectados entre sí mediante una estructura informática que permite de forma sencilla subir una secuencia al servidor y obtener resultados sobre su análisis en unos pocos segundos. Los componentes bioinformáticos utilizados a lo largo de la implementación web están implementados en módulos dentro de paquetes escritos en lenguaje Python, apoyados en la plataforma de herramientas bioinformáticas Biopython.

**Comentario [L2]:** Esto qué es? En las aplicaciones, a veces se documenta las posibles utilidades que tienen. El caso de uso principal en esta aplicación se refiere al análisis de una sec. problema.

Una descripción del resultado final de la estructura, y de cómo se relacionan los módulos dentro de los paquetes, puede leerse en el Anexo I, sección Estructura Bioinformática.

El producto final de este trabajo consiste en una aplicación web accesible desde internet, a través de la url <http://gevolutiva.fcien.edu.uy/>. Básicamente, el flujo de su caso de uso principal es como sigue:

1. La pantalla inicial de la web muestra varias pestañas, entre las cuales está la encargada del análisis de una secuencia problema. Se muestra un formulario, donde se permite al usuario insertar un texto o subir un archivo, correspondiente a la secuencia del virus que desea analizar, en formato FASTA (Figura 9). Se selecciona el tipo de virus (IBDV o IBV).
2. Se realizan validaciones de secuencia y de formato.
3. Se envía la información al servidor, donde se vuelven a realizar validaciones.
4. Se crea un valor numérico único que permitirá identificar inequívocamente el análisis del usuario.
5. Se procede a correr un programa de alineamiento múltiple de secuencias, alineando la secuencia ingresada contra el alineamiento de referencia.
6. Se corre un algoritmo de inferencia filogenética, que dependerá del tipo de virus a analizar.
7. La salida del paso anterior es un archivo representando las relaciones filogenéticas entre las secuencias, en formato Newick. Este archivo se utiliza como input para la visualización del árbol filogenético.
8. Finalizado el paso anterior, se realiza una serie de análisis comparativos de marcadores aminoacídicos.
9. Se despliegan los resultados en la misma pantalla en la que se inició el análisis. Estos resultados comprenden: información al respecto de la cepa según haya sido clasificada mediante análisis filogenéticos; una representación visual del árbol filogenético inferido, resaltando la secuencia problema; información referida al análisis de marcadores. Las figuras 10 a 17 muestran los resultados de analizar secuencias obtenidas del GenBank: para IBDV corresponden a las secuencias con número de acceso en AJ878638 y AF533673; para IBV, el número de acceso es AY851295.



VirusAviáres    Análisis    Información    Publicaciones    Contacto

### Cargar secuencias en formato FASTA

**Virus de Gumboro:** secuencia correspondiente a la región hipervariable de VP2. Largo mínimo de 300 nucleótidos.  
**Virus de la Bronquitis infecciosa:** secuencia correspondiente a la región codificante de S1. Largo mínimo de 1500 nucleótidos.

[Seleccionar archivo](#) No se eligió archivo

[Eliminar archivos](#)

**Secuencia FASTA:**

**Virus:**

[Analizar](#)

**Figura 9** – Formulario de análisis de la aplicación.

### RESULTADOS DEL ANÁLISIS

Información de la cepa    [Filogenias](#)    [Marcadores](#)

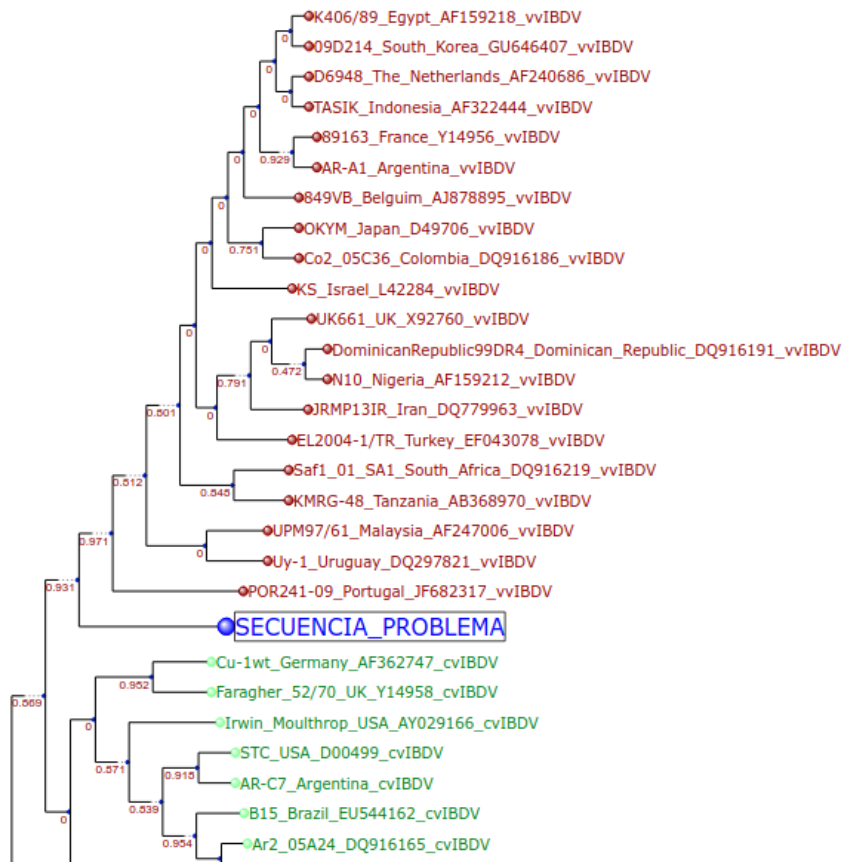
A la fecha, se conocen 5 linajes de las cepas del virus de Gumboro. La cepa analizada se agrupa dentro del cluster de las **Hipervirulentas**.

**Breve descripción:**  
 En este linaje se encuentran aquellas cepas que provocan los signos clínicos más evidentes y altas tasas de mortalidad de entre 50-100%. Luego de su surgimiento a finales de los años 1980s se dispersaron rápidamente por todos los continentes. Su control se basa en el uso de vacunas "calientes" derivadas de cepas clásicas virulentas, existiendo también vacunas desarrolladas a partir de propias cepas hipervirulentas, como la cepa vacunal mb derivada de la cepa de campo ks.

Sección Genética Evolutiva. Facultad de Ciencias. Universidad de la República    Responsable: Ruben Pérez  
 Contacto: [rperez@fcien.edu.uy](mailto:rperez@fcien.edu.uy)

**Figura 10** – Sección con información sobre la cepa analizada, correspondiente a la secuencia AJ878638

## Inferencia filogenética



**Figura 11** - Sección de los resultados luego del análisis filogenético, mostrando la secuencia problema resaltada (AJ878638).

Se realizó un análisis de los marcadores moleculares presentes en la secuencia problema, y los resultados fueron comparados con información obtenida de un análisis exhaustivo de las secuencias de Gumboro presentes en Genbank.

De esta forma, se intenta determinar la similitud de la cepa problema, tomando en cuenta el porcentaje de aparición de ciertos marcadores aminoacídicos, en determinadas posiciones. En particular, resultan interesantes las posiciones 253 y 284; los marcadores H253 y T284 han sido asociados con cepas atenuadas y/o adaptadas a cultivo celular, mientras que los aminoácidos Q253 y A284 se asocian con las cepas de campo.

**Resultados:**

La cepa problema presenta mayor similitud con las **Hipervirulentas**

Marcadores aminoacídicos														
Cepa/Posición	222	242	249	253	254	256	272	279	284	289	290	294	296	330
vvIBDV	A	I	V	Q	G	I	I	D	A	L	M	I	I	S
cvIBDV	P	V/I	Q	Q	G	V	I	D	A/T	L	M	M	I	S
atIBDV	P	V	Q	H	G	V	I	N	T	L	M	I/L	I	R
vaIBDV	T	V	K	Q	S/N	V	I	N	A	L	M	L	I	S
dlIBDV	S	V	Q	Q	N/D	V	T	N	A	P	I	L	P	S
SECUENCIA PROBLEMA	Q	I	Q	Q	G	I	I	D	A	L	M	L	I	S

Se indican los marcadores moleculares más representativos de cada linaje, resaltando los aminoácidos presentes en la cepa problema, y las posiciones asociadas a cepas atenuadas

**Figura 12** – Se muestran los marcadores moleculares aminoacídicos presentes en cada cepa, indicando la posición correspondiente. Las posiciones 253 y 284 se resaltan para indicar posible asociación con cepas atenuadas. Análisis correspondiente a la secuencia AJ878638

**RESULTADOS DEL ANÁLISIS**

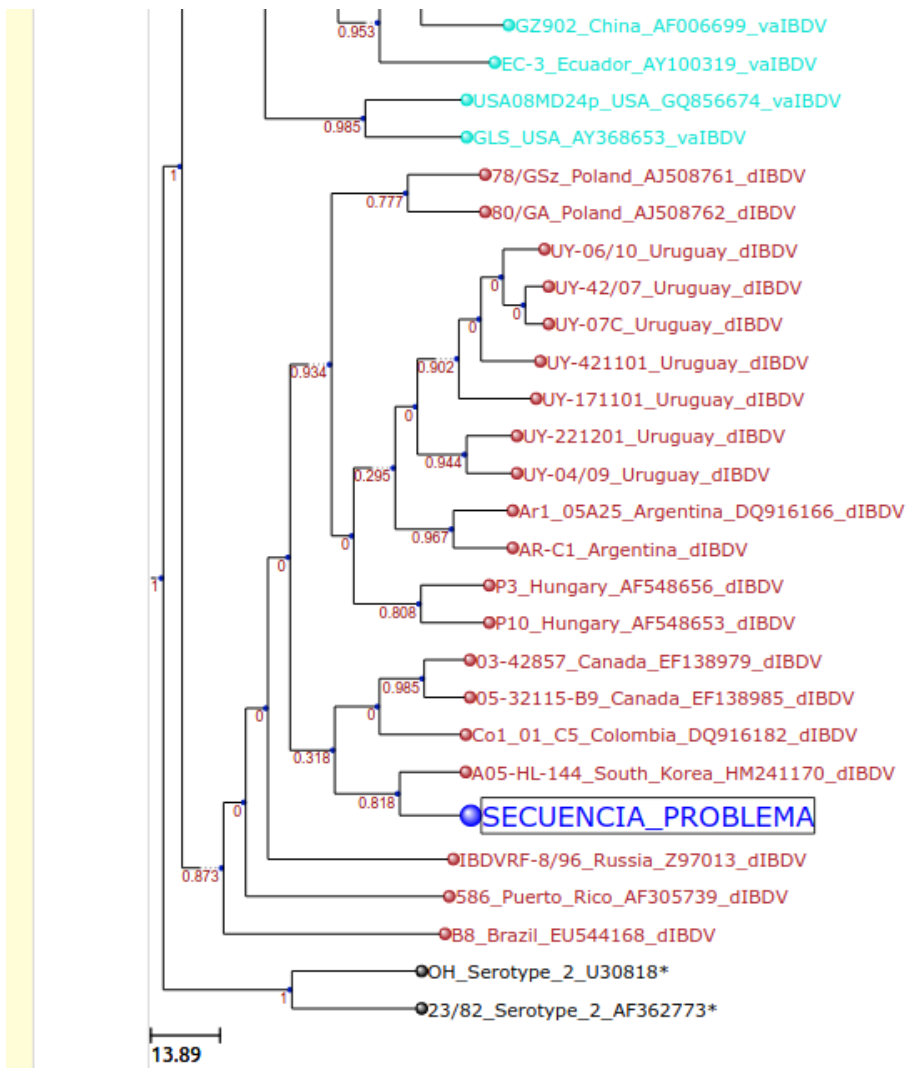
Información de la cepa Filogenias Marcadores

A la fecha, se conocen 5 linajes de las cepas del virus de Gumboro. La cepa analizada se agrupa dentro del cluster de las **Distinct**.

**Breve descripción:**  
 Este linaje está formado por cepas que se caracterizan por ser subclínicas, generando atrofia de la bursa de Fabricius e inmunosupresión pero sin causar mortalidad. Las cepas más antiguas son europeas, pero existen aislados en Asia, África y América. Los reportes más recientes indican que hay una circulación activa e estos virus principalmente en Argentina, Brasil, Canadá y Uruguay. Aún no se conoce la relación antigénica de estos virus con el resto de los linajes.

Sección Genética Evolutiva. Facultad de Ciencias. Universidad de la República Responsable: Ruben Pérez  
Contacto: rperez@fcien.edu.uy

**Figura 13** – Sección con información sobre la cepa analizada, correspondiente a la secuencia AF533673



**Figura 14** - Sección de los resultados luego del análisis filogenético, mostrando la secuencia problema resaltada (AF533673).

Información de la cepa Filogenias Marcadores

Se realizó un análisis de los marcadores moleculares presentes en la secuencia problema, y los resultados fueron comparados con información obtenida de un análisis exhaustivo de las secuencias de Gumboro presentes en Genbank.

De esta forma, se intenta determinar la similitud de la cepa problema, tomando en cuenta el porcentaje de aparición de ciertos marcadores aminoacídicos, en determinadas posiciones. En particular, resultan interesantes las posiciones 253 y 284; los marcadores H253 y T284 han sido asociados con cepas atenuadas y/o adaptadas a cultivo celular, mientras que los aminoácidos Q253 y A284 se asocian con las cepas de campo.

**Resultados:**

La cepa problema presenta mayor similitud con las **Distinct**

Marcadores aminoacídicos														
Cepa/Posición	222	242	249	253	254	256	272	279	284	289	290	294	296	330
vvlBDV	A	I	V	Q	G	I	I	D	A	L	M	I	I	S
cviBDV	P	V/I	Q	Q	G	V	I	D	A/T	L	M	M	I	S
atIBDV	P	V	Q	H	G	V	I	N	T	L	M	I/L	I	R
valBDV	T	V	K	Q	S/N	V	I	N	A	L	M	L	I	S
dIBDV	S	V	Q	Q	N/D	V	T	N	A	P	I	L	P	S
SECUENCIA PROBLEMA	S	V	Q	Q	S	V	T	N	A	P	I	L	F	S

Se indican los marcadores moleculares más representativos de cada linaje, resaltando los aminoácidos presentes en la cepa problema, y las posiciones asociadas a cepas atenuadas

**Figura 15** – Se muestran los marcadores moleculares aminoacídicos presentes en cada cepa, indicando la posición correspondiente. Las posiciones 253 y 284 se resaltan para indicar posible asociación con cepas atenuadas. Análisis correspondiente a la secuencia AF533673

**RESULTADOS DEL ANÁLISIS**

Información de la cepa Filogenias

**Genotipo al que pertenece la cepa analizada:** GI-1

**Características de las cepas que lo componen:** A este linaje pertenece el primer serotipo identificado. Es al día de hoy uno de los mejor estudiados y más ampliamente distribuido, probablemente debido al extenso uso de vacunas derivadas de una de sus cepas.

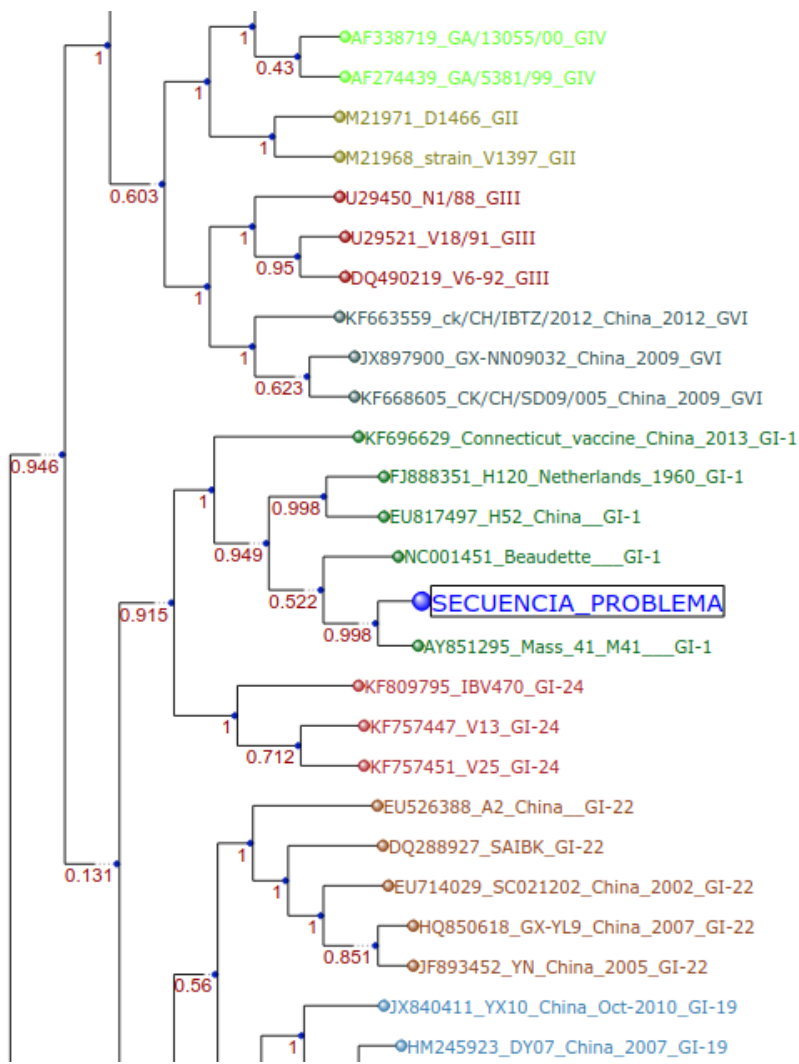
**Cepa prototipo:** Beaudette

**Número de acceso en GenBank:** M95169

Sección Genética Evolutiva. Facultad de Ciencias. Universidad de la República

Responsable: Ruben Pérez  
Contacto: rperez@fcien.edu.uy

**Figura 16** – Sección con información sobre la cepa analizada, correspondiente a la secuencia AY851295



**Figura 17** - Sección de los resultados luego del análisis filogenético, mostrando la secuencia problema resaltada (AY851295).

Todos los puntos que atraviesan el flujo mencionado están unidos entre sí por funciones y métodos auxiliares, necesarios para el correcto funcionamiento de la aplicación. Estas funciones incluyen, por ejemplo, la conversión entre formatos de alineamientos compatibles, limpieza de secuencias, manejo de posibles errores, flujo de información entre el servidor y el cliente, entre otras.

Además del flujo principal comentado previamente, la aplicación dispone de otras páginas con funcionalidades diversas:

- Descripción del proyecto en general, con comentarios breves sobre su propósito y sus expectativas (página inicial)
- Información y antecedentes sobre las cepas, linajes y genotipos (Figuras 18 y 19).
- Visualización de filogenias de referencia, establecidas para cada tipo de virus (Figuras 7 y 8).
- Información sobre marcadores moleculares (Tabla 2).
- Enlaces de interés (nuevas publicaciones en NCBI, herramientas útiles, links para descarga de publicaciones abiertas, etc.) (Figura 20).
- Formulario de contacto (Figura 21).

Cepas de Gumboro

Las cepas de Gumboro del serotipo 1 se agrupan en 5 linajes según el análisis filogenético de la región hipervariable de la proteína VP2: cepas clásicas virulentas, cepas clásicas atenuadas, cepas variantes, cepas hipervirulentas y cepas distinct. En Australia no circula ninguna de estas cepas, sino que circula un tipo de cepa particular restringida a ese país, las cuales forman un linaje aparte dentro del serotipo 1.

Linaje:

Nombre:  
**Clásicas virulentas**

Cepas de referencia:  
**STC (EEUU), Edgar (EEUU), F52/70 (Gran Bretaña), Cu-1wt (Alemania)**

Información general:  
**A este linaje pertenecen las primeras cepas del virus de Gumboro en ser reportadas. Se diseminaron rápidamente por varios continentes y hoy en día tienen distribución mundial. Estas cepas se caracterizan por provocar tasas de mortalidad de entre 10-50% en aves SPF. Muchas vacunas utilizadas mundialmente son realizadas a partir de cepas clásicas virulentas, como la cepa Lukert (deriva de la cepa Edgar) o la cepa Winterfield 2512.**

Sección Genética Evolutiva. Facultad de Ciencias. Universidad de la República Responsable: Ruben Pérez  
 Contacto: [rperez@fcien.edu.uy](mailto:rperez@fcien.edu.uy)

**Figura 18 - Sección de información sobre las cepas de Gumboro**

Cepas de Bronquitis Infecciosa

Las cepas de IBV se agrupan en 32 linajes, clasificados dentro de 6 genotipos. Puede seleccionar cualquiera de los linajes a continuación, y se mostrará información sobre su período de circulación, distribución geográfica, y una breve descripción sobre sus cepas.

Genotipo:

Período de circulación:  
**1975-2009**

Cepa prototipo:  
**UFMG/G**

País de origen:  
**Brazil**

Fecha de colecta:  
**1975**

Información general:  
**Estas cepas se han detectado solamente en América del Sur. Si bien la mayoría de ellas son de Brasil, han aparecido recientemente cepas en Argentina y Uruguay.**

Sección Genética Evolutiva. Facultad de Ciencias. Universidad de la República Responsable: Ruben Pérez  
 Contacto: [rperez@fcien.edu.uy](mailto:rperez@fcien.edu.uy)

**Figura 19 - Sección de información sobre las cepas de Bronquitis Infecciosa**



VirusAviáres    Análisis    Información ▾    Publicaciones    Contacto

En esta sección disponemos de links para descargar artículos publicados de libre acceso, enlaces a sitios vinculados a la temática, etc.

- *Development of an RT-qPCR assay for the specific detection of a distinct genetic lineage of the infectious bursal disease virus* [Tomás et al., 2017]
- *Whole-genome characterization of Uruguayan strains of avian infectious bronchitis virus reveals extensive recombination between the two major South American lineages.* [Marandino et al., 2017]
- *Development of RT-qPCR assays for the specific identification of two major genotypes of avian infectious bronchitis virus* [Marandino et al., 2016]
- *Genome Sequence of a Distinct Infectious Bursal Disease Virus* [Tomás et al., 2015]
- *Genetic characterization of South American infectious bursal disease virus reveals the existence of a distinct worldwide-spread genetic lineage* [Hernández et al., 2015]
- *Genome Sequence of a Distinct Infectious Bursal Disease Virus* [Tomás et al., 2015]
- *Phylogenetic analysis of avian infectious bronchitis virus in South America* [Marandino et al., 2015]
- *Development and validation of a TaqMan-MGB real-time RT-PCR assay for simultaneous detection and characterization of infectious bursal disease virus* [Tomás et al., 2012]
- *Novel Multiplex RT-PCR/RFLP Diagnostic Test to Differentiate Low- from High- Pathogenic Strains and to Detect Reassortant Infectious Bursal Disease Virus* [Hernández et al., 2011]
- *Sequence variability and evolution of the terminal overlapping VP5 gene of the infectious bursal disease virus* [Hernández et al., 2010]
- *Detection of very virulent strains of infectious bursal disease virus (vvIBDV) in commercial broilers from Uruguay* [Hernández et al., 2006]

Sección Genética Evolutiva. Facultad de Ciencias. Universidad de la República      Responsable: Ruben Pérez  
 Contacto: rperez@fcien.edu.uy

**Figura 20** - Sección de la página donde se muestran links a publicaciones de la Sección Genética Evolutiva

Contacto

**Nombre**

**Email**

**Asunto**

**Mensaje**

**Figura 21** - Formulario de contacto

## Discusión

### Importancia de la genotipificación

La industria avícola de todo el mundo debe enfrentar constantemente la ocurrencia de enfermedades que castigan duramente la producción, repercutiendo negativamente en los aspectos económicos y sociales de las regiones afectadas. Dos de los patógenos más problemáticos en este sentido son el virus de Gumboro y el virus de la Bronquitis Infecciosa, de los cuales hay reportes que estiman como cuantiosas las pérdidas económicas asociadas a ambos virus (Colvero, Villarreal, Torres, & Brando, 2015; Zachar et al., 2016). Ambos patógenos son enfrentados activamente a través del empleo de medidas biosanitarias y aplicación de vacunas, pero sin embargo, siguen siendo dos de los principales causantes de problemas sanitarios a nivel mundial. Para poder ejecutar planes de control adecuados, es fundamental conocer qué variante viral es la que está circulando en cada región. Con esa información se pueden ajustar las medidas biosanitarias y los esquemas de vacunación a las realidades de cada región, con lo cual se da un paso importante hacia el control más efectivo del patógeno en cuestión. Intentando colaborar con este aspecto, en el presente trabajo se desarrolló una aplicación web de fácil utilización y acceso, que permite analizar secuencias de los virus de Gumboro y Bronquitis Infecciosa, obteniendo como resultado la genotipificación de la cepa e información adicional asociada a la misma.

### Secuencias de referencia

Parte fundamental del correcto funcionamiento de la aplicación está basada en el grupo de secuencias incluidas en los alineamientos de referencia. Estos cumplen con los requerimientos necesarios para realizar los análisis: a) incluyen un número adecuado de secuencias de cada uno de los linajes/genotipos previamente descritos para cada virus; b) las secuencias incluidas pertenecen a aislamientos de referencia de cada uno de los linajes/genotipos, los cuales en su mayoría han sido muy utilizados en trabajos previos y son ampliamente conocidos; y c) están basados en regiones genómicas muy usadas, cuya utilidad en la genotipificación fue demostrada previamente, como es el caso de la región hipervariable de la proteína VP2 y el gen S1 (Hernández et al., 2015; C Le Nouen, Rivallan, Toquin, & Eterradosi, 2005; Valastro et al., 2016). Estos tres puntos, principalmente el primero y el segundo, no se cumplen en todos los trabajos donde se

realiza la genotipificación de IBV e IBDV, lo cual lleva a que un aislado sea clasificado erróneamente dentro de un linaje/genotipo previamente descrito, o peor aún, a que los autores propongan la existencia de nuevos linajes/genotipos. En el presente trabajo, estos puntos se tuvieron muy en cuenta, utilizando secuencias de referencia que probaron ser adecuadas para la genotipificación de ambos virus.

### **Alineamientos e inferencias filogenéticas**

Los métodos de alineamiento comúnmente utilizados son muy variados, y no siempre dan los mismos resultados. Teniendo esto en cuenta, en este trabajo se utilizó el programa MAFFT para realizar los alineamientos de referencia, el cual ha sido evaluado como uno de los más precisos (Pais, Ruy, Oliveira, & Coimbra, 2014). Si bien este programa es un poco más costoso en términos computacionales y puede ser más lento que otros similares, no representa un inconveniente en nuestro trabajo, principalmente por dos motivos: en primer lugar, las dimensiones de nuestros alineamientos no implican cambios notorios en la eficiencia de estos programas, y segundo, solamente se utiliza este programa en la etapa previa al análisis de una secuencia problema. Para alinear una secuencia problema frente al alineamiento de referencia se utiliza el programa Clustal Omega. Este programa se destaca por su sencillez en cuanto a la configuración de opciones relevantes en nuestro análisis, como la posibilidad de mantener intactas las columnas del alineamiento (se lo considera un perfil), o su fácil integración con la plataforma Biopython. A través del programa jModelTest se determinaron los modelos evolutivos que mejor se adaptan a los alineamientos, lo cual es fundamental para obtener árboles con la topología que mejor represente las relaciones de parentesco entre las secuencias estudiadas (Posada & Crandall, 2001). Finalmente, las inferencias filogenéticas se realizaron utilizando algoritmos de máxima verosimilitud, los cuales son hoy en día considerados los más adecuados para inferir filogenias a partir de datos moleculares (ADN). Estos métodos requieren altas capacidades computacionales, por lo tanto, trabajan analizando solamente un conjunto de topologías de árboles diferentes (nunca se analizan todas; la cantidad de árboles posibles crece exponencialmente con el número y el largo de las secuencias a estudiar), y comparan las probabilidades de producir los alineamientos observados. En general, se han desempeñado con mejores resultados que otros métodos muy utilizados pero que han ido perdiendo popularidad, como el método basado en distancias *Neighbor Joining* (Kuhner & Felsenstein, 1994; Schmidt & von Haeseler, 2009. (Van de Peer, 2009). En

el caso de IBV, se realizaron pruebas que determinaron la validez (comparando la similitud de los resultados obtenidos) del programa Fast Tree, el cual utiliza un método de máxima verosimilitud aproximada.

### **Marcadores moleculares aminoacídicos**

El análisis de marcadores moleculares aminoacídicos de IBDV que incluye la aplicación permite a los usuarios obtener información adicional a la genotipificación. Por un lado, y a través del análisis realizado por un algoritmo diseñado en este trabajo, la aplicación es capaz de determinar a qué linaje es más similar una secuencia problema en base a los aminoácidos contenidos en las posiciones seleccionadas como de interés. Si bien el método de genotipificación más adecuado y que debería considerarse primordial es el filogenético, este es un método secundario incluido en la aplicación que sirve como soporte al primero y también para detectar secuencias con características raras. Para poder ponderar el peso de cada uno de los marcadores, se generó una base de datos sin precedentes incluyendo absolutamente todas las secuencias de la región hipervariable de VP2 incluidas en la base de datos del GenBank (n = 2472), a partir de la cual se pudo generar una tabla con datos objetivos acerca de la robustez de cada marcador (Tabla 2). Por otro lado, la aplicación informa al usuario acerca de ciertas características que tiene su secuencia y que pueden ser de interés, como por ejemplo qué aminoácido tiene en la posición 222, importante desde un punto de vista antigénico, o qué aminoácido tiene en las posiciones 253 y 284, muy relevantes desde el punto de vista patogénico. Cada una de las 14 posiciones aminoacídicas incluidas en el análisis brinda información importante de conocer con respecto a un virus. De hecho, son comúnmente analizadas en trabajos de genotipificación de IBDV (Islam, Le, Rahman, & Islam, 2012; Mohamed, Elzanaty, Bakhit, & Safwat, 2014; Parede et al., 2003; Vera et al., 2015).

### **Características y ventajas de la aplicación VirusAviaries**

La genotipificación de un virus, en este caso IBV e IBDV, implica necesariamente contar con conocimientos previos que permitan realizarla de forma adecuada, desde qué secuencias utilizar como referencia o qué región genómica analizar, hasta cuáles programas bioinformáticos son los más adecuados. La aplicación VirusAviaries permite tanto a expertos como no-expertos realizar este tipo de análisis.

Si bien esta aplicación realiza la clasificación de los virus en base a una región genómica acotada, es necesario destacar que para lograr una caracterización genética completa debería analizarse el genoma completo. En el caso de IBDV, es conocida su capacidad para reordenar segmentos genómicos y generar cepas con características biológicas particulares (Gao, Wang, Gao, & Fu, 2007; Cyril Le Nouen et al., 2006; Sun et al., 2003; Wei et al., 2006, 2008), mientras que en el caso de IBV se sabe que la recombinación juega un rol fundamental en la generación de nuevas variantes (Cavanagh, Davis, & Cook, 1992; Lee & Jackwood, 2000). Sin embargo, también es cierto que la secuenciación genómica para la caracterización viral aún está lejos de utilizarse de forma rutinaria, y el abordaje a través del análisis parcial del genoma es aún más económico, sencillo y principalmente útil.

Existen otras herramientas web que permiten realizar análisis de secuencias y que son muy utilizadas por investigadores expertos y no-expertos de todo el mundo. Galaxy (“Galaxy,” n.d.), Viramp (un producto basado en Galaxy, adaptado para el estudio de ciertos virus particulares) (Wan, Renner, Albert, & Szpara, 2015), servidores online como phylogeny.fr (Dereeper et al., 2008), Phylemon 2.0 (Sánchez et al., 2011), POWER (Lin et al., 2005), o la plataforma del European Bioinformatics Institute <https://www.ebi.ac.uk/services>, son solo algunos ejemplos; estas plataformas funcionan proveyendo al usuario de un conjunto de herramientas bioinformáticas de acceso abierto, mediante las cuales se pueden crear *pipelines* con las funcionalidades deseadas. Hasta donde sabemos, VirusAviaries es la primera aplicación dedicada exclusivamente al análisis de dos de los patógenos más relevantes de la avicultura, característica que le confiere varias ventajas frente a servidores y plataformas web ampliamente conocidos y utilizados. Su mayor fortaleza radica en la incorporación del conocimiento adquirido previamente por un equipo de trabajo dedicado desde hace al menos una década al estudio de estos patógenos.

## **Perspectivas**

En cuanto a futuras modificaciones en el sistema, la aplicación fue diseñada considerando la posibilidad de añadir nuevas herramientas de análisis sobre los virus estudiados. Por ejemplo, para IBDV se podría incluir el análisis de regiones que abarquen el gen de la proteína VP1, con el objetivo de estudiar aspectos relacionados al reordenamiento de los segmentos A y B. Similarmente, se podría añadir el gen 1 y el

gen N de los virus de Bronquitis Infecciosa, y así estudiar fenómenos de recombinación. Por otro lado, estos análisis podrían adaptarse con relativa sencillez al estudio de nuevos patógenos. Además de estas modificaciones, se podrían implementar módulos para la recopilación automática de información referida a publicaciones sobre este tema, mediante scripts de búsqueda externos a nuestro servidor, desde donde se tomarían los links hacia enlaces de publicaciones recientes. En este sentido, el paquete biopython ofrece funcionalidades para conectarse a las bases de datos del NCBI y obtener información, formatearla, y exportarla hacia otros sistemas, las cuales podrían ser aprovechadas por nuestro sistema.

En paralelo a estos cambios, podrían implementarse scripts funcionales a tareas de uso cotidiano en nuestro laboratorio, como la posibilidad de formatear grandes cantidades de secuencias obtenidas en formato GenBank, para luego ser utilizadas en otros programas que admiten formatos estrictos, como BEAST.

Finalmente, consideramos la posibilidad de crear y mantener versiones del sitio web en inglés.

### **Agradecimientos**

A mi familia y a Analía, por la paciencia y el apoyo de siempre. A los compañeros del departamento de Genética Evolutiva. A mis amigos de toda la vida. A Diego, y especialmente a Matías, por todo lo que aprendí con ellos.

## Bibliografía

- Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology*, *55*(4), 539–552. <https://doi.org/10.1080/10635150600755453>
- Ben Abdeljelil, N., Khabouchi, N., Kassar, S., Miled, K., Boubaker, S., Ghram, A., & Mardassi, H. (2014). Simultaneous alteration of residues 279 and 284 of the VP2 major capsid protein of a very virulent Infectious Bursal Disease Virus (vvIBDV) strain did not lead to attenuation in chickens. *Virology Journal*, *11*, 199. <https://doi.org/10.1186/s12985-014-0199-7>
- Berg, T. P. (2000). Acute infectious bursal disease in poultry: a review. *Avian Pathology: Journal of the W.V.P.A.*, *29*(September 2013), 175–194. <https://doi.org/10.1080/03079450050045431>
- Berry, D. M., Cruickshank, J. G., Chu, H. P., & Wells, R. J. H. (1964). The structure of infectious bronchitis virus. *Virology*, *23*(3), 403–407. [https://doi.org/https://doi.org/10.1016/0042-6822\(64\)90263-6](https://doi.org/https://doi.org/10.1016/0042-6822(64)90263-6)
- Bing, G.-X., Liu, X., Pu, J., Liu, Q.-F., Wu, Q.-M., & Liu, J.-H. (2007). Different genotypes of nephropathogenic infectious bronchitis viruses co-circulating in chicken population in China. *Virus Genes*, *35*(2), 333–337. <https://doi.org/10.1007/s11262-007-0100-5>
- Cavanagh, D. (2003). Severe acute respiratory syndrome vaccine development: experiences of vaccination against avian infectious bronchitis coronavirus. *Avian Pathology*, *32*(6), 567–582. Retrieved from <http://proxy.timbo.org.uy:443/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=lbh&AN=20043011568&lang=es&site=eds-live>
- Cavanagh, D., Davis, P. J., & Cook, J. K. (1992). Infectious bronchitis virus: evidence for recombination within the Massachusetts serotype. *Avian Pathology: Journal of the W.V.P.A.*, *21*(3), 401–408. <https://doi.org/10.1080/03079459208418858>
- Colvero, L. P., Villarreal, L. Y. B., Torres, C. A., & Brando, P. E. (2015). Assessing the economic burden of avian infectious bronchitis on poultry farms in Brazil. *Revue Scientifique et Technique (International Office of Epizootics)*, *34*(3), 993–999.
- Cook, C. E., Bergman, M. T., Finn, R. D., Cochrane, G., Birney, E., & Apweiler, R. (2016). The European Bioinformatics Institute in 2016: Data growth and integration. *Nucleic Acids Research*, *44*(D1), D20–6. <https://doi.org/10.1093/nar/gkv1352>
- Crinion, R. A., Ball, R. A., & Hofstad, M. S. (1971). Pathogenesis of oviduct lesions in immature chickens following exposure to infectious bronchitis virus at one day old. *Avian Diseases*, *15*(1), 32–41.
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat Meth*, *9*(8), 772. Retrieved from <http://dx.doi.org/10.1038/nmeth.2109>
- de Wit, J. J., Nieuwenhuisen-van Wilgen, J., Hoogkamer, A., van de Sande, H., Zuidam, G. J., & Fabri, T. H. F. (2011). Induction of cystic oviducts and protection against early challenge with infectious bronchitis virus serotype D388 (genotype QX) by maternally derived antibodies and by early vaccination. *Avian Pathology*, *40*(5), 463–471. <https://doi.org/10.1080/03079457.2011.599060>

- Delgui, L., Gonzalez, D., & Rodriguez, J. F. (2009). Infectious bursal disease virus persistently infects bursal B-lymphoid DT40 cells. *The Journal of General Virology*, 90(Pt 5), 1148–1152. <https://doi.org/10.1099/vir.0.008870-0>
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., ... Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, 36(Web Server issue), 465–469. <https://doi.org/10.1093/nar/gkn180>
- Etteradossi, N., Arnauld, C., Tekaia, F., Toquin, D., Le Coq, H., Rivallan, G., ... Skinner, M. A. (1999). Antigenic and genetic relationships between European very virulent infectious bursal disease viruses and an early West African isolate. *Avian Pathology*, 28(1), 36–46. <https://doi.org/10.1080/03079459995028>
- Etteradossi, N., Arnauld, C., Toquin, D., & Rivallan, G. (1998). Critical amino acid changes in VP2 variable domain are associated with typical and atypical antigenicity in very virulent infectious bursal disease viruses. *Archives of Virology*, 143(8), 1627–1636. <https://doi.org/10.1007/s007050050404>
- Etteradossi, N., & Saif NM. (2008). Infectious bursal disease. *Diseases of Poultry*, 185–208.
- Fahey, K. J., Erny, K., & Crooks, J. (1989). A conformational immunogen on VP-2 of infectious bursal disease virus that induces virus-neutralizing antibodies that passively protect chickens. *The Journal of General Virology*, 70 ( Pt 6), 1473–1481. <https://doi.org/10.1099/0022-1317-70-6-1473>
- Galaxy. (n.d.). Retrieved from <https://galaxyproject.org/>
- Gao, H.-L., Wang, X.-M., Gao, Y.-L., & Fu, C.-Y. (2007). Direct Evidence of Reassortment and Mutant Spectrum Analysis of a Very Virulent Infectious Bursal Disease Virus. *Avian Diseases*, 51(4), 893–899. Retrieved from <http://www.jstor.org/stable/27565668>
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, 59(3), 307–321. <https://doi.org/10.1093/sysbio/syq010>
- Hernández, M., Tomás, G., Hernández, D., Villegas, P., Banda, A., Maya, L., ... Pérez, R. (2011). Novel Multiplex RT-PCR/RFLP Diagnostic Test to Differentiate Low- from High-Pathogenic Strains and to Detect Reassortant Infectious Bursal Disease Virus. *Avian Diseases*, 55(3), 368–74. <https://doi.org/10.1637/9672-013111-Reg.1>
- Hernández, M., Tomás, G., Marandino, A., Iraola, G., Maya, L., Mattion, N., ... Pérez, R. (2015). Genetic characterization of South American infectious bursal disease virus reveals the existence of a distinct worldwide-spread genetic lineage. *Avian Pathology: Journal of the W.V.P.A.*, 44(3), 212–21. <https://doi.org/10.1080/03079457.2015.1025696>
- Hirai, K., Funakoshi, T., Nakai, T., & Shimakura, S. (1981). Sequential changes in the number of surface immunoglobulin-bearing B lymphocytes in infectious bursal disease virus-infected chickens. *Avian Diseases*, 25(2), 484–496.
- Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution*. <https://doi.org/10.1093/molbev/msw046>
- Islam, M. T., Le, T. H., Rahman, M. M., & Islam, M. A. (2012). Molecular characterization of two Bangladeshi infectious bursal disease virus isolates using the hypervariable sequence of VP2 as a genetic marker. *Journal of Veterinary Science*, 13(4), 405–412. <https://doi.org/10.4142/jvs.2012.13.4.405>
- Ismail, N. M., Saif, Y. M., & Moorhead, P. D. (1988). Lack of pathogenicity of five



- serotype 2 infectious bursal disease viruses in chickens. *Avian Diseases*, 32(4), 757–759.
- Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059–3066. <https://doi.org/10.1093/nar/gkf436>
- Ladman, B. S., Loupos, A. B., & Gelb, J. J. (2006). Infectious bronchitis virus S1 gene sequence comparison is a better predictor of challenge of immunity in chickens than serotyping by virus neutralization. *Avian Pathology: Journal of the W.V.P.A.*, 35(2), 127–133. <https://doi.org/10.1080/03079450600597865>
- Lai, M. M., & Cavanagh, D. (1997). The molecular biology of coronaviruses. *Adv Virus Res*, 48, 1–100.
- Le Nouen, C., Rivallan, G., Toquin, D., Darlu, P., Morin, Y., Beven, V., ... Etteradossi, N. (2006). Very virulent infectious bursal disease virus: reduced pathogenicity in a rare natural segment-B-reassorted isolate. *The Journal of General Virology*, 87(Pt 1), 209–216. <https://doi.org/10.1099/vir.0.81184-0>
- Le Nouen, C., Rivallan, G., Toquin, D., & Etteradossi, N. (2005). Significance of the genetic relationships deduced from partial nucleotide sequencing of infectious bursal disease virus genome segments A or B. *Archives of Virology*, 150(2), 313–325. <https://doi.org/10.1007/s00705-004-0409-5>
- Lee, C. W., & Jackwood, M. W. (2000). Evidence of genetic diversity generated by recombination among avian coronavirus IBV. *Archives of Virology*, 145(10), 2135–2148.
- Lemey, P., Salemi, M., & Vandamme, A. (2009). *The phylogenetic handbook. Book*. <https://doi.org/10.1017/CBO9780511819049>
- Letzel, T., Coulibaly, F., Rey, F. A., Delmas, B., Jagt, E., van Loon, A. A. M. W., & Mundt, E. (2007). Molecular and structural bases for the antigenicity of VP2 of infectious bursal disease virus. *Journal of Virology*, 81(23), 12827–12835. <https://doi.org/10.1128/JVI.01501-07>
- Lin, C. Y., Lin, F. K., Lin, C. H., Lai, L. W., Hsu, H. J., Chen, S. H., & Hsiung, C. A. (2005). POWER: Phylogenetic Web Repeater - An integrated and user-optimized framework for biomolecular phylogenetic analysis. *Nucleic Acids Research*, 33(SUPPL. 2), 553–556. <https://doi.org/10.1093/nar/gki494>
- Liu, S., Zhang, X., Wang, Y., Li, C., Han, Z., Shao, Y., ... Kong, X. (2009). Molecular characterization and pathogenicity of infectious bronchitis coronaviruses: complicated evolution and epidemiology in china caused by cocirculation of multiple types of infectious bronchitis coronaviruses. *Intervirology*, 52(4), 223–234. <https://doi.org/10.1159/000227134>
- McFerran, J. B., McNulty, M. S., McKillop, E. R., Connor, T. J., McCracken, R. M., Collins, D. S., & Allan, G. M. (1980). Isolation and serological studies with infectious bursal disease viruses from fowl, turkeys and ducks: demonstration of a second serotype. *Avian Pathology: Journal of the W.V.P.A.*, 9(3), 395–404. <https://doi.org/10.1080/03079458008418423>
- Mohamed, M. A., Elzanaty, K. E. S., Bakhit, B. M., & Safwat, M. M. (2014). Genetic Characterization of Infectious Bursal Disease Viruses Associated with Gumboro Outbreaks in Commercial Broilers from Asyut Province, Egypt. *ISRN Veterinary Science*, 2014, 916412. <https://doi.org/10.1155/2014/916412>
- Muller, H., Mundt, E., Etteradossi, N., & Islam, M. R. (2012). Current status of vaccines against infectious bursal disease. *Avian Pathology: Journal of the W.V.P.A.*, 41(2), 133–139. <https://doi.org/10.1080/03079457.2012.661403>
- Muneer, M. A., Newman, J. A., Halvorson, D. A., Sivanandan, V., & Coon, C. N.

- (1987). Effects of avian infectious bronchitis virus (Arkansas strain) on vaccinated laying chickens. *Avian Diseases*, 31(4), 820–828.
- Noor, M., Mahmud, M. S., Ghose, P. R., Roy, U., Nooruzzaman, M., Chowdhury, E. H., ... Müller, H. (2014). Further evidence for the association of distinct amino acid residues with in vitro and in vivo growth of infectious bursal disease virus. *Archives of Virology*, 159(4), 701–709. <https://doi.org/10.1007/s00705-013-1885-2>
- Pais, F. S.-M., Ruy, P. de C., Oliveira, G., & Coimbra, R. S. (2014). Assessing the efficiency of multiple sequence alignment programs. *Algorithms for Molecular Biology: AMB*, 9, 4. <https://doi.org/10.1186/1748-7188-9-4>
- Parede, L. H., Sapats, S., Gould, G., Rudd, M., Lowther, S., & Ignjatovic, J. (2003). Characterization of infectious bursal disease virus isolates from Indonesia indicates the existence of very virulent strains with unique genetic changes. *Avian Pathology: Journal of the W.V.P.A.*, 32(5), 511–518. <https://doi.org/10.1080/0307945031000154116>
- Posada, D., & Crandall, K. A. (2001). Selecting the best-fit model of nucleotide substitution. *Systematic Biology*, 50(4), 580–601.
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution*, 26(7), 1641–1650. <https://doi.org/10.1093/molbev/msp077>
- Rudd, M. F., Heine, H. G., Sapats, S. I., Parede, L., & Ignjatovic, J. (2002). Characterisation of an Indonesian very virulent strain of infectious bursal disease virus. *Archives of Virology*, 147(7), 1303–1322. <https://doi.org/10.1007/s00705-002-0817-3>
- Sánchez, R., Serra, F., Tárraga, J., Medina, I., Carbonell, J., Pulido, L., ... Dopazo, H. (2011). Phylemon 2.0: A suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Research*, 39(SUPPL. 2), 470–474. <https://doi.org/10.1093/nar/gkr408>
- Schalk, A. ., & Hawn, M. C. (1931). An apparently new respiratory disease of baby chicks. *J. Am. Vet. Med. Assoc*, 78, 413–422.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7, 539. <https://doi.org/10.1038/msb.2011.75>
- Stern, D. F., & Sefton, B. M. (1982). Coronavirus proteins: biogenesis of avian infectious bronchitis virus virion proteins. *Journal of Virology*, 44(3), 794–803. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC256336/>
- Sun, J.-H., Lu, P., Yan, Y.-X., Hua, X.-G., Jiang, J., & Zhao, Y. (2003). Sequence and analysis of genomic segment A and B of very virulent infectious bursal disease virus isolated from China. *Journal of Veterinary Medicine. B, Infectious Diseases and Veterinary Public Health*, 50(3), 148–154.
- Tomas, G., Hernandez, M., Marandino, A., Panzera, Y., Maya, L., Hernandez, D., ... Perez, R. (2012). Development and validation of a TaqMan-MGB real-time RT-PCR assay for simultaneous detection and characterization of infectious bursal disease virus. *Journal of Virological Methods*, 185(1), 101–107. <https://doi.org/10.1016/j.jviromet.2012.06.012>
- Valastro, V., Holmes, E. C., Britton, P., Fusaro, A., Jackwood, M. W., Cattoli, G., & Monne, I. (2016). S1 gene-based phylogeny of infectious bronchitis virus: An attempt to harmonize virus classification. *Infection, Genetics and Evolution*, 39, 349–364. <https://doi.org/10.1016/j.meegid.2016.02.015>
- Van Loon, A. A. W. M., De Haas, N., & Mundt, E. (2002). Alteration of amino acids in

- VP2 of very virulent infectious bursal disease virus results in tissue culture adaptation and attenuation in chickens. *Journal of General Virology*, 83(1), 121–129. <https://doi.org/10.1099/0022-1317-83-1-121>
- Vera, F., Craig, M. I., Olivera, V., Rojas, F., Konig, G., Pereda, A., & Vagnozzi, A. (2015). Molecular characterization of infectious bursal disease virus (IBDV) isolated in Argentina indicates a regional lineage. *Archives of Virology*, 160(8), 1909–1921. <https://doi.org/10.1007/s00705-015-2449-4>
- Wan, Y., Renner, D. W., Albert, I., & Szpara, M. L. (2015). VirAmp: a galaxy-based viral genome assembly pipeline. *GigaScience*, 4, 19. <https://doi.org/10.1186/s13742-015-0060-y>
- Wei, Y., Li, J., Zheng, J., Xu, H., Li, L., & Yu, L. (2006). Genetic reassortment of infectious bursal disease virus in nature. *Biochemical and Biophysical Research Communications*, 350(2), 277–287. <https://doi.org/10.1016/j.bbrc.2006.09.040>
- Wei, Y., Yu, X., Zheng, J., Chu, W., Xu, H., Yu, X., & Yu, L. (2008). Reassortant infectious bursal disease virus isolated in China. *Virus Research*, 131(2), 279–282. <https://doi.org/10.1016/j.virusres.2007.08.013>
- Winterfield, R. W., & Hitchner, S. B. (1962). Etiology of an infectious nephritis-nephrosis syndrome of chickens. *American Journal of Veterinary Research*, 23, 1273–1279.
- Xia, R. X., Wang, H. Y., Huang, G. M., & Zhang, M. F. (2008). Sequence and phylogenetic analysis of a Chinese very virulent infectious bursal disease virus. *Archives of Virology*, 153(9), 1725–1729. <https://doi.org/10.1007/s00705-008-0140-8>
- Zachar, T., Popowich, S., Goodhope, B., Knezacek, T., Ojkic, D., Willson, P., ... Gomis, S. (2016). A 5-year study of the incidence and economic impact of variant infectious bursal disease viruses on broiler production in Saskatchewan, Canada. *Canadian Journal of Veterinary Research = Revue Canadienne de Recherche Veterinaire*, 80(4), 255–261.

## Anexo I

### Desarrollo de la aplicación web

La infraestructura de la aplicación se basa en un modelo Cliente - Servidor, el cual consta de 3 partes fundamentales:

- Un servidor web
- Clientes que acceden a la información alojada en el servidor (a través de navegadores, o *browsers*)
- Conexión entre el cliente y el servidor

En este modelo de tres capas, los navegadores solicitan información al servidor vía el protocolo HTTP; el servidor provee los archivos solicitados, los cuales son desplegados por el navegador (cliente) y permiten la interacción del usuario final con los datos.

Actualmente, una de las arquitecturas más utilizadas para el desarrollo de aplicaciones es un modelo basado en 3 capas:

1. Capa de presentación
2. Capa de lógica
3. Capa de datos

Estrictamente, el sitio web desarrollado consta de dos aplicaciones:

#### ***Cliente***

Esta es la aplicación de *front-end* que renderiza las vistas en HTML (capa de presentación) y es la encargada de la interacción con el usuario. Está implementada en AngularJS (“AngularJS,” 2016), toma librerías de Bootstrap (“Bootstrap,” n.d.) y utiliza el servidor ExpressJS (“ExpressJS,” n.d.) de NodeJS (“NodeJS,” n.d.). Para el diseño y desarrollo de esta aplicación se utilizó el IDE Atom (“Atom Editor,” n.d.)

AngularJS es una plataforma de uso libre y gratuito basado puramente en Javascript, desarrollado en 2009 y mantenido actualmente por Google. Propone una estructura fuertemente orientada al patrón de diseño conocido como Model-View-Controller; su principal cometido es la separación de la presentación, la lógica de la aplicación, y el manejo del modelo de datos en capas bien definidas y poco acopladas. Este patrón de diseño permite trabajar de forma más limpia, y poder mantener y modificar un componente particular del sistema sin afectar el resto, o con un impacto mínimo. En este sentido, mientras se mantenga una estructura acorde a esta filosofía, se puede añadir el

análisis de nuevos tipos de virus sin demasiados cambios. Una característica importante de AngularJS es que su diseño permite la vinculación de los datos de la aplicación con la presentación de una manera muy sencilla, a través de directivas propias de AngularJS que extienden las etiquetas convencionales de HTML. AngularJS es compatible con los navegadores más utilizados en la industria; el manejo del código Javascript está diseñado para adaptarse automáticamente a los navegadores actuales, ya sea Mozilla Firefox, Google Chrome, Edge, etc.

La estructura resultante se concentra dentro de un directorio *app*, junto con otros directorios accesorios que alojan librerías necesarias para el funcionamiento del servidor de *NodeJS*, más archivos de configuración con especificaciones sobre el arranque de la aplicación, puertos, variables, etcétera. Está desarrollada utilizando los lenguajes de programación Javascript, HTML y CSS.

Dentro de *app* tenemos una estructura bastante simple, tomada de lo que se conoce como un *seed* de AngularJS, con los siguientes directorios:

- *css: cascade style sheets*; archivos de descripción de la presentación de las vistas (archivos HTML). Permiten darles estética a las vistas, y determinan cómo se renderizan estos archivos
- *imgs*: imágenes que utiliza el sitio (fondos, logos, filogenias estáticas, etcétera)
- *lib*: librerías utilizadas por AngularJS
- *recursos*: archivos *json* con información estática, utilizados para cargar información de manera sencilla
- *views*: vistas (páginas en HTML)
- *js*: tal vez el directorio más importante. Contiene la estructura propiamente dicha de la aplicación cliente, siguiendo la filosofía de AngularJS. Dentro de esta carpeta tenemos los *controllers*, *factories*, *services*, y *directives*, más el archivo de configuración *app.js*. Los *controllers* son archivos que se encargan de manejar la lógica del modelo. Típicamente, tenemos un archivo para cada vista. El modelo de datos es llevado al cliente por medio de las *factories*, con lo que tenemos un patrón de diseño conocido como Model-View-Controller. En *app.js* se declara el módulo principal de toda la aplicación, y se determina qué controlador maneja qué vista, y las rutas de cada vista. Se importan algunas de

las librerías utilizadas, como ser la encargada de manejar los archivos con las secuencias subidas mediante el formulario principal.

### ***Servidor***

La capa de lógica del sistema corre del lado del servidor, y está desarrollada enteramente en python, mediante el uso del paquete bioinformático Biopython (Cock et al., 2009). Python es un lenguaje con las características apropiadas para el tipo de aplicación que nos ocupa. Es un lenguaje conciso, de fácil interpretación, multiparadigma (permite la programación orientada a objetos, así como la programación procedural) y multiplataforma. Tornado es un mini-framework liviano desarrollado en python, con la capacidad de funcionar asincrónicamente, compatible con la herramienta de *front-end* propuesta. Dado que el principal cometido de la aplicación involucra el manejo de información biológica, una de las opciones que aparecen como posibles soluciones es el paquete de aplicaciones Biopython, de libre uso y con una amplia gama de módulos para la implementación de programas bioinformáticos. Biopython es un paquete de desarrollo *open-source*, con una licencia muy poco restrictiva, y cuenta con una gran comunidad de desarrolladores que aportan nuevas implementaciones y conforman un recurso valioso en cuanto a la transferencia de conocimiento. Este paquete cuenta con módulos que permiten el manejo de secuencias de ácidos nucleicos, el alineamiento de las mismas, la inferencia y la visualización de árboles filogenéticos, la conexión con bases de datos, y la manipulación de información de secuencias proteicas, entre muchas otras herramientas.

El proyecto está conformado por paquetes (similares a directorios), los cuales agrupan módulos (archivos .py con diversas funcionalidades). Así, tenemos paquetes encargados del manejo de secuencias, alineamientos, filogenias, y análisis de marcadores. Por otro lado, se agrupan módulos con funcionalidades utilitarias: manejo de constantes, errores, validaciones, limpieza de archivos, y funciones utilizadas por el resto de los paquetes. Por último, se tiene un paquete con módulos encargados de recibir y derivar las peticiones del cliente.

El servidor y *framework* donde corren estos programas es Tornado (“Tornado,” 2016). El desarrollo de esta aplicación fue hecho sobre el IDE Eclipse (“Eclipse IDE,” n.d.).

## ***Hosting***

El sistema en su conjunto está levantado y accesible en internet de forma permanente (las 24 horas) mediante dos servicios: *supervisor* mantiene AngularJS corriendo y se inicia automáticamente ante eventuales apagones de la máquina, y *pm2*, encargado de levantar y monitorear Tornado. Se accede al sitio desde internet mediante un servidor HTTP denominado *nginx* (“Nginx,” n.d.).

Se configuraron repositorios donde se aloja el código fuente ([www.bitbucket.org](http://www.bitbucket.org), (“Bitbucket,” n.d.) y se utiliza el sistema de versionado *git* (<https://git-scm.com/>, (“Github,” n.d.)) para trabajar de forma limpia y ordenada en ambientes de desarrollo y producción.

Todos los componentes del sistema se encuentran completamente alojados en un servidor virtual Linux de la Facultad de Ciencias, Universidad de la República.

## **Estructura bioinformática**

La aplicación servidor gira en torno a Tornado, por lo tanto, esta estructura contiene y se asocia con los paquetes y módulos encargados de procesar los análisis bioinformáticos. A continuación, una breve explicación de cómo están empaquetadas las diferentes funcionalidades del sistema.

### **Módulo de secuencias**

Se encarga de tomar los datos recibidos desde el cliente, y transformarlos en objetos *SeqSecord* de *Biopython*. Realiza algunas validaciones propias de una secuencia de nucleótidos, y ejecuta ciertas restricciones con el objetivo de estandarizar el método de análisis y la interpretación de los resultados: deben ser secuencias de ADN; se aceptan solamente secuencias en formato FASTA; se restringe la longitud de las secuencias ingresadas. De estas restricciones, probablemente la más importante sea la de controlar el largo mínimo que deben tener las secuencias ingresadas para ser consideradas válidas. En el caso de utilizar secuencias significativamente más largas que nuestro alineamiento de referencia, los programas de MSA y de inferencia filogenética se encargan de quitar información sobrante. Por otro lado, se apoya en el módulo utilitario para validaciones más generales.

## Módulo de alineamientos

Este módulo tiene funciones para trabajar con objetos de tipo `AlignIO.Alignment`, y maneja los *wrappers* para los programas de alineamiento múltiple, como `ClustalOmegaCommandLine`, o `MuscleCommandLine`.

## Módulo de inferencias filogenéticas

Aquí tenemos todo el manejo de los algoritmos de inferencia filogenética; funciones que reciben alineamientos y devuelven estructuras en formato Newick que representan las relaciones filogenéticas, accesibles por programas de visualización de árboles. Se relaciona con otros módulos utilitarios, necesarios para realizar conversiones de formato, por ejemplo, de FASTA a PHYLIP (único formato considerado válido para el programa de inferencia filogenética PhyML).

## Módulo de visualización de árboles

Se encarga de procesar árboles en formato Newick, y devuelve imágenes con determinado formato, resaltando la secuencia problema. Utiliza la librería *ete2* de la plataforma ETE ToolKit.

## Módulo de análisis de marcadores moleculares

En este módulo se ubica todo el manejo correspondiente a la obtención de los aminoácidos presentes en la secuencia problema, y su posterior análisis. Consiste en pequeños *scripts* que se ocupan de recorrer secuencias de ADN, traducirlas, ubicarlas en los marcos de lectura correctos, y calcular similitudes con respecto a la configuración de marcadores presentes en cada linaje de las cepas de Gumboro.

## Algoritmos utilizados para el análisis filogenético de una secuencia problema

Para **IBDV**, utilizamos PhyML. Se debe convertir el alineamiento en formato FASTA, a un formato aceptado por PhyML (PHYLIP). Se utiliza el *wrapper* de biopython, con los siguientes parámetros:

```
datatype = "nt"  
modelo_sust = "GTR"  
bootstrap = "-4"  
params_opt = "lr"  
alpha = "e"  
prop_invar = "e"
```



Para **IBV**, se utiliza el *wrapper* para FastTree, con los siguientes parámetros:

```
nt = "True"  
gtr = "True"  
gamma = "True"  
cat = 4
```

## Bibliografía

---

---

AngularJS. (2016). Retrieved from <https://angularjs.org/>

Atom Editor. (n.d.). Retrieved from <https://atom.io/>

Bitbucket. (n.d.).

Bootstrap. (n.d.). Retrieved from <http://getbootstrap.com>

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., ... de Hoon, M. J. L. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), 1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>

Eclipse IDE. (n.d.). Retrieved from <https://www.eclipse.org/>

ExpressJS. (n.d.).

Github. (n.d.). Retrieved from <https://git-scm.com/>

Nginx. (n.d.). Retrieved from <https://nginx.org/en/>

NodeJS. (n.d.). Retrieved from <https://nodejs.org/es/>

Tornado. (2016). Retrieved from <http://www.tornadoweb.org/en/stable/>

## Anexo II

Tabla 1- Cepas de IBDV incluidas en el alineamiento de referencia.

<b>Número de acceso</b>	<b>Nombre</b>	<b>País</b>	<b>Linaje</b>
EU544168	B8	Brazil	dIBDV
DQ916182	Co1_01_C5	Colombia	dIBDV
HM241170	A05-HL-144	South-Korea	dIBDV
AF305739	586	Puerto-Rico	dIBDV
Z97013	IBDVRF-8/96	Russia	dIBDV
EF138985	05-32115-B9	Canada	dIBDV
EF138979	03-42857	Canada	dIBDV
KM659888	AR-C1	Argentina	dIBDV
DQ916166	Ar1_05A25	Argentina	dIBDV
KP274853	UY-04/09	Uruguay	dIBDV
KP274852	UY-221201	Uruguay	dIBDV
AF548653	P10	Hungary	dIBDV
AF548656	P3	Hungary	dIBDV
KM659887	UY-171101	Uruguay	dIBDV
KM659893	UY-421101	Uruguay	dIBDV
KM659891	UY-07C	Uruguay	dIBDV
KM659892	UY-42/07	Uruguay	dIBDV
KM659890	UY-06/10	Uruguay	dIBDV
AJ508762	80/GA	Poland	dIBDV
AJ508761	78/GSz	Poland	dIBDV
AB368970	KMRG-48	Tanzania	wIBDV
DQ779963	JRMP13IR	Iran	wIBDV
JF682317	POR241-09	Portugal	wIBDV
AF322444	TASIK	Indonesia	wIBDV
DQ916219	Saf1_01_SA1	South-Africa	wIBDV
EF043078	EL2004-1/TR	Turkey	wIBDV
AF159212	N10	Nigeria	wIBDV
DQ916191	DominicanRepublic99DR4	Dominican-Republic	wIBDV
DQ297821	Uy-1	Uruguay	wIBDV
AF247006	UPM97/61	Malaysia	wIBDV
L42284	KS	Israel	wIBDV
DQ916186	Co2_05C36	Colombia	wIBDV
X92760	UK661	UK	wIBDV
D49706	OKYM	Japan	wIBDV
KM659881	AR-A1	Argentina	wIBDV
Y14956	89163	France	wIBDV
GU646407	09D214	South-Korea	wIBDV
AF159218	K406/89	Egypt	wIBDV
AF240686	D6948	The-Netherlands	wIBDV
AJ878895	849VB	Belguim	wIBDV
DQ916224	Saf2_05SA10	South-Africa	vaIBDV
AY100319	EC-3	Ecuador	vaIBDV
AF006699	GZ902	China	vaIBDV
M64285	A/DEL	USA	vaIBDV

AF413070	BX	China	vaIBDV
DQ916171	Bolivia05B61	Bolivia	vaIBDV
AF281238	T1	USA	vaIBDV
GU646410	09D243	South-Korea	vaIBDV
DQ916184	Co3_01_C9	Colombia	vaIBDV
AF498635	U-28	USA	vaIBDV
X54858	E/DEL	USA	vaIBDV
DQ916212	Mexico04M84	Mexico	vaIBDV
AY368653	GLS	USA	vaIBDV
GQ856674	USA08MD24p	USA	vaIBDV
AF091097	3212	USA	vaIBDV
DQ916239	Spain01_S9	Spain	vaIBDV
DQ916275	Ve1_05VEN33	Venezuela	vaIBDV
DQ916265	Ve2_01_VEN8	Venezuela	vaIBDV
EU544158	Cevac-Gumbo-L	Brazil	atIBDV
GQ985435	P98/05	Taiwan	atIBDV
KM659886	UY-139D	Uruguay	atIBDV
AJ427340	KT1/99	India	atIBDV
KM659883	UY-01/09	Uruguay	atIBDV
EU544151	B4	Brazil	atIBDV
KM659884	UY-130D	Uruguay	atIBDV
KM659885	UY-131D	Uruguay	atIBDV
EF646853	J1C7	Japan	atIBDV
AF499929	D78	USA	atIBDV
D00868	PBG-98	UK	atIBDV
DQ202329	B87	China	atIBDV
AJ310185	CT	France	atIBDV
A33255	Edgar	USA	cvIBDV
AF279288	2512	USA	cvIBDV
DQ916165	Ar2_05A24	Argentina	cvIBDV
EU544162	B15	Brazil	cvIBDV
KM659882	AR-C7	Argentina	cvIBDV
D00499	STC	USA	cvIBDV
AY029166	Irwin_Moulthrop	USA	cvIBDV
Y14958	Faragher_52/70	UK	cvIBDV
AF362747	Cu-1wt	Germany	cvIBDV
AF362773	23/82_Serotype_2	Germany	serotipo 2
U30818	OH_Serotype_2	USA	serotipo 2

Tabla 2- Cepas de IBV incluidas en el alineamiento de referencia.

Número de acceso	Nombre	País	Genotipo
AY851295	Mass	USA	GI-1
FJ888351	H120	The Netherlands	GI-1
EU817497	H52	China	GI-1
NC001451	Beaudette	USA	GI-1
KF696629	Connecticut	China	GI-1
GU361608	Iowa609	USA	GI-2
GU393336	Holte	USA	GI-2
DQ070840	SDW	China	GI-2

## Hoja1

GU393338	JMK	USA	GI-3
GU393334	Gray	USA	GI-3
AF520606	IA/10623/99	USA	GI-3
KC577394	56GX-98I	China	GI-4
AY251816	GX2-98	China	GI-4
L18988	Holte	USA	GI-4
U29522	N1/62	Australia	GI-5
DQ490215	V2-02	Australia	GI-5
U29453	N3/62	Australia	GI-5
U29523	N2/75	Australia	GI-6
KF460437	VicS-v	Australia	GI-6
KF931628	VicS-del	Australia	GI-6
KT946798	CK/CH/GD/GZ14	China	GI-7
KP790143	CK/CH/LDL/140520	China	GI-7
DQ646405.2	TW2575/98	Taiwan	GI-7
JQ964066	L613	USA	GI-8
JQ964071	L919	USA	GI-8
JQ964070	L905	USA	GI-8
AY514485	Cal99	USA	GI-9
GQ504720	Arkansas	USA	GI-9
GQ504721.2	Arkansas	USA	GI-9
AF151959	K87	New Zeland	GI-10
AF151954	B_S1	New Zeland	GI-10
AF151960	T6	New Zeland	GI-10
KM658237	UY/11/CA/18	Uruguay	GI-11
KM658244	AR/11/ER/33	Argentina	GI-11
JX182777	UFMG/12.2	Brazil	GI-11
FN182276	NGA/295/2006	Nigeria	GI-12
X04723	6/82	USA	GI-12
AJ441314	RF/01/02	Russia	GI-12
KF377577	4/91	The Netherlands	GI-13
KP036502	ck/CH/LHLJ/140906	China	GI-13
JX195176	ck/CH/LZJ/111113	China	GI-13
KJ135013	IBVUkr27-11	Ukraine	GI-13
KR231009	B1648	Belgium	GI-14
FN182277	NGA/324/2006	Nigeria	GI-14
AY257062	K281-01	South Korea	GI-15
JQ977697	SNU8067	South Korea	GI-15
FJ807932	B4	South Korea	GI-15
KM658245	UY/09/CA/01	Uruguay	GI-16
KM658255	AR/11/BA/28	Argentina	GI-16
JX195177	ck/CH/LDL/97I	China	GI-16
KP780179	gammaCoV/Ck/Italy/I2022/13	Italy	GI-16
AF419315	CA/Machado/88	USA	GI-17
AF510656	AL/6609/98	USA	GI-17
AF305595	PA/Wolgemuth/98	USA	GI-17
AY296744	JP8127	Japan	GI-18
KC577388	48SD-96VI	China	GI-18
AY296745	JP8443	Japan	GI-18
HM245923	DY07	China	GI-19
JX840411	YX10	China	GI-19

## Hoja1

JF732903	Sczy3	China	GI-19
AF349620	Qu16	Canada	GI-20
AF349621	Qu	Canada	GI-20
DQ901377	It/497/02	Italy	GI-21
AJ457137	Italy-02	Italy	GI-21
DQ386105	Spain/04/5438	Spain	GI-21
HQ850618	GX-YL9	China	GI-22
EU526388	A2	China	GI-22
EU714029	SC021202	China	GI-22
JF893452	YN	China	GI-22
DQ288927	SAIBK	China	GI-22
EU780077	IS/1494/06	Israel	GI-23
KC533682	Eg/1265B/2012	Egypt	GI-23
AF093796	variant	USA	GI-23
KF809795	IBV470	India	GI-24
KF757447	V13	Australia	GI-24
KF757451	V25	Australia	GI-24
EU637854	CK/CH/LSD/05I	China	GI-25
KP085597	GA/13485/2013	USA	GI-25
KM660636	GA/10216/2010	USA	GI-25
FN182243	NGA/B401/2006	Nigeria	GI-26
FN430415	NGA/A116E7/2006	Nigeria	GI-26
FN182269	NGA/N544/2006	Nigeria	GI-26
GU301925	Georgia	USA	GI-27
KM660635	GA/13384/2013	USA	GI-27
KM660631	GA/10231/2010	USA	GI-27
M21971	D1466	The Netherlands	GII
M21968	strain_V1397	The Netherlands	GII
U29521	V18/91	Australia	GIII
U29450	N1/88	Australia	GIII
DQ490219	V6-92	Australia	GIII
GU393332	Delaware	USA	GIV
AF338719	GA/13055/00	USA	GIV
AF274439	GA/5381/99	USA	GIV
DQ059620	N4/03	Australia	GV
DQ059619	N5/03	Australia	GV
DQ059618	N4/02	Australia	GV
JX897900	GX-NN09032	China	GVI
KF668605	CK/CH/SD09/005	China	GVI
KF663559	ck/CH/IBTZ/2012	China	GVI