

# CAPÍTULO I: INTRODUCCIÓN

## 1.1 Formulación del problema

El género *Phytophthora* se presenta hoy en día como una de las mayores plagas para las plantas dicotiledóneas (Armstrong *et al.*, 2005). La especie *Phytophthora infestans* ataca a varios representantes de la familia solanaceae, destacándose la infección al cultivo de la papa (*Solanum tuberosum*) y del tomate (*Solanum lycopersicum*) por su importancia económica.

La preocupación por este género se remonta a los años 1845 y 1846, donde se produjo la devastación de los campos de papa irlandeses por la acción de *Phytophthora infestans*. Este suceso provocó una hambruna que diezmó a gran parte de la población local y obligó a muchos otros a migrar, principalmente hacia Estados Unidos (Edwin y Ribeiro, 1996). En el Ecuador, la enfermedad producida por el *P. infestans* es conocida como lancha de la papa o tizón tardío. Se piensa que esta plaga es originaria de México por la variabilidad que presenta el patógeno en la zona, especialmente en hospederos silvestres (Drenth, 1994). Por esta razón se cree que *P. infestans* ha migrado desde este país hacia todo el mundo, provocando diversas epidemias.

En los últimos diez años se han registrado varios cambios importantes en la población ecuatoriana de *Phytophthora infestans* (posiblemente producidos por migraciones ocurridas desde México). Estos cambios han generado importantes variaciones en los genotipos y por ende en la patogenicidad producida por estos parásitos. Así también la resistencia a fungicidas comúnmente utilizados se ha incrementado, aumentando los costos de producción (por concepto de plaguicidas), disminuyendo el rendimiento de cada cosecha e incrementando el daño ecológico (Andrade *et al.*, 1998).

La alta variabilidad genética y la capacidad del género *Phytophthora* para cambiar, desarrollar resistencia a plaguicidas y colonizar nuevos hospederos han provocado que en los últimos años el interés por la genética de estos organismos aumente. Entender los modelos de interacción planta-patógeno resulta indispensable para generar eficientes sistemas de manejo en cultivos atacados por estas plagas. En la actualidad, el estudio de efectores concita la atención de los investigadores ya que estos son pieza clave en el reconocimiento y por ende el proceso defensivo de la planta contra los patógenos. Una clase de efectores llamados RXLR han sido estudiados fuertemente desde hace un par de años, observados inicialmente en *Hyaloperonospora parasitica*, *P. infestans* y *P. sojae*, estos efectores mantienen un motivo conservado de: arginina (R), cualquier aminoácido (X), leucina (L) y arginina (R), que al parecer está implicado en la translocación del parásito a la célula hospedera. Estas proteínas presentan dos dominios característicos: el dominio N-terminal que involucra al péptido señal y el motivo RXLR y el dominio C-terminal que se encarga de realizar la acción elicitora y operativa dentro de la célula huésped (Lamour *et al.*, 2007). Cada uno de estos dominios se encuentran sometidos a diferentes presiones evolutivas, así el dominio N-terminal sumamente conservado es diferente al motivo C-terminal casi siempre sometido a selección positiva (Win *et al.*, 2006).

## 1.2 Justificación del problema

Según el III Censo Nacional Agropecuario del 2002, el Ecuador cultiva un total de 47.494 hectáreas anuales de papa de las cuales Pichincha cultiva 2.758 de estas. Este cultivo es básico en la alimentación ecuatoriana y a nivel mundial se lo considera como el más importante. En volumen de producción el cultivo de papa se encuentra en el cuarto lugar después del maíz, el trigo y el arroz (Hawkes, 1990).

*Phytophthora infestans* es uno de los principales problemas en papa, se estima que el tizón tardío produce pérdidas globales en países en vías de

desarrollo de alrededor de \$5.3 billones anuales (FAO-CIP, 1995), de las 85 millones de toneladas que producen anualmente estos mismos países (CIP, 1996). Se destina a la protección de lancha entre 8 y 20 % del valor comercial de la producción de papa (Oyarzum *et al.*, 2001).

La variabilidad presentada por los patógenos del género *Phytophthora* hace que las estrategias comunes de control resulten ineficientes después de cierto periodo de aplicación y los esfuerzos en este campo van dirigidos a la obtención de nuevas, eficientes y ecológicas estrategias de control para estas plagas.

Los mecanismos de infección y la variabilidad presentada por el género *Phytophthora* no solo concitan el interés científico sino que puede ser la clave para producir cultivos que expresen resistencia duradera a enfermedades como el tizón tardío. A través de los genes de resistencia (*R*) las plantas pueden defenderse efectivamente de los patógenos ya que el reconocimiento mediado por estos genes identifica proteínas específicas de los patógenos como las proteínas elicitoras y de este modo la planta detiene la infección (sección 1.4.2). El estudio de los genes efectores de estos patógenos facilita el desarrollo de estrategias de mejoramiento. El producir plantas con distintos genes (*R*) disminuirá el uso de plaguicidas, minimizará las pérdidas en la cosecha y evitará que se genere resistencia en los patógenos, impulsando una producción más limpia y ecológica.

### 1.3 Objetivos de la investigación

#### 1.3.1 Objetivo general

Clonar y caracterizar genes efectores de especies patogénicas del género *Phytophthora* que atacan cultivos de la familia solanaceae en los andes del Ecuador

### 1.3.2 Objetivos específicos

- Clonar genes efectores de varias especies de *Phytophthora* en *Escherichia coli*.
- Obtener la secuencia de los genes clonados.
- Generar alineamientos y árboles filogenéticos que muestren la relación entre los diferentes alelos encontrados.
- Analizar las fuerzas evolutivas a las que se encuentran sometidos los distintos genes estudiados.
- Comparar las secuencias proteicas provenientes de los genes analizados para determinar diferencias estructurales entre alelos.

### 1.4 Marco teórico

#### 1.4.1 Generalidades sobre *Phytophthora*

El género *Phytophthora* forma parte del phylum oomycota, subclase peronosporomycetida, orden pythiales, familia pythiaceae. Estos organismos son más parecidos a las diatomeas y a las algas pardas que a los hongos inferiores, grupo en el cual fue clasificado originalmente (Vega, 1999). Los oomycetes poseen paredes celulares conformadas por celulosa y  $\beta$ -glucanos mayoritariamente, con muy poca quitina (principal componente de la pared celular fúngica). Esta es una de las diferencias más importantes que agrupa a *Phytophthora* dentro de los oomycetes (Kamoun, 2003).

Durante su ciclo de vida, las especies de *Phytophthora* son capaces de producir dos tipos de estructuras reproductivas: las zoosporas asexuales (producidas dentro de esporangióforos) que son estructuras reniformes con dos flagelos quimiotáctiles que les permiten nadar, siendo esta una de las principales características que los diferencian de sus antepasados (como los anisocontos y heterocontos). Las oosporas sexuales (de forma generalmente globosa) provienen

de la unión de dos gametangios, el anteridio componente masculino y el oogonio componente femenino (Erwin y Ribeiro, 1996).

*Phytophthora* causa un enorme daño económico en importantes cultivares como papa, tomate, pimiento, soya y alfalfa, de igual forma produce un considerable daño ambiental en ecosistemas naturales. La mayor parte de las plantas dicotiledóneas son afectadas por una o más especies de *Phytophthora*, y varias especies de monocotiledóneas son afectadas también (Kamoun, 2003).

El tizón tardío provocado por el *P. infestans* es sin duda la enfermedad fúngica (se la considera de esta forma aunque el agente que la produce no es un hongo) más seria en papa (*Solanum tuberosum*). Afecta a las hojas (pequeñas manchas de color verde oscuro o claro que se convierten en lesiones pardas o negras según el grado de humedad), las raíces (decoloración superficial e irregular), tuberculos y tallos (se debilitan y mueren de la lesión hacia arriba). Es capaz de devastar un campo en pocos días y causar pérdidas del 100%. Se desarrolla más rápidamente en temperaturas bajas y alto grado de humedad (Henfling, 1987).

Al igual que *P. infestans*, existen especies relacionadas que atacan especies cultivadas y silvestres, entre ellas tenemos a: *P. mirabilis*, *P. phaseoli*, *P. ipomeae* y *P. andina*, todas ellas pertenecen al grupo morfológico IV dentro de la filogenia de *Phytophthora* (Martin y Tooley, 2003).

*Phytophthora mirabilis* es originario de México y fue reportado originalmente como *P. infestans* por Servin (1958), más tarde se notaron diferencias que llevaron a clasificarlo como una especie aparte. Se sabe que está muy relacionado con *P. infestans*, con la diferencia que se lo ha reportado infectando únicamente a *Mirabilis jalapa* y exclusivamente dentro de México (Erwin y Ribeiro, 1996).

*Phytophthora phaseoli* fue descrito por primera vez en Estados Unidos por Thaxter (1889) tras producir la enfermedad del moho plumoso en *Phaseolus lunatus* (Frijoles de lima), de ahí en adelante se lo ha reportado en por lo menos 7 hospederos más, incluyendo tomate. Es por esto que este tipo de *Phytophthora* representa una plaga de consideración para varios cultivos de interés comercial (Edwin y Ribeiro, 1996).

*Phytophthora ipomoeae* es otra especie proveniente de México, específicamente de las montañas del centro de este país. Varias investigaciones con marcadores moleculares revelan que se trata de una nueva especie, relacionada cercanamente con *P. infestans* y *P. mirabilis*. Fue recogida por primera vez por Sánchez (1968) de las hojas de *Ipomoea longipedunculata*. Esta especie es causante del tizón de hoja, al igual que *P. mirabilis* (Flier, 2001).

*P. andina*, fue recolectada por primera vez de *Solanum brevifolium* y al parecer es nativa de los Andes. En la caracterización realizada con marcadores moleculares, *P. andina* parece estar cercanamente emparentada con *P. infestans* (Adler *et al.*, 2004; Oliva *et al.*, 2007). El análisis filogenético sugiere que *P. andina* se encuentra emparentada con *P. infestans*, *P. mirabilis*, *P. ipomoeae* y *P. phaseoli* (Kroon *et al.*, 2004). De acuerdo a las investigaciones realizadas hasta el momento, se ha visto a esta especie atacando a *Solanum betaceum* y al complejo de solanáceas *Anarrhichomenum* (Ordoñez *et al.*, 2000; Adler *et al.*, 2004; Flier *et al.*, 2004, datos no publicados).

#### 1.4.2 Reacción de hipersensibilidad en plantas

La interacción entre patógenos y hospederos ocasiona que las plantas desarrollen sistemas de defensa ante sus agresores, así por ejemplo tenemos: tricomas, callos, fitoalexinas, cutículas, suberinas, ligninas, entre otros. Por otro lado, las plantas han implementado mecanismos moleculares de defensa, entre los cuales se registra a los genes de resistencia. Esta clase de genes son capaces

de expresar proteínas que combaten efectivamente a los patógenos (Agrios, 1988).

Los genes de resistencia son divididos normalmente en:

Genes menores.- aquellos que confieren a la planta resistencia inespecífica ante varios agresores, también llamada resistencia horizontal, la cual tiene carácter cuantitativa (Ordóñez, 1993).

Genes mayores (*R*).- son aquellos que confieren resistencia específica a la planta frente a un patógeno, también llamada resistencia vertical, tiene carácter cualitativo (Ordóñez, 1993). Casi siempre son dominantes y su genes recesivos (*r*) son llamados de susceptibilidad (Agrios, 1988).

Por otro lado los patógenos (entre ellos *Phytophthora*) producen sustancias encargadas de reprogramar, infectar y promover la enfermedad en las células que llegan a parasitar. A estas moléculas se las ha denominado efectores y comprenden a las toxinas (moléculas que causan muerte celular para facilitar la colonización del hospedero) y a los elicitores (moléculas que desencadenan una respuesta por parte del hospedero que resulta en la mejora de la resistencia al patógeno). Dentro del grupo de los elicitores se hallan las proteínas de avirulencia (AVR), identificadas en varios organismos tales como *H. parasitica* y varias clases de *Phytophthora* (Kamoun, 2006).

La interacción planta-patógeno está caracterizada por una relación gen-por-gen (Flor, 1971). En la cual, a cada gen de avirulencia (*Avr*) o elicitor (proteína elicitor) del patógeno le corresponde un gen de resistencia (*R*) en el genoma de la planta. Cuando existe el reconocimiento de una elicitor por parte de una proteína (*R*) se desemboca una cascada de procesos que culmina con la muerte de la célula y el parásito que lleva consigo, a este suceso se lo ha denominado Reacción de Hipersensibilidad (HR). Durante la HR, la célula infectada pierde

permeabilidad en las membranas celulares, incrementa su respiración, acumula y oxida compuestos fenólicos y produce fitoalexinas (Agrios, 1988).

Cuando el patógeno expresa un gen *Avr* y la planta expresa el gen específico *R* correspondiente, la HR es inducida y la infección se detiene. A este fenómeno se lo llama interacción incompatible. Cuando un gen *Avr* es expresado sin que tenga su gen *R* correspondiente en la planta, la HR no es activada y la enfermedad se produce. A este fenómeno se lo denomina interacción compatible (Taylor, 1996).

La reacción de hipersensibilidad (HR) es una respuesta rápida y fulminante contra los parásitos que son capaces de franquear los sistemas de defensa iniciales. Se caracteriza por su velocidad y especificidad el momento de matar células infectadas (Goodman y Novacky, 1994).

Existen dos modelos planteados sobre la interacción entre genes *R* y *Avr*. El modelo "receptor-ligando" indica que las proteínas AVR son reconocidas directamente por las proteínas R. Mientras que el modelo denominado "hipótesis guardia" sugiere que las proteínas R detectan la acción de las proteínas AVR, cuando estas han entrado en la célula hospedera y se encuentran efectuando su acción elicitora. Según estos modelos, las proteínas R determinan si se trata de una proteína de avirulencia, y si el reconocimiento es positivo se desencadena la reacción de hipersensibilidad (Dangl y McDowell, 2006).

Existen varios tipos de proteínas R, siendo las más comunes las del tipo NBS-LRR (sitio de unión nucleotídica con repeticiones ricas en leucina: siglas en inglés). Se piensa que la porción LRR está encargada del reconocimiento específico ya que puede presentar gran diversidad dentro de la planta (Sthali y Bishop, 2000). La diversidad alélica de estas proteínas R está relacionada con la capacidad evolutiva de las proteínas AVR correspondientes. Se han visto proteínas del tipo NBS-LRR implicadas en los dos tipos de reconocimientos, tanto

de forma directa (receptor-ligando) como de forma indirecta (hipótesis guardia) (Dangl y McDowell, 2006).

Tanto los genes *R* como los genes *Avr*, están sometidos a fuerte presión selectiva, dado que mantienen una interacción co-evolutiva constante. Se ha propuesto el modelo de “carrera armamentista” (Thomas y Stephen, 1999; Sthl y Bishop, 2000), en el cual cada vez que un nuevo gen *R* aparece genera una presión selectiva sobre el patógeno. La presencia del “nuevo gen *R*” hace que los individuos avirulentos sean seleccionados en contra y los alelos virulentos poco comunes incrementen su frecuencia hasta alcanzar un equilibrio (McDonald, 2004).

#### 1.4.3 Genes de avirulencia en patógenos

Aquellos genes cuyos productos provocan una respuesta de hipersensibilidad en la planta se los ha denominado genes de avirulencia (*Avr*). Sin embargo, no se ha demostrado con exactitud si estos genes participan directamente en la patogenicidad del organismo. Se sospecha que los genes de avirulencia (*Avr*) cumplen funciones específicas en el patógeno y que muchos de ellos ayudan de una forma indirecta a provocar enfermedades (Taylor, 1996).

Las proteínas de avirulencia pueden ser reconocidas por los receptores en el citoplasma o en el apoplasto de la planta. El mecanismo por el cual los genes de avirulencia provocan la reacción de hipersensibilidad no está del todo dilucidado y hoy en día es un apasionante tema de estudio (Taylor, 1996).

En la década de los 80's se realizaron varios descubrimientos sobre proteínas elicitoras (elicitinas) y genes efectores, es así que se empieza a identificar varios genes de avirulencia (un gen de avirulencia es un gen efector, pero los genes efectores no son necesariamente de avirulencia) en distintos parásitos como bacterias y hongos (incluyendo oomycetes), por ejemplo el gen *avrD* de *Pseudomonas syringae* y el gen *avr9* de *Cladosporium fulvum*. De igual

forma se descubrieron nuevas elicinas provenientes de diversas especies de *Phytophthora*, entre ellas *P. parasitica*, *P. cinnamomi*, *P. cactorum* (Kamoun, 1993).

El primer gen *Avr* clonado en *P. infestans* fue el denominado *Avr3a* (Amsntrong *et al*, 2005). El gen que codifica la proteína AVR3a de *P. infestans* tiene 2 formas alélicas representativas que difieren en 3 aminoácidos, una de las cuales AVR3a<sup>KI</sup> (C<sup>19</sup>, K<sup>80</sup>, I<sup>103</sup>) se expresa en aislados avirulentos, mientras que el AVR3a<sup>EM</sup> (S<sup>19</sup>, E<sup>80</sup>, M<sup>103</sup>) se expresa en aislados virulentos. El alelo avirulento AVR3a<sup>KI</sup> es reconocido positivamente por el gen de resistencia *R3a* (*S. tuberosum*), mientras que la variante virulenta AVR3a<sup>EM</sup> de alguna forma escapa al reconocimiento del hospedero. Esto sugiere que los genes de avirulencia se encuentran en continua selección para tratar de escapar al reconocimiento de la planta (Bos, *et al.*, 2006).

A las proteínas efectoras (proteínas de avirulencia incluidas) se las ha clasificado en dos grupos según el sitio de reconocimiento en el hospedero (figura 1.1):

Efectores apoplasmáticos; son secretados dentro del espacio extracelular de la planta, ahí interactúan con los objetivos extracelulares y los receptores de membrana.

Efectores citoplasmáticos; se traslocan dentro de la célula vegetal, aparentemente por estructuras especializadas como vesículas de infección y haustorios (Kamoun, 2006).

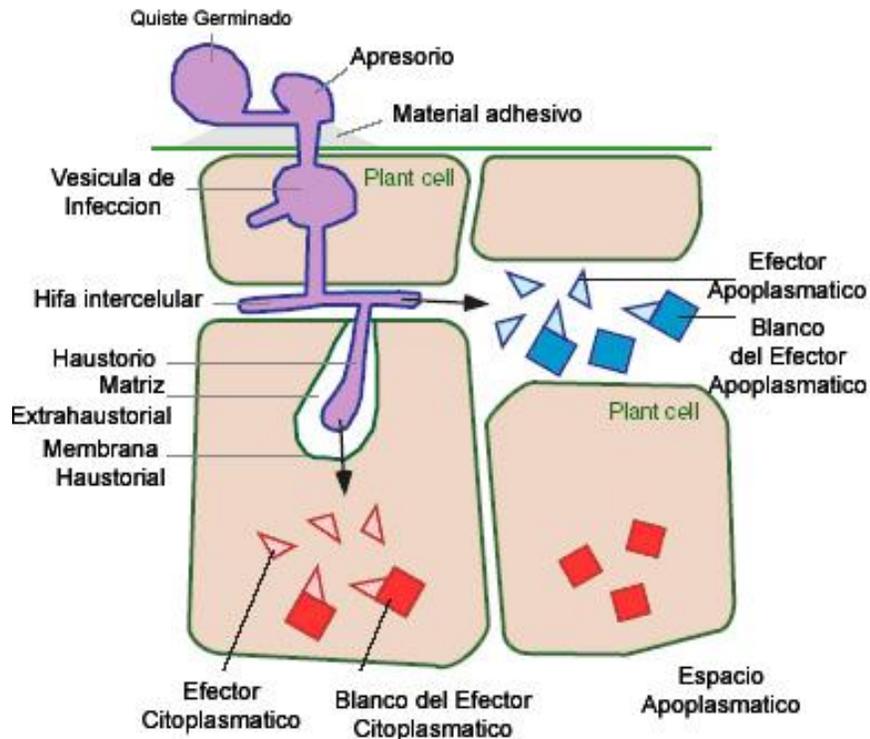


Figura 1.1 Vista esquemática de las etapas tempranas en la infección de *Phytophthora infestans* a una célula huésped. Se ilustran los sitios de acción de los efectores apoplasmáticos y citoplasmáticos (Kamoun, 2006).

La familia de proteínas RXLR (un tipo de efectores citoplasmáticos) son un grupo de efectores que llevan consigo el motivo conservado RXLR (arginina, cualquier aminoácido, leucina, arginina). Usualmente se sugiere que el motivo RXLR funciona como señal para mediar el paso de la proteína desde el haustorio hacia el interior de la célula vegetal (el mecanismo por el cual la elicitorina pasa del haustorio al citoplasma vegetal y a partir de ahí como esta encuentra sus blancos en la célula vegetal es aún un misterio) (Birch *et al.*, 2006). Las proteínas y genes *Avr* que provienen de oomycetes pertenecen en gran parte a la familia de proteínas RXLR. Dentro de este grupo se ha descrito por ejemplo a las proteínas ATR1 (*H. parasitica*), ATR13 (*H. parasitica*), AVR3a (*P. infestans*), AVR1b (*P. sojae*). Todas ellas provocan un reconocimiento en la planta huésped, a través de las proteínas R (Kamoun, 2006). El motivo RXLR es similar al de las proteínas encargadas de la infección en la malaria (provocada por el género *Plasmodium*), lo

cual sugiere que los mecanismos utilizados por patógenos (eucariotas) de animales y plantas están emparentados (Kamoun, 2006).

Actualmente se sigue encontrando nuevos genes de avirulencia y utilizando la información disponible se ha diseñado programas bioinformáticos para buscar genes efectores dentro de las librerías de expresión de *Phytophthora*. Así por ejemplo tenemos a los genes candidatos RD2, RD6, RD11, RD17, RD39/40, entre otros, los cuales están siendo sometidos a pruebas de funcionalidad (efecto de sus proteínas dentro de las plantas) para ser catalogados dentro del grupo de los genes de avirulencia.

Además de los genes *Avr* existe gran cantidad de información sobre genes y proteínas efectoras, tanto apoplasmáticas como citoplasmáticas. Es así que se conocen en mayor o menor grado a diferentes grupos de proteínas como: inhibidores de glucanasas (GIP), inhibidores de serin-proteasas (EPI), inhibidores de cistein-proteasas, pequeñas proteínas ricas en cisteína, elicinas INF1, transglutaminasas, proteínas CRN, entre otras, que conforman una pequeña parte de las más de 100 proteínas que se piensa están dedicadas a la manipulación estructural y funcional de las células huésped, y que de una u otra forma tienen la función de facilitar la colonización de la célula hospedera (Kamoun, 2006).

El conocimiento de los genes de *Phytophthora* implicados en la infección de plantas, permitirá diseñar estrategias efectivas en el control de las enfermedades que aquejan a los cultivos y nos proporcionará información básica sobre la complejidad genética de estos patógenos.

#### 1.4.4 Análisis bioinformático de secuencias

La bioinformática es un conjunto de algoritmos matemáticos y técnicas informáticas desarrolladas para permitir el avance de las ciencias de la vida a través del manejo, simulación y almacenamiento de datos. A partir del desarrollo

de esta disciplina se ha podido ver adelantos increíbles dentro de la biología y ramas afines. El análisis de secuencias a través de programas informáticos se ha convertido en parte fundamental de la biología molecular y hoy en día los análisis *in silico* permiten inferir resultados muy precisos a partir únicamente de secuencias (nucleotídicas o proteicas). Es así que se pueden realizar análisis de funcionalidad, evolución, filogenia, estructura, entre otros, en un tiempo mínimo y con bajo un bajo margen de error.

En la actualidad existen diferentes programas capaces de realizar análisis de secuencias, muchos de ellos especializados en un propósito particular, por ejemplo: CLUSTAL W (programa que permite comparar secuencias con bases de datos de instituciones dedicadas a la investigación biológica), BioEdit (programa diseñado para facilitar la tarea de edición de secuencias), MEGA 3 (programa que permite realizar reconstrucciones filogenéticas, análisis de neutralidad entre otros), DnaSP (programa que permite analizar polimorfismos, neutralidad y realizar simulaciones de coalescencia), TOPALI (programa que permite realizar análisis de recombinación de secuencias), PAML (programa que permite detectar sustituciones, ya sea a nivel de nucleótidos o codones de acuerdo a distintos modelos), YASARA (programa modelador de moléculas complejas) entre otros.

#### 1.4.4.1 Alineamiento de secuencias

Generalmente las secuencias de ADN son perfectamente copiadas durante el proceso de replicación, pero en ciertas ocasiones se producen errores que causan polimorfismos en las secuencias. A estos cambios con respecto al molde original se los denomina mutaciones y se los puede clasificar según el tipo de evento que los ha causado en: sustituciones (reemplazo de un nucleótido por otro), deleciones (pérdida de uno o más nucleótidos), inserciones (introducción de uno más nucleótidos en la cadena) e inversiones (rotación de un segmento de dos o más nucleótidos en ciento ochenta grados) (Li y Graur, 1991).

La comparación de dos o más secuencias homólogas, involucra la identificación de mutaciones que pueden ocurrir en cualquiera de las secuencias (estas mutaciones representan la divergencia hacia un ancestro común). A esta comparación se la conoce como alineamiento de secuencias (Li y Graur, 1991).

Las deleciones o inserciones que ocurren en una secuencia u otra son reflejadas en los denominados gap's (agujeros en la secuencia). Cualquier otro tipo de mutación identificada será expresada en el alineamiento como una desigualdad o "mismatch".

Se han creado varios algoritmos para el alineamiento de secuencias, los mismos que pueden resumirse en los siguientes métodos: matriz de comparación, alineamientos globales, alineamientos locales e inspección visual. De estos cuatro, la inspección visual no tiene un algoritmo bien definido. Para comparar múltiples secuencias generalmente se usa un alineamiento en pares o "pairwise alignment" (Hillis y Moritz, 1990)

En el método de la matriz de comparación se coloca a una de las dos secuencias en el eje vertical de un sistema coordenado y a la otra secuencia en el eje horizontal, de esta forma se va dibujando puntos en los lugares de semejanza o caso contrario el espacio queda en blanco (Hillis y Moritz, 1990).

Los alineamientos globales se refieren a la búsqueda y alineamiento de regiones homólogas enteras (en general secuencias similares con aproximadamente los mismos tamaños), usando un balance entre las coincidencias de nucleótidos y gap's. Este método logra extender las secuencias (ingresando gap's) para que terminen con el mismo tamaño. De esta manera se ubican los lugares en donde las secuencias no coinciden ya que cada carácter (nucleótido o aminoácido) está emparejado a un carácter idéntico un gap o una desigualdad (mismatch). Los gap's necesitan ser penalizados para que el alineamiento pueda darse, ya que no es lo mismo el tamaño del gap (espacios que

ocupa dentro de la secuencia) que el número de gap's a lo largo de la secuencia (generalmente más fuertemente penalizados) (Hillis y Moritz, 1990).

Los alineamientos locales encuentran todas las coincidencias de subsecuencias en una secuencia entera. Este algoritmo es el más utilizado en los buscadores de secuencias (como BLAST por ejemplo) (Hillis y Moritz, 1990).

Muchos de los programas especializados en alineamientos, trabajan con programación dinámica, inspirados en el método de Needleman y Wunsch, en el cual se utiliza calificaciones positivas para las coincidencias entre pares de bases (generalmente 1), cero para las no coincidencias (mismatch) y valores negativos para los gap's. Los gap's son necesarios para maximizar la puntuación de similaridad al sumar todas las posiciones a lo largo de la cadena (también se puede calcular el puntaje de disimilaridad) (Hillis y Moritz, 1990).

#### 1.4.4.2 Árboles filogenéticos

Existen varios métodos para realizar árboles filogenéticos (ilustraciones de la relación evolutiva entre grupos de Unidades Taxonómicas Operacionales [OTU's]), cada uno de estos métodos tienen características y algoritmos de construcción definidos y de una u otra forma representan una relación evolutiva bastante aproximada entre unidades taxonómicas. Un árbol filogenético está compuesto por brazos y nodos. Los nodos representan las unidades taxonómicas y los brazos definen la relación entre estas. La longitud del brazo representa usualmente la cantidad de cambios ocurridos entre los nodos que conecta. Las unidades taxonómicas representadas por los nodos del árbol pueden ser en general: poblaciones, especies, individuos o genes (Li y Graur, 1991). Algunos de los métodos más utilizados para realizar reconstrucciones filogenéticas se describen a continuación.

## Método UPGMA

La reconstrucción a través de UPGMA (método no ponderado de apareamiento de grupos a través de la media aritmética, siglas en inglés), es el método de matriz de distancias más simple de todos, se utiliza particularmente cuando la medida de la distancia de cada valor esperado es aproximadamente proporcional al tiempo evolutivo. En simulaciones computacionales se ha mostrado que cuando las distancias estimadas están sujetas a una gran cantidad de errores estocásticos (errores atribuidos al azar), el UPGMA es usualmente superior a otros métodos de matriz de distancias para recobrar el árbol verdadero (árbol que representa la verdadera historia evolutiva de sus unidades taxonómicas) (Nei, 1987).

De manera general este algoritmo consiste en identificar los OTU's más parecidos entre sí, unirlos y seguir comparando, partiendo de una raíz hipotética. De esta forma se van formando grupos de OTU's que pueden ser comparados con otros grupos antes formados o con OTU's individuales. Cuando dos OTU's tienen el menor valor dentro de la matriz de distancias son unidos entre sí por un nodo (Backeljau *et al.* 1996).

## Método Neighbor Joining

Este método no asume que los valores de divergencia de las unidades taxonómicas sean iguales (diferente tiempo evolutivo por brazo) y puede incluir diferencias taxón por taxón en rangos promedio. Este algoritmo produce una matriz de distancias modificada donde la diferencia entre cada par de nodos es ajustada sobre la base de sus promedios de divergencia con relación al resto de nodos (normalización de la divergencia de cada taxón) (Hillis y Moritz, 1990).

## Método de evolución mínima

Este método asume que el árbol con la menor extensión de sus brazos ( $S$ ), es el más cercano a ser el real. Matemáticamente es estimado a través de mínimos cuadrados. El primer paso en este método consiste en generar un árbol Neighbor Joining (NJ) y a partir de este generar nuevos árboles (o combinaciones de datos con respecto al original), de los cuales se va calculando y registrando el valor de  $S$  (suma de la extensión de los brazos), al final se computa el árbol con menor valor de  $S$  (Rzhetsky y Nei, 1993).

## Métodos de parsimonia

Existen varios métodos de parsimonia (por ejemplo los métodos de Fitch y Wagner, Dollo, Camin-sokal y Transversión), pero el más utilizado en el análisis de secuencias consiste en inferir en los aminoácidos o nucleótidos de los ancestros de los taxones y escoger el árbol con el menor número de cambios mutacionales. El resultado de este proceso es llamado árbol de máxima parsimonia. La distancia de las brazos del árbol no es obtenible salvo bajo ciertas asunciones (Nei, 1987).

## Bootstrap

Los análisis con bootstrap se basan en el muestreo al azar de datos y se utilizan complementariamente a los árboles filogenéticos. Se realizan sustituciones en los valores originales (datos de agrupamiento en un árbol) para proporcionar una colección de nuevos datos que difieren de los originales, de estos se calcula los límites de confianza y se vuelve a muestrear, de esta manera se escogen los valores (en el caso del alineamiento de secuencias los valores resultarían de la divergencia calculada entre secuencias) estadísticamente más adecuados que se

encuentran dentro de los límites de confianza. A través de este criterio estadístico se puede concluir que si un grupo de datos se encuentra unido en un 95% o más de los árboles obtenidos luego del remuestreo, el grupo es aceptado en un nivel del 95 % (Hillis y Moritz, 1990). En el caso de los programas de análisis de secuencias como el MEGA 3 el valor de remuestreo es escogido por el usuario. Este algoritmo es generalmente combinado con los métodos de matriz de distancias (diferencias numéricas entre las distintas secuencias, expresadas en una matriz) para obtener valores de veracidad en los nodos que agrupan los taxones.

#### 1.4.4.3 Neutralidad y selección natural

A partir de los descubrimientos de Charles Darwin, el mundo científico se ha venido preguntando cuál es el mecanismo real que la evolución maneja para generar nuevas especies. La selección natural no lograba responder a muchas de las preguntas que los datos empíricos a nivel molecular arrojaban, así que fueron surgiendo nuevos modelos para explicar muchos fenómenos no entendidos hasta el momento. Es así como se formula el Neo-Darwinismo o teoría sintética de la evolución, que reformula la teoría Darwiniana proponiendo que las mutaciones son la primera fuente de variación genética, pero el mayor rol para crear nuevas especies viene dado por la selección natural. Por otro lado con la aparición de modernas técnicas moleculares, algunos datos obtenidos de estas no parecían concordar totalmente con la teoría sintética, por ejemplo, se notaba en algunos alelos la presencia constante de polimorfismos, los cuales al parecer no se encontraban bajo selección pero sin embargo se mantenían en el genoma. Este tipo de mutaciones llamadas neutras, mantenían un perfil bajo en la teoría Neo-Darwiniana, pero al observar que su presencia era significativa en los genomas, se genera una nueva hipótesis llamada teoría neutral (propuesta por Motoo Kimura), la que proclama que la mayoría de mutaciones a nivel molecular ocurren por fluctuaciones genéticas al azar (deriva génica), además que gran parte de las

variaciones genéticas entre poblaciones son neutras o cercanamente neutras (Nei, 1987).

Fueron necesarios entonces modelos estadísticos y matemáticos para determinar en cada caso si el modelo Neo-Darwiniano o la teoría neutral determinaban la historia evolutiva de las secuencias. Las pruebas de neutralidad fueron desarrolladas para definir si las secuencias se encuentran actuando de una manera neutral o si la selección natural está actuando sobre ellas, ya sea de una manera positiva (una mutación confiere una característica que aumenta la capacidad del organismo de sobrevivir y de reproducirse, dicha mutación se mantiene y se favorece) o purificadora (si la mutación afecta negativamente al organismo en su capacidad de reproducirse y sobrevivir, donde la mutación es desechada) (Li y Graur, 1991).

#### Método de máxima probabilidad para identificar selección positiva

Por definición, una sustitución sinónima de nucleótido es aquella que no genera un cambio de aminoácido en la traducción a proteína, mientras que una sustitución no sinónima es aquella que produce dicho cambio. Generalmente en la comparación de dos secuencias los parámetros  $dS$  y  $dN$  son utilizados, los cuales se definen como el número de sustituciones sinónimas y no sinónimas por sitio en la secuencia, respectivamente. Si el cambio de un aminoácido en un alelo es neutral, este cambio debería ser fijado con el mismo índice que una sustitución sinónima, lo que matemáticamente se expresa como  $dN/dS = 1$ , por el contrario si el cambio de aminoácido es deletéreo, el parámetro  $dN$  será menor que  $dS$  y  $dN/dS$  será  $<1$ . Solamente cuando un cambio no sinónimo resulta ventajoso para el organismo, la fijación del índice  $dN$  será mayor que  $dS$  y por tanto  $dN/dS$  será  $>1$  (Yang y Bielawski, 2000).

El método de máxima probabilidad (ML) (Maximum Likelihood, siglas en inglés), se ha convertido en el medio más sensible para identificar selección positiva, superando a métodos aproximados tales como: Nei y Gojobori, Li, Ina, Yang y Nielsen, entre otros. El método ML toma en cuenta la divergencia entre las secuencias ( $t$ ), el índice de transiciones/transversiones ( $k$ ) y el índice  $dN/dS$  ( $w$ ), parámetros que son estimados de los datos proporcionados por las secuencias (Yang y Bielawski, 2000). El paquete informático PAML a implementado en su programa codeml, modelos (de sustitución de codones) a partir de ML que permiten detectar distintas presiones selectivas a lo largo de la proteína (Yang *et al.*, 2000). El método conlleva dos pasos principales. El primero a cargo de la prueba de índice de probabilidad (LRT) (Likelihood ratio test, siglas en inglés), la cual compara estadísticamente un modelo nulo (detecta mutaciones neutras y/o deletéreas) (M0, M1, M7) contra un modelo alternativo que informa de los sitios bajo selección positiva (M2, M3, M8). Los 6 modelos recomendados por Yang *et al.*, (2000) que son utilizados para detectar selección positiva son: M0 (modelo nulo que asume un  $w$  igual para todos los sitios de la proteína), M1 (modelo nulo que asume una proporción  $p_0$  de sitios conservados con un  $w_0=0$  y una proporción  $p_1=1-p_0$  de sitios neutrales  $w_1=1$ ), M2 (modelo alternativo que además de las asunciones de M1, incorpora una tercera clase de sitios de frecuencia  $p_2=1-p_0-p_1$  con un  $w_2$  calculado a partir de los datos de la secuencia), los modelos M3, M7 y M8 trabajan con diferentes formas de distribución de  $w$  observadas de datos reales, así, el modelo M3 asume una distribución discreta para explicar los distintos índices de  $w$  en la secuencia y asume 3 clases de sitios ( $K$ ) en la secuencia, de proporciones  $p_0$ ,  $p_1$  y  $p_2$ . El modelo M7 asume una distribución beta, de donde calcula las clases de sitios en la secuencia ( $K=10$ ) y el modelo M8 asume igualmente una distribución beta mas un parámetro  $w_1$  extra (asumiendo que  $w_1>1$ ), con proporción  $p_1=1-p_0$  ( $K=11$ , 10 sitios calculados de la distribución beta + 1 sitio  $w_1>1$ ). El

doble de la diferencia de los logaritmos del índice de probabilidad ( $2 \Delta \ln L$ ), entre el modelo nulo y el modelo alternativo, son comparados a la distribución chi-cuadrado  $X^2$  con grados de libertad igual a la diferencia entre el número de parámetros estimados por el par de modelos ( $p$  y  $w$ ). Esta prueba estadística determina si el modelo alternativo describe mejor los datos que el modelo nulo (Liu *et al.*, 2004).

El segundo paso consiste en definir que aminoácido se encuentra bajo selección positiva a través del teorema empírico de Bayes, el cual calcula la probabilidad posterior (probabilidad de que un evento suceda antes de coleccionar los datos) de que un aminoácido esté bajo una clase definida de selección. La probabilidad posterior resulta de la división del índice  $w$  de cada sitio de la proteína para el índice  $w$  del sitio  $K$  de proporción  $p$  más alta (Liu *et al.*, 2004).

#### 1.4.4.4 Análisis de proteínas

La secuencia de una proteína se obtiene cuando se identifica el marco de lectura de la región codificante de una secuencia nucleotídica. Es común analizar a las proteínas de acuerdo a sus niveles estructurales, esto es: estructura primaria o el orden de los aminoácidos en la secuencia proteica o polipeptídica, estructura secundaria (identificación de estructuras de plegamiento entre aminoácidos, como  $\alpha$ -hélices,  $\beta$ -láminas, y otros enrollamientos), estructura terciaria (conformación espacial de la proteína) y si es posible la estructura cuaternaria (si existen varios polipéptidos que se unen para formar una conformación más compleja).

Dado que los aminoácidos no tienen las mismas propiedades entre sí, gracias a la cadena R característica de cada uno de ellos, la estructura primaria aporta ciertas pistas del posible plegamiento que puede tener la proteína en su estructura secundaria y terciaria. El punto isoeléctrico y el peso molecular son generalmente deducidos de la estructura primaria. Comúnmente los 20

aminoácidos esenciales son clasificados por su cadena R en: no polares, polares, ácidos (cargados negativamente), básicos (cargados positivamente) y aminoácidos aromáticos (que comparten características con el grupo de los no polares, por participar en interacciones hidrófobas). De la interacción entre los distintos aminoácidos se puede predecir el plegamiento secundario que presentaría la proteína, ya que cada aminoácido tiene una posibilidad establecida de formar una  $\alpha$  - hélice, una  $\beta$  -lámina u otro enrollamiento no específico. Los niveles estructurales primario y secundario determinan el nivel terciario de conformación proteica, el cual contiene diversos y complejos patrones, los cuales resultan difíciles de predecir si no se cuenta con la información necesaria (Lehninger *et al.*, 1993).

Existen programas especializados en realizar predicciones de estructura primaria y secundaria, algunos de ellos se encuentran libres en Internet, a través del Expasy de Swissprot (2007) (<http://www.expasy.ch/tools/>) por ejemplo. Para determinar la estructura terciaria de una proteína es necesario realizar una aproximación experimental y generar un archivo con la información suficiente para modelar la molécula. Por ejemplo los archivos PDB (en base al Protein Data Bank, siglas en inglés) contienen información de cada átomo de la proteína (distancias, enlaces, entre otros), información que puede ser leída por modeladores de macromoléculas como YASARA, ChemSW, Rasmol, entre otros (Berman *et al.*, 2000).

### 1.5 Sistema de hipótesis

- Las secuencias de los genes de avirulencia provenientes de diferentes especies de *Phytophthora* presentan diferencias.
- Los alelos encontrados en cada gen de avirulencia presentan variaciones intra-específicas.

- Cada uno de los genes estudiados se encuentran sometidos a diferentes tipos de presión evolutiva.
- Estructuralmente los alelos encontrados son diferentes entre sí.

## **CAPÍTULO II: MÉTODOS Y TÉCNICAS**

### 2.1 Participantes

El presente proyecto contó con el respaldo del Centro Internacional de la Papa Quito (CIP-Quito), a través del Biólogo Ricardo Oliva M.Sc investigador adjunto de dicha institución, quien fue el promotor y tutor de la investigación en su totalidad.

Por otro lado el departamento de Patología Vegetal de la Universidad Estatal de Ohio en Estados Unidos, a través del Dr. Sophien Kamoun y la Dra. Carla García, colaboraron cercanamente con este proyecto, proporcionando material de trabajo y asesoría técnica.

Dentro de la Escuela Politécnica del Ejército se contó con el apoyo de la Dra. Karina Proaño y el Ing. Marco Vergara, quienes desempeñaron las funciones de directora y co-director respectivamente.

### 2.2 Zona de estudio

Tanto el trabajo realizado en laboratorio como los análisis realizados en computadora fueron realizados en el Centro Internacional de la Papa-Quito, ubicado en el Cantón Mejía, Parroquia Cutuglahua, Estación Experimental Santa Catalina-INIAP, Panamericana Sur, Km 1.

### 2.3 Periodo de tiempo de investigación

El proyecto fue realizado en 9 meses, iniciándose en el mes de diciembre del 2006 y culminándose en el mes de septiembre del 2007.

## 2.4 Procedimientos

### 2.4.1 Transformación de *Escherichia coli*

La cepa utilizada para la transformación, *Escherichia coli* DH10B fue donada por la Universidad San Francisco de Quito. Las células fueron mantenidas en agar LB Gibco, en refrigeración a 4°C.

Cuatro insertos fueron utilizados en la transformación bacteriana: el gen de avirulencia *Avr3a* (también llamado RD7) y los genes candidatos RD2, RD6 (también llamado IPIO) y RD39/40. Cada uno de los genes fue extraído de aislados particulares de *P. infestans*, *P. andina*, *P. mirabilis* y *P. ipomoeae* (Tabla 2.1). Los fragmentos fueron amplificados con primers específicos e insertados dentro del plásmido pGR106. Las construcciones utilizadas (plásmido+inserto) fueron creadas en el laboratorio del departamento de Patología Vegetal de la Universidad Estatal de Ohio (OSU) en Estados Unidos y donadas a CIP-Quito. Las construcciones fueron mantenidas en agua destilada-desionizada, en congelación a -15°C.

Tabla 2.1: Insertos utilizados en la transformación bacteriana, de acuerdo al gen y al aislado.

Número de Laboratorio	*Código del Aislado	Especie	Hospedero	Gen Utilizado	Código del Inserto
P1	3163	<i>P. andina</i>	<i>Anarrichomenum</i>	RD2	2P1
P1	3163	<i>P. andina</i>	<i>Anarrichomenum</i>	<i>Avr3a</i>	3P1
P1	3163	<i>P. andina</i>	<i>Anarrichomenum</i>	RD6	6P1
P1	3163	<i>P. andina</i>	<i>Anarrichomenum</i>	RD39/40	40P1
P2	3650	<i>P. andina</i>	<i>Anarrichomenum</i>	RD2	2P2

P2	3650	<i>P. andina</i>	<i>Anarrichomenum</i>	Avr3a	3P2
P2	3650	<i>P. andina</i>	<i>Anarrichomenum</i>	RD6	6P2
P2	3650	<i>P. andina</i>	<i>Anarrichomenum</i>	RD39/40	40P2
P3	3399	<i>P. andina</i>	<i>Anarrichomenum</i>	RD2	2P3
P3	3399	<i>P. andina</i>	<i>Anarrichomenum</i>	Avr3a	3P3
P3	3399	<i>P. andina</i>	<i>Anarrichomenum</i>	RD6	6P3
P3	3399	<i>P. andina</i>	<i>Anarrichomenum</i>	RD39/40	40P3
<b>Número de Laboratorio</b>	<b>*Código del Aislado</b>	<b>Especie</b>	<b>Hospedero</b>	<b>Gen Utilizado</b>	<b>Código del Inserto</b>
P4	3598	<i>P. andina</i>	<i>S. betaceum</i>	RD2	2P4
P4	3598	<i>P. andina</i>	<i>S. betaceum</i>	RD6	3P4
P4	3598	<i>P. andina</i>	<i>S. betaceum</i>	RD39/40	40P4
P5	3644	<i>P. andina</i>	<i>S. betaceum</i>	RD2	2P5
P5	3644	<i>P. andina</i>	<i>S. betaceum</i>	Avr3a	3P5
P5	3644	<i>P. andina</i>	<i>S. betaceum</i>	RD6	6P5
P5	3644	<i>P. andina</i>	<i>S. betaceum</i>	RD39/40	40P5
P6	3843	<i>P. infestans</i>	<i>S. tuberosum</i>	RD2	2P6
P6	3843	<i>P. infestans</i>	<i>S. tuberosum</i>	Avr3a	3P6
P6	3843	<i>P. infestans</i>	<i>S. tuberosum</i>	RD6	6P6
P6	3843	<i>P. infestans</i>	<i>S. tuberosum</i>	RD39/40	40P6
P7	3805	<i>P. infestans</i>	<i>S. muricatum</i>	RD2	2P7
P7	3805	<i>P. infestans</i>	<i>S. muricatum</i>	Avr3a	3P7
P7	3805	<i>P. infestans</i>	<i>S. muricatum</i>	RD6	6P7
P7	3805	<i>P. infestans</i>	<i>S. muricatum</i>	RD39/40	40P7
P8	3842	<i>P. infestans</i>	<i>S. habrochaites</i>	RD2	2P8
P8	3842	<i>P. infestans</i>	<i>S. habrochaites</i>	Avr3a	3P8
P8	3842	<i>P. infestans</i>	<i>S. habrochaites</i>	RD6	6P8
P8	3842	<i>P. infestans</i>	<i>S. habrochaites</i>	RD39/40	40P8
P9	3620	<i>P. infestans</i>	<i>S. ochanthum</i>	RD2	2P9
P9	3620	<i>P. infestans</i>	<i>S. ochanthum</i>	Avr3a	3P9
P9	3620	<i>P. infestans</i>	<i>S. ochanthum</i>	RD6	6P9
P9	3620	<i>P. infestans</i>	<i>S. ochanthum</i>	RD39/40	40P9
P10	3561	<i>P. infestans</i>	<i>S. quitoense</i>	RD2	2P10
P10	3561	<i>P. infestans</i>	<i>S. quitoense</i>	Avr3a	3P10
P10	3561	<i>P. infestans</i>	<i>S. quitoense</i>	RD6	6P10
P10	3561	<i>P. infestans</i>	<i>S. quitoense</i>	RD39/40	40P10
P11	3381	<i>P. infestans</i>	<i>S. lycopersicum</i>	RD2	2P11
P11	3381	<i>P. infestans</i>	<i>S. lycopersicum</i>	Avr3a	3P11
P11	3381	<i>P. infestans</i>	<i>S. lycopersicum</i>	RD6	6P11
P11	3381	<i>P. infestans</i>	<i>S. lycopersicum</i>	RD39/40	40P11
P12	3707	<i>P. infestans</i>	<i>S. quitoense</i>	RD2	2P12
P12	3707	<i>P. infestans</i>	<i>S. quitoense</i>	Avr3a	3P12
P12	3707	<i>P. infestans</i>	<i>S. quitoense</i>	RD6	6P12
P12	3707	<i>P. infestans</i>	<i>S. quitoense</i>	RD39/40	40P12

P15	A1477.1	<i>P. ipomoaea</i>	<i>I.longipendunculata</i>	RD2	2P15
P15	A1477.1	<i>P. ipomoaea</i>	<i>I.longipendunculata</i>	Avr3a	3P15
P15	A1477.1	<i>P. ipomoaea</i>	<i>I.longipendunculata</i>	RD39/40	40P15
P16	A1479.1	<i>P. mirabilis</i>	<i>Mirabilis jalapa</i>	RD2	2P16
P16	A1479.1	<i>P. mirabilis</i>	<i>Mirabilis jalapa</i>	Avr3a	3P16
P16	A1479.1	<i>P. mirabilis</i>	<i>Mirabilis jalapa</i>	RD39/40	40P16

\*Código correspondiente al banco de germoplasma de CIP-Quito

El Sílabo de Laboratorio de Venancio Arahana (2006) para transformación de bacterias fue utilizado como base en el desarrollo del protocolo final utilizado.

Para este protocolo, las bacterias *E. coli* cepa DH10B fueron incubadas en tubos de ensayo con 5 ml de LB líquido (peptona 10 g/l; extracto de levadura 5 g/l; NaCl 5g/l; pH 7) en agitación por 3 h, posteriormente se transfirió la suspensión celular a microtubos de 2 ml, los cuales fueron mantenidos en hielo con anterioridad (se debe llenar los tubos). Los tubos fueron centrifugados por 10 min a 3500 rpm a 4 °C. Se resuspendió el pellet en 2 ml de CaCl<sub>2</sub> (autoclavado, pH 7.5, 60 mM) (se debe trabajar en hielo). Se volvió a centrifugar el pellet por 10 min a 3500 rpm a 4 °C y se lo resuspendió en 2 ml de CaCl<sub>2</sub> (autoclavado, pH 7.5, 60 mM) (se debe trabajar en hielo). El pellet fue centrifugado una vez más por 10 minutos a 3500 rpm a 4°C y resuspendido en 200ul de CaCl<sub>2</sub> (autoclavado, pH 7.5, 60 mM) (se debe trabajar en hielo). A la suspensión se le añadió 3 ul del plásmido (pGR106+inserto) y se colocaron los tubos en hielo por 10 minutos. Una vez enfriados los tubos de reacción, estos fueron calentados a 42 °C en un termobloque (TAITEC DTU-1B) por 2 min exactos e inmediatamente transferidos a hielo. Se añadió a cada tubo 1 ml de LB líquido (peptona 10g/l; extracto de levadura 5g/l; NaCl 5g/l; pH 7) precalentado y se incubó a 37 °C por 45 min. Se colocó de 100 a 200 ul de la suspensión en una caja petri con agar LB+Kanamicina (LB Gibco; Kanamicina Gibco 65.5 ug/ml), esparciendo el líquido por toda la superficie del agar con la ayuda de una varilla de vidrio. Por último las cajas fueron incubadas a 37 °C por 18 horas y luego de este tiempo almacenadas en refrigeración a 4 °C.

#### 2.4.2 Respaldo de los clones transformados

Las colonias obtenidas durante la investigación contaron con un respaldo, el cual fue mantenido en agar LB+Kanamicina (LB Gibco; Kanamicina Gibco 65.5 ug/ml). Se procedió a extraer las colonias de su caja de cultivo original, con la ayuda de una punta de micropipeta estéril, para luego ser sembradas en un medio nuevo. La caja petri fue dividida en celdas (aproximadamente 20 por caja). En cada una de las celdas se colocó una colonia pura, la misma que recibió un código específico. Posteriormente las cajas fueron incubadas por 18 horas a 37 °C y luego almacenadas en refrigeración a 4 °C.

#### 2.4.3 Extracción de ADN

Se realizó una extracción rápida de ADN bacteriano según el método descrito en Mayoral *et al.*, (2005). La punta de micropipeta estéril, utilizada en el respaldo de los clones transformados (sección 2.4.2) fue sumergida en un microtubo de 250 ul con 25 ul de agua destilada estéril. En cada tubo se colocó material de 5 colonias. A continuación se colocaron los tubos en un termociclador PT-100 (MJ Research, INC.) por 10 min a 100 °C para lisar la membrana celular y liberar el ADN de cada colonia. Los tubos fueron almacenados a -15 C°.

#### 2.4.4 Selección de clones positivos

La selección de clones fue realizada mediante la técnica de PCR, patentada por Kary Mullis en 1987 (Innis y Gelfald, 1990). Para diferenciar los clones que presentaron la construcción inserto+plásmido (construcción positiva) de los clones que llevaban consigo solamente el plásmido (construcción negativa), se utilizaron los primers PVX-F (5' AATCAATCACAGTGTTGGCTTGC 3') Y PVX-R (5' AGTTGACCCTATGGGCTGTGTTG 3'), los mismos que fueron diseñados en base al plásmido pGR106 (plásmido en el que se insertaron los fragmentos de interés).

El resto de reactivos necesarios para la reacción de PCR fueron proporcionados por la casa comercial Promega.

Los reactivos fueron mezclados (“master mix”) conforme el protocolo de CIP-Quito (CIP, 1999):

<b>Reactivo</b>	<b>Concentración inicial</b>	<b>Volumen</b>
Agua		3.38 ul
Buffer 10X (Promega)	10 X	1 ul
MgCl <sub>2</sub> (Promega)	25.0 mM	1.40 ul
dNTP's (Promega)	2.50 mM	0.20 ul
Primers (PVX-R Y PVX-F)	10 mM	0.20 ul
Taq Polimerasa (Promega)	5 U/ul	0.13 ul
DNA		3.5 ul

El volumen total de cada uno de los tubos fue calculado para una sola reacción de 10 ul. La amplificación se llevó a cabo en un termociclador PT-200 (MJ Research, INC.) bajo el siguiente programa:

<b>Pasos</b>	<b>Temperatura</b>	<b>Tiempo</b>	<b>Ciclos</b>	<b>Observaciones</b>
1	94 °C	5 min	1 ciclo	
2	94 °C	45 seg	1 ciclo	
3	55 °C	45 seg	1 ciclo	
4	72 °C	1 min	1 ciclo	
5			36 ciclos	Repetir del paso 2 al 4
6	72 °C	10 min	1 ciclo	
7	4 °C			Terminación

Concluido el programa de PCR, los productos amplificados fueron corridos en geles de agarosa al 1.5%, en donde se diferenció entre colonias positivas (plásmido + inserto) y negativas (plásmido solamente), con la ayuda de un control negativo (plásmido sin inserto) y un marcador de peso molecular 300 pb. El control negativo produjo fragmentos de 281 pb, mientras que las construcciones positivas produjeron fragmentos de entre 350 pb a 500 pb dependiendo del gen insertado.

#### 2.4.5 Secuenciación de clones positivos

Una vez identificados los clones positivos, estos fueron enviados al laboratorio de Fitopatología de la Universidad Estatal de Ohio, Estados Unidos a cargo del Doctor Sophien Kamoun para su secuenciación.

#### 2.4.6 Análisis bioinformático de secuencias

Las secuencias de los clones positivos fueron editadas en el programa BioEdit, el mismo que permitió revisar los cromatogramas obtenidos del secuenciador para determinar zonas de variabilidad (zonas polimórficas) entre clones, e identificar errores de transformación como de secuenciación. El programa también permitió cortar zonas en las secuencias que no fueron útiles para el análisis, como los fragmentos amplificados del plásmido o residuos que se encontraban contaminando la muestra y por ende impidiendo la correcta lectura de la misma. Una vez depurados los errores detectables en las secuencias, se identificaron los sitios de restricción con que fueron creadas las construcciones inserto + plásmido, para determinar el inicio y el final del fragmento correspondiente al inserto (genes RD2, *Avr3a*, RD6 o RD39/40). Una vez editadas las secuencias, se procedió a comparar cada una de ellas con las secuencias modelo enviadas desde la Universidad Estatal de Ohio (OSU) como referente de cada tipo de gen (cedidas por cortesía de Carla Garzón y Sophien Kamoun)

Luego de alinear las secuencias obtenidas para cada gen (RD2, *Avr3a*, RD6 y RD39/40), se utilizó el método de evolución mínima (Rzhetsky y Nei, 1993) para reconstruir la filogenia de cada gen dentro del programa MEGA 3. Los árboles fueron calculados a partir del número de diferencias entre aminoácido con 1000 replicaciones de bootstrap.

Posteriormente, los alineamientos fueron ingresados en el programa codeml del paquete bioinformático PAML 3.14, para generar una prueba de selección positiva. En todos los casos los modelos neutrales M0, M1, M7 (modelos que detectan los sitios bajo selección purificadora y/o neutra) fueron comparados contra los modelos alternativos M3, M2, M8 (modelos que detectan sitios bajo selección positiva) respectivamente, a través de la prueba de índice de probabilidad (LRT) (likelihood ratio test, siglas en inglés). De esta forma se pudo determinar con una base estadística, a qué modelo se ajusta mejor el patrón de mutación de cada gen. Asimismo fueron identificados los sitios con probabilidades de encontrarse bajo selección positiva (de entre el 50 y el 94%) y los sitios con probabilidades altamente significativas de encontrarse bajo dicha selección (>95%) (Liu *et al.*, 2004; Yang *et al.*, 2000).

Finalmente, se produjo una aproximación de la estructura primaria y secundaria de los alelos representativos de cada gen estudiado. Para el análisis de estructura primaria se utilizó el programa Protéine de la universidad de Provence en Marsella Francia (2007), disponible libremente en Internet ([http://www.iut-arles.up.univ.mrs.fr/w3bb/d\\_abim/compo-p.html](http://www.iut-arles.up.univ.mrs.fr/w3bb/d_abim/compo-p.html)). De este programa se obtuvo el peso molecular y el punto isoeléctrico de las proteínas ingresadas. Para el análisis de estructura secundaria se utilizó el programa LOOP versión 3.0 @ CBSU (2007) de la universidad de Cornell, libre en Internet (<http://cbsuapps.tc.cornell.edu/loopp.aspx>). Este programa proporcionó información sobre el posible plegamiento secundario de las secuencias proteicas

analizadas (mediante los símbolos H para  $\alpha$ -hélice, E para  $\beta$ -lámina y C para los plegamientos no determinados).

## **CAPÍTULO III: RESULTADOS**

### **3.1 Transformación y detección de clones positivos**

El protocolo de transformación aplicado (sección 2.4.1) resultó altamente eficiente, un total de 4260 colonias fueron obtenidas, de las cuales 365 no contaron con respaldo por lo que no fueron tomadas en cuenta para la detección realizada con PCR. De las colonias respaldadas se obtuvieron 779 tubos (5 colonias por tubo) los mismos que fueron destinados a la extracción de ADN, de acuerdo a lo descrito en la sección 2.4.3.

La detección de clones positivos realizada a través de PCR se observa en la figura 3.1, donde se puede apreciar la diferencia de peso existente entre el control negativo (plásmido solamente) y los clones positivos (plásmido + inserto).

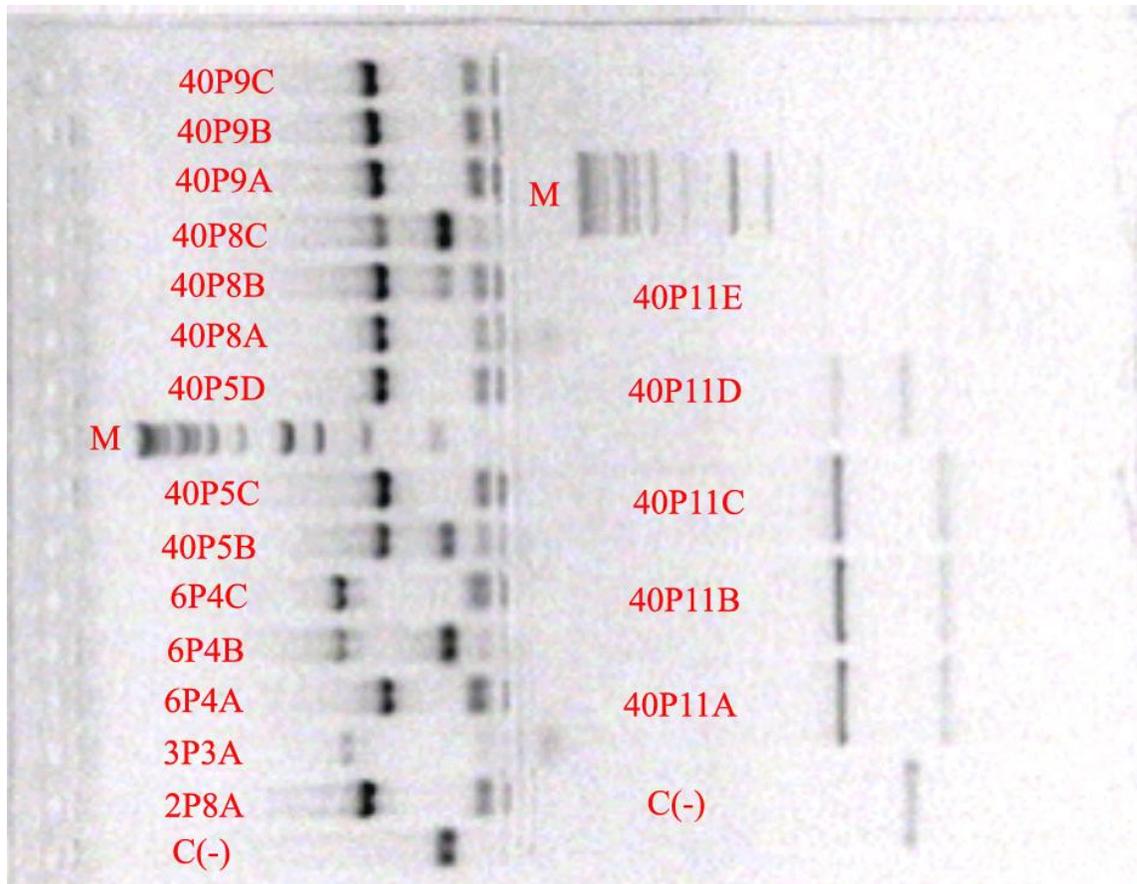


Figura 3.1: Productos amplificados de PCR, corridos en gel de agarosa al 1.5%, donde se muestran 19 construcciones (inserto + plásmido) con distinto peso molecular al control negativo. C (-) Control negativo (únicamente plásmido). M marcador molecular 100 pb.

En total se detectaron 30 clones con inserto, lo que significa que el 0.7% de las transformaciones resultaron positivas. Estas 30 colonias fueron respaldadas según el procedimiento descrito en la sección 2.4.2 y enviadas hacia Ohio State University (OSU) para su secuenciación (Tabla 3.1).

Tabla 3.1: Clones positivos obtenidos y enviados para su secuenciación. Se especifica el nombre de la construcción, el aislado, el orden del clon, el gen y el hospedero de donde fue recolectado el aislado.

#	Nombre de la Construcción	Gen	Orden del Clon	Especie	Hospedero
1	2P1	RD2	A	<i>P. andina</i>	<i>Anarrichomenum</i>
2	2P6	RD2	A	<i>P. infestans</i>	<i>S. tuberosum</i>
3	2P8	RD2	A	<i>P. infestans</i>	<i>S. habrochaites</i>
4	2P16	RD2	A	<i>P. mirabilis</i>	<i>Mirabilis jalapa</i>
5	3P3	Avr3a	A	<i>P. andina</i>	<i>S. betaceum</i>
6	3P6	Avr3a	A	<i>P. infestans</i>	<i>S. tuberosum</i>
7	3P7	Avr3a	A	<i>P. infestans</i>	<i>S. muricatum</i>
8	3P12	Avr3a	A	<i>P. infestans</i>	<i>S. quitoense</i>
9	3P12	Avr3a	B	<i>P. infestans</i>	<i>S. quitoense</i>
10	3P15	Avr3a	A	<i>P. ipomoaea</i>	<i>I.longipendunculata</i>
11	3P16	Avr3a	A	<i>P. mirabilis</i>	<i>Mirabilis jalapa</i>
12	6P1	RD6	A	<i>P. andina</i>	<i>Anarrichomenum</i>
13	6P4	RD6	A	<i>P. andina</i>	<i>S. betaceum</i>
14	6P4	RD6	B	<i>P. andina</i>	<i>S. betaceum</i>
15	6P4	RD6	C	<i>P. andina</i>	<i>S. betaceum</i>
16	40P5	RD39/40	A	<i>P. andina</i>	<i>S. betaceum</i>
17	40P5	RD39/40	B	<i>P. andina</i>	<i>S. betaceum</i>
18	40P5	RD39/40	C	<i>P. andina</i>	<i>S. betaceum</i>
19	40P5	RD39/40	D	<i>P. andina</i>	<i>S. betaceum</i>
20	40P8	RD39/40	A	<i>P. infestans</i>	<i>S. habrochaites</i>
21	40P8	RD39/40	B	<i>P. infestans</i>	<i>S. habrochaites</i>
22	40P8	RD39/40	C	<i>P. infestans</i>	<i>S. habrochaites</i>
23	40P9	RD39/40	A	<i>P. infestans</i>	<i>S. ochanthum</i>
24	40P9	RD39/40	B	<i>P. infestans</i>	<i>S. ochanthum</i>
25	40P9	RD39/40	C	<i>P. infestans</i>	<i>S. ochanthum</i>
26	40P11	RD39/40	A	<i>P. infestans</i>	<i>S. lycopersicum</i>
27	40P11	RD39/40	B	<i>P. infestans</i>	<i>S. lycopersicum</i>
28	40P11	RD39/40	C	<i>P. infestans</i>	<i>S. lycopersicum</i>
29	40P11	RD39/40	D	<i>P. infestans</i>	<i>S. lycopersicum</i>
30	40P11	RD39/40	E	<i>P. infestans</i>	<i>S. lycopersicum</i>

### 3.2 Secuenciación

Las secuencias de 21 clones lograron ser procesadas en OSU, de estas, 18 se encontraban correctamente secuenciadas y se pudo recabar información. Las secuencias 3P7A, 6P1A y 6P4A no coincidieron con el resto de secuencias, por lo que se sospecha en el caso de la secuencia 3P7A la presencia de contaminación. En el caso de las secuencias 6P1A y 6P4A la divergencia pudo deberse a una amplificación inespecífica en la secuenciación o a falsos positivos. Nueve de los 30 clones positivos no pudieron ser secuenciados, ya que la amplificación con primers específicos resultó negativa para dichas colonias (estas 9 secuencias no fueron enviadas desde OSU) (ver tabla 3.2).

Tabla 3.2. Resultado obtenido de la secuenciación realizada a los clones positivos. Se especifica el nombre de la secuencia y el sentido del primer con el que fue amplificado: Forward (F) y Reverse (R).

#	Nombre de la secuencia	Forward (F)	Reverse (R)
1	2P1A	----	X
2	2P6A	X	----
3	2P8A	X	X
4	2P16A	X	----
5	3P3A	X	X
6	3P6A	----	X
7	3P7A	N	----
8	3P12A	X	----
9	3P12B	----	----
10	3P15A	----	----
11	3P16A	----	----
12	6P1A	N	----
13	6P4A	N	N
14	6P4B	----	----
15	6P4C	----	X
16	40P5A	----	X
17	40P5B	----	----
18	40P5C	X	----
19	40P5D	----	X

#	Nombre de la secuencia	Forward (F)	Reverse (R)
20	40P8A	----	----
21	40P8B	----	X
22	40P8C	----	----
23	40P9A	----	----
24	40P9B	----	X
25	40P9C	----	X
26	40P11A	----	X
27	40P11B	X	X
28	40P11C	X	X
29	40P11D	X	X
30	40P11E	----	----

X secuencias útiles para el análisis bioinformático.

---- secuencias no enviadas desde MCIC (centro de secuenciación de Ohio State University OSU).

N secuencias no utilizadas en el análisis bioinformático (resultantes de clones falso-positivos o contaminados).

### 3.3 Análisis bioinformático

Los análisis realizados fueron básicamente 3: Alineamiento de las secuencias y reconstrucción filogenética, análisis de neutralidad y análisis de proteínas. Para el gen RD2 se obtuvieron secuencias pertenecientes a *P. andina*, *P. infestans* y *P. mirabilis*, para el gen *Avr3a* secuencias de *P. andina* y *P. infestans*, para el gen RD6 se obtuvo únicamente una secuencia de *P. andina* y para el gen RD39/40 secuencias de *P. andina* y *P. infestans*.

#### 3.3.1 Alineamiento de secuencias y reconstrucción filogenética

Al realizar el análisis del gen RD2 se encontró un alelo único en *P. andina* (2P1A) que difiere en 11 nucleótidos con respecto a las secuencias del gen RD2 de *P. infestans*. De estos cambios se derivan 8 diferencias a nivel de aminoácidos entre dichos alelos. Además se encontró una diferencia de nucleótido en las secuencias 2P6A y 2P8A provenientes de *P. infestans* en distintos sitios cada una, pero en la traducción a proteínas no implicaron cambios de aminoácidos (cambio sinónimo). La secuencia proveniente de *P. mirabilis* 2P16A no presentó diferencias

con respecto a las secuencias de *P. infestans*, como se puede observar en el alineamiento de la figura 3.2.

```

RD2_osu      CTCTCGACAA CCACGGGTGT TCAGGCTGCT AATTGCTAG GCCCAGCCCA GCGTCTACTG [ 60]
2P1A        .....A.....T..... [ 60]
2P6A-F      .....C..... [ 60]
2P8A-F      ..... [ 60]
2P16A-F     ..... [ 60]
osu-CU1-RD2R ..... [ 60]
osu-CU2-RD2R ..... [ 60]
osu-CU4-RD2R ..... [ 60]

RD2_osu      AGCAAACT ACACGGCAGC TGAAAACGAC GATGACTCTG AAGCAAGGGC TCTGAATACA [120]
2P1A        .....G...T....A.....C... [120]
2P6A-F      ..... [120]
2P8A-F      ..... [120]
2P16A-F     ..... [120]
osu-CU1-RD2R ..... [120]
osu-CU2-RD2R ..... [120]
osu-CU4-RD2R ..... [120]

RD2_osu      GAGAAATGA AAACGATGTT GAAAGCTGGG ATGACTGTTG ACGACTACGC TGCCAAAGCTA [180]
2P1A        .....G.....G..... [180]
2P6A-F      ..... [180]
2P8A-F      .....T..... [180]
2P16A-F     ..... [180]
osu-CU1-RD2R ..... [180]
osu-CU2-RD2R ..... [180]
osu-CU4-RD2R ..... [180]

RD2_osu      AAACCTACCG ACAAGATTGC AGCTGCAGCT AACTCTGCAA GGGCGATGGA AAAGCTTGCC [240]
2P1A        .....A.....A..... [240]

```

Figura 3.2: Alineamiento del gen RD2 donde se muestran 10 de los 11 sitios de variabilidad a nivel de nucleótidos entre la secuencia de *P. andina* (2P1A) y las secuencias pertenecientes a *P. infestans* (2P6A, 2P8A y secuencias OSU) y *P. mirabilis* (2P16A) (secuencias Ohio State University OSU, cortesía de Carla Garzón (RD2-OSU) y Sang Keun Oh (OSU-CU)).

El motivo RLLR fue identificado en todas las secuencias, como se muestra en el alineamiento a nivel de aminoácidos de la figura 3.3.

```

RD2_osu      LSTTTGVQAA NLVGPAQRLL RKHYTAAEND DDSEARALNT EKMKTMLKAG MTVDDYAAKL [ 60]
2P1A        .....M.....G.....P.....W...G..... [ 60]
2P6A-F      ..... [ 60]
2P8A-F      ..... [ 60]
2P16A-F     ..... [ 60]
osu-CU1-RD2R ..... [ 60]
osu-CU2-RD2R ..... [ 60]
osu-CU4-RD2R ..... [ 60]
osu-CU6-RD2R ..... [ 60]

RD2_osu      KLTDKIAAAA NSARAMEKLG ETLKMKLLR YLNYVAEHTA V* [102]
2P1A        ..N.....K.....H..... [102]
2P6A-F      ..... [102]
2P8A-F      ..... [102]
2P16A-F     ..... [102]
osu-CU1-RD2R ..... [102]
osu-CU2-RD2R ..... [102]
osu-CU4-RD2R ..... [102]
osu-CU6-RD2R ..... [102]

```

Figura 3.3: Alineamiento del gen RD2 a nivel de aminoácidos, donde se observan 8 diferencias entre la secuencia de *P. andina* (2P1A) y las secuencias pertenecientes a *P. infestans* (2P6A, 2P8A y secuencias OSU) y *P. mirabilis* (2P16A). Motivo RLLR se encuentra resaltado en el alineamiento (secuencias Ohio State University (OSU), cortesía de Carla Garzón (RD2-OSU) y Sang Keun Oh (OSU-CU)).

La reconstrucción filogenética en base al número de diferencias entre aminoácidos, logró mostrar claramente la diferenciación que existe entre el alelo 2P1A de *P. andina* y el resto de secuencias que corresponden a un mismo alelo conservado (secuencias de *P. infestans* y *P. mirabilis*) (Figura 3.4).

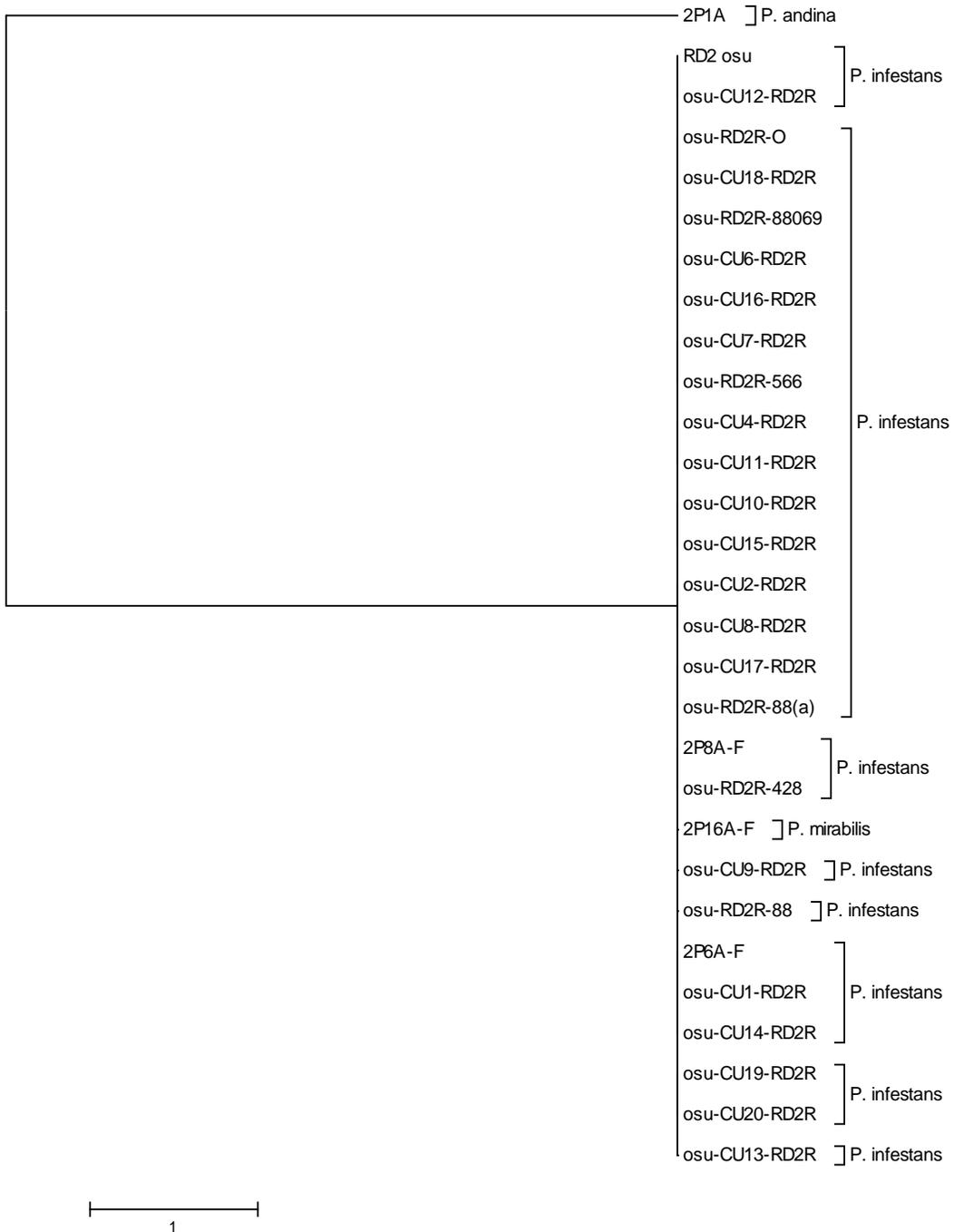


Figura 3.4: Árbol de Evolución mínima para el gen RD2, reconstruido a partir del número de diferencias entre aminoácidos (secuencias Ohio State University OSU, cortesía de Carla Garzón (RD2-OSU) y Sang Keun Oh (OSU-CU)).

En la presente investigación se identificaron los dos alelos del gen *Avr3a* para *P. infestans* descritos por Bos, *et al* (2006). El alelo avirulento AVR3a<sup>KI</sup> (C<sup>19</sup>, K<sup>80</sup>, I<sup>103</sup>) fue identificado en el clon 3P3A (*P. andina*) y los clones de *P. infestans* 3P12A y 3P12B. En el clon 3P6A (*P. infestans*) se encontró el alelo virulento AVR3a<sup>EM</sup> (S<sup>19</sup>, E<sup>80</sup>, M<sup>103</sup>). Dentro de los alelos EM y KI se identificaron discrepancias, como se refleja en el árbol de reconstrucción filogenética, realizado a partir del número de diferencias a nivel de aminoácidos, a través del método de evolución mínima con un valor de 1000 repeticiones de bootstrap. Los valores de bootstrap agrupan con relativa confianza a todas las secuencias KI dentro de un mismo grupo (Figura 3.5).

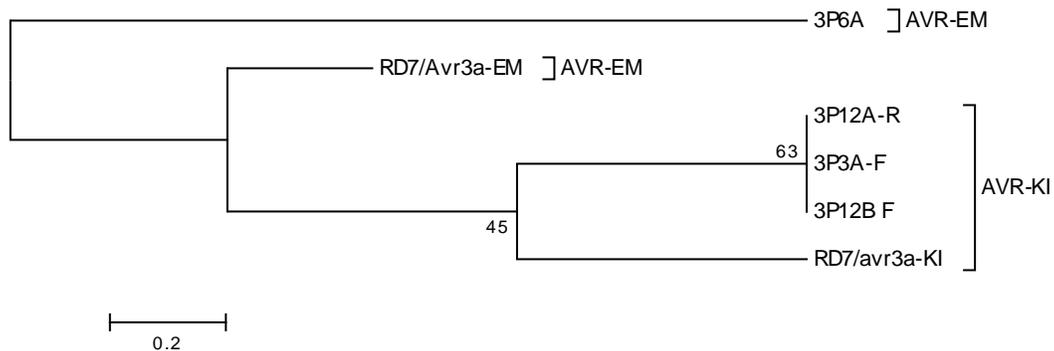


Figura 3.5: Árbol de evolución mínima para el gen *Avr3a*, reconstruido a partir del número de diferencia entre aminoácidos. Se reflejan las diferencias existentes entre los alelos AVR3a-EM (virulento) y AVR3a-KI (avirulento) (secuencias RD7/Avr3a-EM y RD7/avr3a-KI cortesía de Carla Garzón). Los valores de bootstrap son mostrados en los nodos.

De igual forma se determinó la presencia del motivo RXLR, en este caso RLLR, en todas las secuencias. Al realizar el alineamiento de las secuencias a nivel de aminoácidos, se determinó únicamente 3 sitios polimórficos, como lo muestra la figura 3.6.

```

OSU Avr3a-EM -----AP NFNLANLNEE MFNVAALTER [ 60]
OSU avr3a-KI -----..K. [ 60]
3P3A-F MDQTKVLVYG TPAHYIHDSA GRLLLRKNEE NEETSEER.. [ 60]
3P6A MDQTKVLVYG TPAHYIHDSA GRLLLRKNEE NEETSEER.. [ 60]
3P12B_F MDQTKVLVYG TPAHYIHDSA GRLLLRKNEE NEETSEER.. [ 60]
3P12A-R MDQTKVLVYG TPAHYIHDSA GRLLLRKNEE NEETSEER.. [ 60]

OSU Avr3a-EM ADAKKLAKQL MGNDKLADAA YIWWQHNRVT LDQIDTFLKL ASRKTQGAKY NQIYNSYLMH [120]
OSU avr3a-KI .....M.. [120]
3P3A-F ..... [120]
3P6A .....M.. [120]
3P12B_F ..... [120]
3P12A-R ..... [120]

OSU Avr3a-EM LGLTGY* [127]
OSU avr3a-KI ..... [127]
3P3A-F ..... [127]
3P6A ..... [127]
3P12B_F ..... [127]
3P12A-R ..... [127]

```

Figura 3.6: Alineamiento entre proteínas AVR3a, donde se aprecian los alelos AVR3a<sup>EM</sup> (3P6A, OSU Avr3a-EM) (virulentos) y AVR3a<sup>KI</sup> (3P3A, 3P12A, 3P12B y OSU avr3a-KI) (avirulentos). El motivo RLLR se encuentra resaltado en gris (secuencias Ohio State University (OSU) cortesía de Carla Garzón).

A nivel de nucleótidos se observan 3 diferencias entre los alelos AVR3a<sup>KI</sup> (C<sup>19</sup>, K<sup>80</sup>, I<sup>103</sup>) y AVR3a<sup>EM</sup> (S<sup>19</sup>, E<sup>80</sup>, M<sup>103</sup>) (Figura 3.7).

```

OSU_Avr3a-EM AATTTCAATT TGGCGAATCT AAATGAGGAG ATGTTTAATG TGGCTGCCTT GACGGAGAGA [180]
OSU_avr3a-KI .....A..... [180]
3P3A-F .....A..... [180]
3P6A ..... [180]
3P12E_F .....A..... [180]
3P12A-R .....A..... [180]

OSU_Avr3a-EM GCAGATGCCA AAAAGCTAGC GAAACAGCTT ATGGGTAATG ATAAGCTGGC GGATGCTGCA [240]
OSU_avr3a-KI ..... [240]
3P3A-F ..... [240]
3P6A ..... [240]
3P12E_F ..... [240]
3P12A-R ..... [240]

OSU_Avr3a-EM TACATTTGCT GGCAGCACAA CAGGGTTACG CTAGACCAGA TTGACACCTT CCTGAAGCTT [300]
OSU_avr3a-KI ..... [300]
3P3A-F ..... [300]
3P6A .....G..... [300]
3P12E_F ..... [300]
3P12A-R ..... [300]

OSU_Avr3a-EM GCAAGCCGCA AGACGCAAGG CGCAAAGTAC AATCAGATCT ACAATAGCTA CCTGATGCAC [360]
OSU_avr3a-KI .....A..... [360]
3P3A-F ..... [360]
3P6A .....A..... [360]
3P12E_F ..... [360]
3P12A-R ..... [360]

```

Figura 3.7: Alineamiento del gen *Avr3a* a nivel de ADN, donde se aprecian 3 cambios de nucleótidos entre los alelos *AVR3a<sup>EM</sup>* (3P6A) (virulento) y *AVR3a<sup>KI</sup>* (3P3A, 3P12A y 3P12B) (avirulentos)(secuencias Ohio State University OSU cortesía de Carla Garzón).

Del gen RD6 solo se pudo recuperar el clon 6P4C, perteneciente a *P. andina*, el cual fue comparado con 3 secuencias del Ohio State University (OSU) pertenecientes a *P. infestans* (cortesía de Carla Garzón). Del alineamiento de las secuencias antes mencionadas, se lograron identificar 3 alelos, que difieren entre sí en 38 sitios a nivel de ADN (Figura 3.8).

```

RD6_OSU_IPIO_ADN GTTTCATCCA ATCTCAACAC CGCCGTGAAT TACGCTTCCA CATCCAAGAT TCGCTTTCTG [ 60]
RD6_OSU_CU3_ADN -----.....G....G..... [ 60]
RD6_OSU_88069_ADN ..... [ 60]
RD6_6P4C .....G....G..... [ 60]

RD6_OSU_IPIO_ADN TCGACTGAGT ACAACGCCGA TGA AAAAAGA AGCTTGCGAG GTGACTACAA CAATGAGGTC [120]
RD6_OSU_CU3_ADN .....G..... [120]
RD6_OSU_88069_ADN ..... [120]
RD6_6P4C .....G..... [120]

RD6_OSU_IPIO_ADN ACAAAAAGAC CCAACACGTC TGACGAAGAG CGGGCGTTTT CTATCTCAAA GTCTGCCGAA [180]
RD6_OSU_CU3_ADN .....G.....C...T.... [180]
RD6_OSU_88069_ADN ..... [180]
RD6_6P4C .....G.....C...T.... [180]

RD6_OSU_IPIO_ADN TACGTGAAGA TGCTACTTTA TGGATTCAAA CTTGGATTTT CTCCTCGCAC TCAGTCCAAG [240]
RD6_OSU_CU3_ADN A.A.....T...G.T.G...C...A..G A...CT....C..A..... [240]
RD6_OSU_88069_ADN ..... [240]
RD6_6P4C A.A.....T...G.T.G...C...A..G A...CT....C..... [240]

RD6_OSU_IPIO_ADN ACGGTGTTGC GATACGAAGA TAAACTGTTT ACGGCTCTCT ATAAATCCGG AGACACGCCG [300]
RD6_OSU_CU3_ADN .....TA....C..... [300]
RD6_OSU_88069_ADN .....C.....TA... [300]
RD6_6P4C .....TA....C..... [300]

RD6_OSU_IPIO_ADN ACAAGCCTAA GGACCAAGCA TCTCGATAAG GCTTCCGCTA GCGTATTTTT CAACAGATTC [360]
RD6_OSU_CU3_ADN GCT...TAC. A..A.....C.....G..... [360]
RD6_OSU_88069_ADN .T.....G..... [360]
RD6_6P4C GCT...TAC. A..A.....C.....G..... [360]

```

Figura 3.8: Alineamiento del gen RD6 a nivel de ADN, donde se observan 37 sitios de variabilidad a nivel de nucleótidos entre los diferentes alelos encontrados. La secuencia de *P. andina* 6P4C difiere únicamente en un aminoácido con respecto a la secuencia RD6\_OSU\_CU3 (secuencias Ohio State University (OSU) cortesía de Carla Garzón).

Además 22 sitios polimórficos a nivel de aminoácidos fueron identificados, al igual que el motivo RSLR en todas las secuencias (Figura 3.9).

```

RD6_OSU_IPIO_ADN VSSNLNTAVN YASTSKIRFL STEYNADEKR SLRGGDYNNEV TKEPNTSDEE RAFSISKSAE [ 60]
RD6_OSU_CU3_ADN -----G. D..... [ 60]
RD6_OSU_88069_ADN ..... [ 60]
6P4C .....G. D.....A... Y.V. [ 60]

RD6_OSU_IPIO_ADN YVKMVLGFK LGFSPRTQSK TVLRYEDKLF TALYKSGETP RSLRTKHLDK ASASVFFNRF [120]
RD6_OSU_CU3_ADN K..LG..AL. IA..... Y.H..... A.YKN..P.. ..G..... [120]
RD6_OSU_88069_ADN .....L.... I.....G..... [120]
6P4C K..LG..AL. IA..... Y.H..... A.YKN..P.. ..G..... [120]

RD6_OSU_IPIO_ADN KNYVDKNVGP S*LVPVKSVH L*HRI [145]
RD6_OSU_CU3_ADN ..... [145]
RD6_OSU_88069_ADN ..... [145]
6P4C ..... [145]

```

Figura 3.9: Alineamiento a nivel de aminoácidos del gen RD6, donde se observan 22 sitios de variabilidad. El motivo RSLR se resalta en gris (secuencias Ohio State University OSU cortesía de Carla Garzón).

El clon 6P4C perteneciente a *P. andina* resultó ser idéntico a nivel de aminoácidos con la secuencia RD6\_OSU\_CU3 de *P. infestans* y mantiene una sola diferencia sinónima a nivel de ADN. El árbol de evolución mínima reconstruido respalda esta observación, con un valor bootstrap de 100 en unión de estas dos secuencias (Figura 3.10).

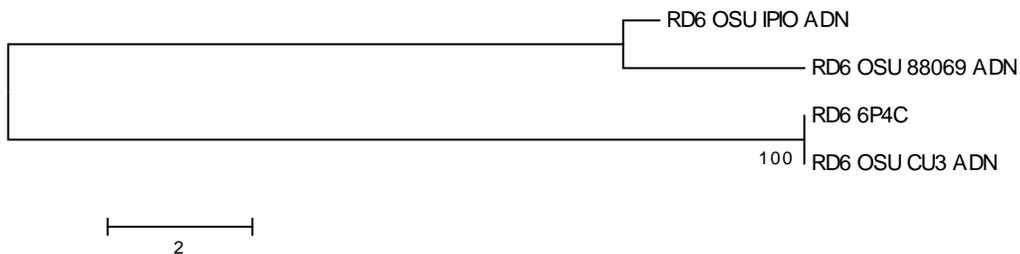


Figura 3.10: Árbol de evolución mínima para el gen RD6, reconstruido a partir del número de diferencias entre aminoácidos (secuencias Ohio State University (OSU) cortesía de Carla Garzón). Los valores de bootstrap calculados son mostrados en cada nodo.

El gen RD39/40 resultó ser altamente polimórfico, ya que en 13 secuencias obtenidas se registran 6 alelos diferentes. El árbol de evolución mínima

reconstruido de estos datos, obtuvo valores altos de bootstrap (entre 86 y 100) para determinar las secuencias dentro de cada grupo (Figura 3.11). Comparando con 120 secuencias RD39/40 de *P. infestans* enviadas desde la Ohio State University (OSU) (cortesía de Sophien Kamoun y Carla Garzón), se pudo determinar que *P. andina* presenta un alelo único en los clones 40P5C y 40P5D de entre 66 alelos encontrados en un total de 114 secuencias analizadas. El alineamiento realizado a las secuencias RD39/40 utilizadas dio un total de 101 sitios de variabilidad con respecto a ADN y 53 sitios con respecto a aminoácidos. No se pudo encontrar el motivo RXLR en las secuencias ecuatorianas, pero fue identificado el motivo RSLR en las secuencias donadas por OSU (lo que indica que este gen si posee el motivo). Al parecer esto se debe a una amplificación incompleta el momento de copiar el gen del patógeno a través de PCR o el momento de digerir e insertar la secuencia en el plásmido, ya que se puede ver el sitio de restricción justo después de la posición correspondiente al motivo. Cabe resaltar que el gen RD39/40 fue el único en presentar una diversidad alélica tan importante como para mostrar distintos alelos en un mismo aislado. Como es el caso de los aislados: 40P5 (alelos 40P5A y 40P5C) y 40P11 (alelos 40P11A, 40P11B-R, 40P11C-F, 40P11C-R, 40P11D).

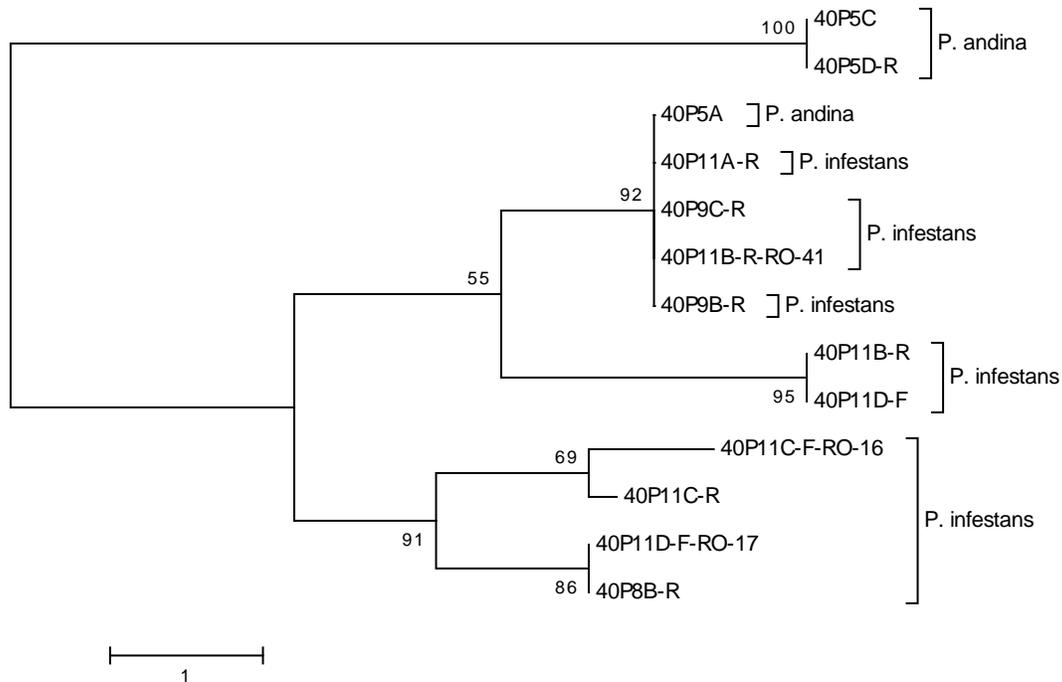


Figura 3.11: Árbol de evolución mínima para el gen RD39/40, reconstruido a partir del número de diferencias entre aminoácidos. Se puede apreciar 6 alelos diferentes identificados en una muestra de 13 aislados clonados en Ecuador. Se muestran los valores de bootstrap en cada nodo.

### 3.3.2 Análisis de neutralidad

El análisis de neutralidad realizado en el programa codeml del paquete PAML 3.14, a través del método de máxima probabilidad (ML) (Maximum Likelihood, siglas en inglés), fue utilizado para detectar si existe selección positiva actuando sobre los distintos genes analizados, y si es el caso, identificar específicamente sobre que sitios de la proteína se encuentra actuando dicha selección. Tres pares de modelos de sustitución a nivel de codones fueron utilizados: M0 vs M3, M1 vs M2 y M7 vs M8.

El gen RD2 no presenta diferencias significativas entre los modelos alternativos M2, M3, M8 (de selección positiva) y los modelos de neutralidad M0,

M1, M7, una vez realizada la prueba de razón de probabilidad (LRT) (Likelihood-ratio Test, siglas en inglés), lo que significa que las dos clases de modelos se comportan de la misma forma, es decir, entre neutros y deletéreos ( $0 < w < 1$ ), mas no como alternativos ( $w > 1$ , selección positiva). Se identificaron únicamente 3 aminoácidos bajo selección positiva en el modelo M8 ( $\beta + w$ ) con una probabilidad posterior de entre el 50 y el 52%, los cuales no son tomados en cuenta ya que el modelo M8 no describe correctamente a los datos (Tabla 3.3).

Tabla 3.3: Análisis de máxima probabilidad (ML) para el gen RD2, donde se muestran los valores obtenidos en la prueba de índice de probabilidad (LRT) para cada modelo, y la comparación del parámetro  $2\Delta \ln L$  con la distribución de chi-cuadrado ( $\chi^2$ ), para determinar si el gen se encuentra descrito por un modelo nulo o un modelo alternativo. Se muestran también los sitios de selección positiva obtenidos en cada modelo.

<i>Modelo</i> <sup>a</sup>	<i>lnL</i> <sup>b</sup>	<i>Comparación de modelos</i>	$2\Delta \ln L$ <sup>c</sup>	<i>Valor crítico de <math>\chi^2</math> (1%)</i>	<i>Grados de libertad</i>	<i>Sitios bajo selección positiva</i> <sup>d</sup>
M0: un radio	-417.587821					No presenta
M1: neutral	-417.587821					No presenta
M2: selección	-417.587821	M1 vs M2	0 <sup>NS</sup>	9.21	2	Ninguno
M3: discreto	-417.587821	M0 vs M3	0 <sup>NS</sup>	13.28	4	Ninguno

<b>M7: beta</b>	<b>-417.588106</b>					<b>No presenta</b>
<b>M8:beta+w</b>	<b>-417.588107</b>	<b>M7 vs M8</b>	<b>-1x10<sup>-6</sup> NS</b>	<b>9.21</b>	<b>2</b>	<b>40 T, 47 L</b>
						<b>83 L</b>

<sup>a</sup> Modelos nulos (M0, M1, M7), modelos alternativos (M2, M3, M8) (determinan selección positiva).

<sup>b</sup> Logaritmo del índice de probabilidad.

<sup>c</sup> Doble de la diferencia en el logaritmo del índice de probabilidad entre un modelo nulo y el modelo alternativo, comparado a la distribución de chi-cuadrado.

<sup>d</sup> Sitios bajo selección positiva con probabilidades del 50 al 52%

El gen *Avr3a* no presentó diferencias significativas entre los modelos alternativos y los modelos de neutralidad una vez realizada la LRT, aunque a nivel de residuos específicos, los modelos alternativos M2, M3 y M8 presentaron 3 sitios de selección positiva, con una probabilidad posterior de entre el 95 y 99% (Tabla 3.4). Estos datos indicarían que la acción de la presión selectiva es ejercida sobre ciertos lugares específicos en la secuencia pero no sobre la proteína en su totalidad.

Tabla 3.4: Análisis de máxima probabilidad (ML) para el gen *Avr3a*, donde se muestran los valores obtenidos en la prueba de índice de probabilidad (LRT) para cada modelo, y la comparación del parámetro  $2\Delta\ln L$  con la distribución de chi-cuadrado ( $\chi^2$ ), para determinar si el gen se encuentra descrito por un modelo nulo o un modelo alternativo. Se muestran también los sitios de selección positiva obtenidos en cada modelo.

<b>Modelo<sup>a</sup></b>	<b><math>\ln L</math><sup>b</sup></b>	<b>Comparación de modelos</b>	<b><math>2\Delta L</math><sup>c</sup></b>	<b>Valor crítico de <math>\chi^2</math> (1%)</b>	<b>Grados de libertad</b>	<b>Sitios bajo selección positiva<sup>d</sup></b>
<b>M0: un radio</b>	<b>-487.330777</b>					<b>No presenta</b>

<b>M1:neutral</b>	<b>-488.204354</b>					<b>No presenta</b>
<b>M2: selección</b>	<b>-485.678835</b>	<b>M1 vs M2</b>	<b>5.05</b> <sup>NS</sup>	<b>9.21</b>	<b>2</b>	59 E, 82 I, <u>118 L</u>
<b>M3: discreto</b>	<b>-485.678604</b>	<b>M0 vs M3</b>	<b>3.30</b> <sup>NS</sup>	<b>13.28</b>	<b>4</b>	<u>59 E, 82 I, 118 L</u>
<b>M7: beta</b>	<b>-488.221479</b>					<b>No presenta</b>
<b>M8:beta+w</b>	<b>-485.678743</b>	<b>M7 vs M8</b>	<b>5.08</b> <sup>NS</sup>	<b>9.21</b>	<b>2</b>	<u>59 E, 82 I, 118 L</u>

<sup>a</sup> Modelos nulos (M0, M1, M7), modelos alternativos (M2, M3, M8) (determinan selección positiva).

<sup>b</sup> Logaritmo del índice de probabilidad.

<sup>c</sup> Doble de la diferencia en el logaritmo del índice de probabilidad entre un modelo nulo y el modelo alternativo, comparado a la distribución de chi-cuadrado.

<sup>d</sup> Aminoácidos en negrilla presentan probabilidad >95%, aminoácidos subrayados y en negrilla presentan probabilidad >99%

El gen RD6 reflejó diferencias significativas entre los modelos alternativos y los modelos de neutralidad en la prueba de razón de probabilidad LRT. En la comparación de los modelos M0 vs M3 la LRT ( $2 \Delta \ln L = 2 \times [-672.509607 - (-682.499901)] = 19.99$ , resultó mayor al valor crítico de chi-cuadrado (13.28 al 1% de significación con 4 grados de libertad). En la comparación de los modelos M1 vs M2, el valor de LRT ( $2 \Delta \ln L$ ) fue igual a 14.60, cifra mayor al valor crítico de chi-cuadrado (9.21 al 1% de significación con 2 grados de libertad). En la comparación de los modelos M7 vs M8, el valor de LRT ( $2 \Delta \ln L$ ) fue igual a 14.70, cifra mayor al valor crítico de chi-cuadrado (9.21 al 1% de significación con 2 grados de libertad). Esto quiere decir que los modelos alternativos ( $w > 1$ ) describen mejor a las secuencias, por ende se habla de alelos bajo selección positiva. También fueron

identificados 7 sitios bajo selección positiva en los 3 modelos alternativos, con probabilidad posterior de entre el 50 y 99% (Tabla 3.5).

Tabla 3.5: Análisis de máxima probabilidad (ML) para el gen RD6, donde se muestran los valores obtenidos en la prueba de índice de probabilidad (LRT) para cada modelo, y la comparación del parámetro  $2\Delta \ln L$  con la distribución de chi-cuadrado ( $\chi^2$ ), para determinar si el gen se encuentra descrito por un modelo nulo o un modelo alternativo. Se muestran también los sitios de selección positiva obtenidos en cada modelo.

---

<i>Modelo</i> <sup>a</sup>	<i>lnL</i> <sup>b</sup>	<i>Comparación de modelos</i>	$2\Delta L$ <sup>c</sup>	<i>Valor crítico de <math>\chi^2</math> (1%)</i>	<i>Grados de libertad</i>	<i>Sitios bajo selección positiva</i> <sup>d</sup>
----------------------------	-------------------------	-------------------------------	--------------------------	--	---------------------------	--

<b>M0: un radio</b>	<b>-682.49</b>					<b>No presenta</b>
<b>M1: neutral</b>	<b>-679.80</b>					<b>No presenta</b>
<b>M2: selección</b>	<b>-672.50</b>	<b>M1 vs M2</b>	<b>14.60<sup>**</sup></b>	<b>9.21</b>	<b>2</b>	<b>61 Y , 66 L, 72 G, 92 A, 96 S, <u>101 R</u>, <u>103 L</u></b>
<b>M3: discreto</b>	<b>-672.50</b>	<b>M0 vs M3</b>	<b>19.99<sup>**</sup></b>	<b>13.28</b>	<b>4</b>	<b>61 Y , 66 L, 72 G, 92 A, 96 S, <u>101 R</u>, <u>103 L</u></b>
<b>M7: beta</b>	<b>-679.78</b>					<b>No presenta</b>
<b>M8: beta+w</b>	<b>-672.43</b>	<b>M7 vs M8</b>	<b>14.70<sup>**</sup></b>	<b>9.21</b>	<b>2</b>	<b>61 Y , 66 L, 72 G, 92 A, 96 S, <u>101 R</u>, 103 L</b>

<sup>a</sup> Modelos nulos (M0, M1, M7), modelos alternativos (M2, M3, M8) (determinan selección positiva).

<sup>b</sup> Logaritmo del índice de probabilidad.

<sup>c</sup> Doble de la diferencia en el logaritmo del índice de probabilidad entre un modelo nulo y el modelo alternativo, comparado a la distribución de chi-cuadrado.

<sup>d</sup> Aminoácidos no marcados presentan probabilidades entre el 50% al 90%, aminoácidos en negrilla presentan probabilidad >95%, aminoácidos subrayados y en negrilla presentan probabilidad >99%

Al igual que en el caso anterior, el gen RD39/40 presentó diferencias significativas entre los modelos alternativos y los modelos de neutralidad en la prueba de razón de probabilidad LRT. En la comparación de los modelos M0 vs M3 la LRT ( $2 \Delta \ln L$ ) =  $2x[-832.677462 - (-855.127135)] = 44.90$ , resultó mayor al valor crítico de chi-cuadrado (13.28 al 1% de significación con 4 grados de libertad). En la comparación de los modelos M1 vs M2, el valor de LRT ( $2 \Delta \ln L$ ) fue igual a

41.88, cifra mayor al valor crítico de chi-cuadrado (9.21 al 1% de significación con 2 grados de libertad). En la comparación de los modelos M7 vs M8, el valor de LRT ( $2\Delta\ln L$ ) fue igual a 42.50, cifra mayor al valor crítico de chi-cuadrado (9.21 al 1% de significación con 2 grados de libertad). Por lo tanto los modelos de neutralidad no describen a estas secuencias y los modelos alternativos ( $w>1$ ) resultan más adecuados. Fueron identificados 14 sitios bajo selección positiva en los 3 modelos alternativos, con probabilidad posterior de entre el 50 y 99% (Tabla 3.6).

Tabla 3.6: Análisis de máxima probabilidad (ML) para el gen RD39/40, donde se muestran los valores obtenidos en la prueba de índice de probabilidad (LRT) para cada modelo, y la comparación del parámetro  $2\Delta\ln L$  con la distribución de chi-

cuadrado ( $\chi^2$ ), para determinar si el gen se encuentra descrito por un modelo nulo o un modelo alternativo. Se muestran también los sitios de selección positiva obtenidos en cada modelo.

<i>Modelo</i> <sup>a</sup>	<i>lnL</i> <sup>b</sup>	<i>Comparación de modelos</i>	$2\Delta L$ <sup>c</sup>	<i>Valor crítico de <math>\chi^2</math> (1%)</i>	<i>Grados de libertad</i>	<i>Sitios bajo selección positiva</i> <sup>d</sup>
<b>M0: un radio</b>	<b>-855.12</b>					<b>No presenta</b>
<b>M1:neutral</b>	<b>-853.61</b>					<b>No presenta</b>
<b>M2: selección</b>	<b>-832.67</b>	<b>M1 vs M2</b>	<b>41.88**</b>	<b>9.21</b>	<b>2</b>	<u>20 D</u> , <u>32 G</u> , <u>34 T</u> , <u>39 V</u> , <b>56 F</b> , <b>60 V</b> , <u>61 E</u> , <b>62 P</b> , 69 K, <b>70 K</b> , <b>71 P</b> , <u>76 K</u> , 79 Y, 87 S
<b>M3: discreto</b>	<b>-832.67</b>	<b>M0 vs M3</b>	<b>44.90**</b>	<b>13.28</b>	<b>4</b>	<u>20 D</u> , <u>32 G</u> , <u>34 T</u> , <u>39 V</u> , <b>56 F</b> , <b>60 V</b> , <u>61 E</u> , <b>62 P</b> , 69 K, <b>70 K</b> , <b>71 P</b> , <u>76 K</u> , 79 Y, 87 S
<b>M7: beta</b>	<b>-853.95</b>					<b>No presenta</b>
<b>M8:beta+w</b>	<b>-832.67</b>	<b>M7 vs M8</b>	<b>42.54**</b>	<b>9.21</b>	<b>2</b>	<u>20 D</u> , <u>32 G</u> , <u>34 T</u> , <u>39 V</u> , <b>56 F</b> , <b>60 V</b> , <u>61 E</u> , <b>62 P</b> , 69 K, <b>70 K</b> , <b>71 P</b> , <u>76 K</u> , 79 Y, 87 S

<sup>a</sup> Modelos nulos (M0, M1, M7), modelos alternativos (M2, M3, M8) (determinan selección positiva).

<sup>b</sup> Logaritmo del índice de probabilidad.

<sup>c</sup> Doble de la diferencia en el logaritmo del índice de probabilidad entre un modelo nulo y el modelo alternativo, comparado a la distribución de chi-cuadrado.

<sup>d</sup> Aminoácidos no marcados presentan probabilidades entre el 50% al 90%, aminoácidos en negrilla presentan probabilidad >95%, aminoácidos subrayados y en negrilla presentan probabilidad >99%

### 3.3.3 Estudio de proteínas

Para el estudio de proteínas se tomó en cuenta únicamente a las secuencias representativas dentro de cada gen, esto incluye a las secuencias OSU que contienen mayor información y a todas las secuencias ecuatorianas que representan a un alelo en particular.

El estudio de la estructura primaria de las proteínas se enfocó en el tamaño de las secuencias, el peso molecular en kilodaltons y el punto isoeléctrico. Estimaciones que fueron hechas a través del programa Protéine (2007) de la Universidad de Provence en Marsella Francia, disponible libremente en Internet (<http://www.iut-arles.up.univ-mrs.fr/w3bb/dabim/compo-p.html>). Los resultados obtenidos de este análisis, no reflejan mayores diferencias entre secuencias del mismo gen, a excepción de 2P1A (que registra un leve incremento en el punto isoeléctrico con relación al resto de secuencias RD2) y 3P6A que manifiesta una inusual variación en el punto isoeléctrico con relación a secuencias del mismo gen y del mismo peso (7.88 vs 9.51 del resto de secuencias) (Tabla 3.7).

Tabla 3.7: Punto isoeléctrico, peso molecular y número de aminoácidos de las secuencias representativas, obtenidos a partir de la estructura primaria proteica.

<b>Secuencias</b>	<b>Número de aminoácidos</b>	<b>Peso (Kdal)</b>	<b>Punto isoeléctrico</b>
2P1A	103	11.17	10.16
2P6A	104	11.35	9.81
2P8A	104	11.35	9.81

2P16A	104	11.35	9.81
OSU-RD7/Avr3a-EM ADN	88	10.06	9.67
OSU-RD7/avr3a-KI ADN	88	10.08	9.83
3P3A	129	14.83	9.51
<b>Secuencias</b>	<b>Número de aminoácidos</b>	<b>Peso (Kdal)</b>	<b>Punto isoelectrico</b>
3P6A	128	14.72	7.88
3P12A	129	14.89	9.61
3P12B	129	14.87	9.51
RD6_OSU IPIO ADN	143	16.40	10.08
RD6_OSU CU3 ADN	137	15.56	9.93
RD6_OSU 88069 ADN	143	16.37	9.98
6P4C	145	16.38	9.93
40P5A	79	87.16	10.25
40P9B-R	80	89.15	10.31
40P9C-R	80	89.04	10.31
40P11A-R	80	88.72	10.31
40P11B-R-RO-41	77	85.54	10.25
40P5C	79	89.12	10.69
40P5D-R	80	90.80	10.69
40P11C-R	80	88.73	10.34
40P11D-F-RO-17	77	85.44	10.13
40P8B-R	77	85.44	10.13
41P11B-R	77	86.15	10.09
40P11D-F	77	86.15	10.09
40P11C-F-RO-16	77	85.44	10.17

El estudio de estructura secundaria realizado a través del programa LOOP versión 3.0 @ CBSU (2007) de la Universidad de Cornell, libre en Internet

(<http://cbsuapps.tc.cornell.edu/loopp.aspx>), proporcionó información sobre el plegamiento que puede tener cada una de las secuencias de aminoácidos analizadas.

Con respecto al gen RD2 las secuencias analizadas fueron 2P1A (*P. andina*), 2P8A (*P. infestans*) y 2P16A (*P. mirabilis*) y dado que solamente la secuencia 2P1A presenta diferencias con relación al resto de secuencias RD2, el análisis se redujo a revisar las secuencias 2P1A y 2P8A (Figura 3.12).

```

2P1A
CCCCCHHHHCCCCCHHHHHHHCCCCCCCCCCCCCCCCCCCCCHHHCCCCCCCCCCCC
2P8A
CCCCCHHHHCCCCCHHHHHHHCCCCCCCCCCCCCCCCCCCCCHHHCCCCCCCCCCCC
2P1A
CCCCCCCCCCCCCHHHHHHCCCCCHHHHHHHHHHHHHCHHHHH
2P8A
CCCCCCCCCCCCCHHHHHHHHHHHCHHHHHHHHHHHCHHHHH

```

Figura 3.12: Posibles diferencias de estructura secundaria entre las secuencias 2P1A (*P. andina*) y 2P8A (*P. infestans*), alelos representativos del gen RD2. C (enrollamiento), H ( $\alpha$  - hélice).

El programa produjo una conformación secundaria ligeramente diferente entre las dos secuencias (4% más  $\alpha$  - hélices en 2P8A que en 2P1A y 4% menos enrollamientos) (Tabla 3.8).

Tabla 3.8: Porcentaje de estructuras secundarias en el gen RD2 calculado para las secuencias 2P1A (*P. andina*) y 2P8A (*P. infestans*).

Tipo de Estructura	Porcentaje de Estructuras en la Secuencia (%)	
	2P1A	2P8A
$\alpha$ - hélices	36.63	40.59
$\beta$ - láminas	0	0

otros plegamientos	63.36	59.40
--------------------	-------	-------

Cabe aclarar que cada uno de los aminoácidos identificados con cada una de las estructuras (enrollamiento,  $\alpha$  -hélice o  $\beta$  -lámina) son únicamente probabilidades y no se refiere a la forma real de las proteínas, sino más bien a una aproximación (Figura 3.13). Por otro lado se registraron 3 cambios conservativos de aminoácidos (M por L en el sitio 12, W por L en el sitio 47, K por R en el sitio 74), esto quiere decir que el cambio se produjo entre aminoácidos del mismo grupo (tomando en cuenta los cuatro grupos de aminoácidos: Apolar, polar, catiónico y aniónico) y 5 cambios radicales (G por D en 31, P por T en 40, G por D en 55, N por D en 64, H por L en 83), donde se produce el cambio entre aminoácidos de diferentes grupos (por ejemplo un aminoácido de tipo polar por uno de tipo catiónico).

```

> 1 60
LSTTTGVQAANMVGPAQRLLRKHYTAAENDGDSEARALNPEKMKTMWKAGMTVDGYAAKL
CCCCCHHHHCCCCCHHHHHHCCCCCCCCCCCCCCCCCCCCCHHHCCCCCCCCCCCCCCC
65666544445677757776654545567886544456875455445566777776655

> 61 101
KLTNKIAAAANSKAMEKLGETHKMKLLRYLNYVAEHTAV
CCCCCCCCCCCCCHHHHHHCCCCCHHHHHHHHHHHHCHHHHH
55666656667655677655544467777776554554434

```

Figura 3.13: Estructura secundaria probable de la secuencia 2P1A (gen RD2). En la primera fila se presenta la secuencia de aminoácidos, en la segunda fila el tipo de estructura con que se identifica cada aminoácido H ( $\alpha$  - hélice), C (enrollamiento), E ( $\beta$  - lámina), en la tercera fila se indica el nivel de confianza de la segunda fila (escala de 3 a 9, correspondiente a p=0.3 confianza baja y p=0.9 confianza alta, respectivamente).

Con respecto a la proteína AVR3a, fueron tomadas en cuenta las secuencias 3P6A (*P. infestans*) que representa al alelo virulento AVR3a<sup>EM</sup> (S<sup>19</sup>,

E<sup>80</sup>, M<sup>103</sup>) y 3P3A (*P. andina*) que representa al alelo avirulento AVR3a<sup>KI</sup> (C<sup>19</sup>, K<sup>80</sup>, I<sup>103</sup>) (Figura 3.14).

```

3P6A
CCCCCCCCCCCCCCCCCHHHCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHCHHHHHHHHH

3P3A
CCCCCCCCCCCCCCCCCHHCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHCHHHHHHHHH

3P6A
HCCCCHHHHHHHHHHHHHCCCCCHHHHHHHHHHHCCCCCCCCCHHHHHHHHHHHHHCCCCC

3P3A
HCCCCHHHHHHHHHHHHHCCCCCHHHHHHHHHHHCCCCCCCCCHHHHHHHHHHHHHCCCCC

```

Figura 3.14: Posibles diferencias de estructura secundaria entre las secuencias 3P6A (alelo virulento AVR3a<sup>EM</sup>) y 3P3A (alelo avirulento AVR3a<sup>KI</sup>), del gen AVR3a. C (enrollamiento), H ( $\alpha$  - hélice), E ( $\beta$  - lámina).

La estructura secundaria de estas dos secuencias es prácticamente la misma de acuerdo al análisis bioinformático realizado (Figura 3.15). Solamente un cambio radical de aminoácidos se presenta entre estos alelos (entre ácido glutámico (E) y lisina (K)), asimismo 2 cambios conservativos fueron encontrados.

```

> 1 60
MDQTKVLVYGTPAHYIHDSAGRLLLRKNEENEETSEERAPNFNLANLNEEMFNVAALTER
CCCEEEEEEEEECCCCCCCCCCCCCHHHCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHH
65457777678754444455544434678776665557777666676456778888876

> 61 120
ADAKKLAKQLMGNDKLDAAAYMWWQHNRVTLTDQIDTFLKLSRKTQGAKYNQIYNSYMMH
CCHHHHHHHHHHCCCHHHHHHHHHHHHCCCCCHHHHHHHHHHHCCCCCCCCCHHHHHHHHHHHHH
77799999986785688999999875787788999999875776777645666665544

> 121 126
LGLTGY
HCCCCC
344566

```

Figura 3.15: Estructura secundaria probable de la secuencia 3P6A (gen *Avr3a*) correspondiente al alelo virulento AVR3a<sup>EM</sup>. En la primera fila se presenta la secuencia de aminoácidos, en la segunda fila el tipo de estructura con que se



En la tabla 3.9 se analizan los porcentajes de estructuras secundarias probables entre cada secuencia (secuencias OSU presentan entre 12.8% y 14.2% más  $\alpha$  - hélices que 6P4C y 13.73% menos plegamientos, el número de  $\beta$  - láminas se mantiene prácticamente constante).

Tabla 3.9: Porcentaje de estructuras secundarias en el gen RD6 calculado para las secuencias 6P4C (*P. andina*), RD6\_OSU 88069 (*P. infestans*) y RD6\_OSU IPIO (*P. infestans*).

Tipo de Estructura	Porcentaje de Estructuras en la Secuencia (%)		
	6P4C	RD6_OSU 88069	RD6_OSU IPIO
$\alpha$ - hélices	7.63	22.13	20.61
$\beta$ - láminas	70.29	69.46	70.99
otros plegamientos	22.13	8.39	8.39

En la tabla 3.10 se puede observar el número de sustituciones radicales y conservativas entre las secuencias analizadas, 12 cambios radicales de aminoácidos y 9 conservativos se registran entre las secuencias RD6\_OSU IPIO y 6P4C, 13 radicales y 8 conservativos entre las secuencias RD6\_OSU 88069 y 6P4C y solamente 3 radicales entre las dos secuencias RD6\_OSU. Estos datos mantienen la divergencia entre la secuencia 6P4C y las secuencias RD6\_OSU y apoyan las diferencias de conformación secundaria predichas en el análisis bioinformático.

Tabla 3.10: Número de sustituciones radicales y conservativas entre los aminoácidos de las secuencias 6P4C (*P. andina*), RD6\_OSU 88069 (*P. infestans*) y RD6\_OSU IPIO (*P. infestans*) correspondientes al gen RD6.

Secuencias comparadas	Tipo de Sustitución	
	Conservativas	Radicales

6P4C vs RD6_OSU 88069	9	<b>13</b>
RD6_OSU 88069 vs RD6_OSU IPIO	0	<b>3</b>
<b>6P4C vs RD6_OSU IPIO</b>	<b>8</b>	<b>12</b>

Con respecto al gen RD39/40, se utilizó un aislado de cada alelo encontrado en los clones ecuatorianos para determinar su estructura secundaria. Estos fueron: 40P11A-R (*P. infestans*), 40P5C (*P. andina*), 40P11C-R (*P. infestans*), 40P8B (*P. infestans*), 40P11D (*P. infestans*) y 40P11C-F (*P. infestans*) (Figura 3.17).

```

40P11A
CCHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCC

40P5C
CCHHHHCCCCCCCCCCCCCHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHCCCCCCCCCHHHHCCCCCCHHHHCCC

40P11C_R
CCHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCC

40P8B
CCHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHCCCCCCCCCHHCCCCCCCCCCCC

40P11D
CCHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCC

40P11C_F
CCHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHCCCCCCCCCHHCCCCCCCCCCCC

```

Figura 3.17: Posibles diferencias de estructura secundaria entre las secuencias 40P11A-R (*P. infestans*), 40P5C (*P. andina*), 40P11C-R (*P. infestans*), 40P8B (*P. infestans*), 40P11D (*P. infestans*) y 40P11C-F (*P. infestans*), alelos representativos para el gen RD39/40. C (enrollamiento), H ( $\alpha$  - hélice).

La secuencia correspondiente al alelo único de *P. andina* 40P5C, presenta una estructura secundaria visiblemente diferente de los otros alelos analizados correspondientes a *P. infestans*. Esto se refleja también en el número de cambios entre aminoácidos (cambios radicales especialmente), ya que existen de 9 a 11 sustituciones radicales de aminoácidos entre el alelo 40P5C de *P. andina* y los

alelos provenientes de *P. infestans*, y solamente de 0 a 4 sustituciones radicales entre los alelos de *P. infestans* comparados entre sí (Tabla 3.11).

Tabla 3.11: Número de sustituciones radicales y conservativas entre los aminoácidos de las secuencias 40P11A-R (*P. infestans*), 40P5C (*P. andina*), 40P11C-R (*P. infestans*), 40P8B (*P. infestans*), 40P11D (*P. infestans*) y 40P11C-F (*P. infestans*), pertenecientes al gen RD39/40.

Secuencias comparadas	Tipo de Sustitución	
	Conservativas	Radicales
40P11A vs 40P5C	6	11
40P11A vs 40P11C-R	3	1
40P11A vs 40P8B	3	2
40P11A vs 40P11D	2	3
40P11A vs 40P11C-F	4	3
40P5C vs 40P11C-R	7	10
40P5C vs 40P8B	7	10
40P5C vs 40P11D	6	11
40P5C vs 40P11C-F	8	9
40P11C-R vs 40P8B	3	1
40P11C-R vs 40P11D	4	4
40P11C-R vs 40P11C-F	1	1
40P8B vs 40P11D	4	3
40P8B vs 40P11C-F	2	0
40P11D vs 40P11C-F	5	3

Al analizar los porcentajes de posibles estructuras secundarias, se registran 3 grupos de secuencias. El primero grupo conformado por los alelos 40P11A y 40P11C-R, los cuales no presentan diferencias entre ellos, el segundo grupo formado por los alelos 40P8B, 40P11D y 40P11C-F, y el tercer grupo, en donde se

encuentra únicamente el alelo de *P. andina* 40P5C, que no coincide con ninguno de los dos grupos antes mencionados (Tabla 3.12).

Tabla 3.12: Porcentaje de estructuras secundarias para el gen RD39/40, calculado para las secuencias 40P11A-R (*P. infestans*), 40P5C (*P. andina*), 40P11C-R (*P. infestans*), 40P8B (*P. infestans*), 40P11D (*P. infestans*) y 40P11C-F (*P. infestans*).

Tipo de Estructura	Porcentaje de Estructuras en la Secuencia (%)					
	40P11A	40P5C	40P11C-R	40P8B	40P11D	40P11C-F
$\alpha$ - hélices	14.28	38.96	14.28	24.67	24.67	24.67
$\beta$ - láminas	0	0	0	0	0	0
otros plegamientos	85.71	61.03	85.71	75.32	75.32	75.32

## CAPÍTULO IV: DISCUSIÓN

A través de este estudio se ha logrado determinar la variabilidad alélica en los distintos genes de la familia de efectores RXLR analizados, así como pistas para determinar su evolución molecular. Algunos de los hallazgos resultan sumamente importantes en el estudio de *P. andina*, ya que esta especie hace poco descubierta empieza a ser entendida y reconocida (Ordoñez *et al.*, 2000; Wattier *et al.*, 2003; Adler *et al.*, 2004; y Oliva *et al.*, 2007). Dado que es el primer trabajo de este tipo, realizado en especies ecuatorianas, los datos obtenidos apuntan a que existe una diversidad importante en el país. Las reconstrucciones filogenéticas y los análisis de estructura primaria y secundaria, ayudaron a determinar la conservación y la divergencia entre los alelos encontrados en cada especie. *P. andina* se distingue claramente de *P. infestans* en los genes RD2 y RD39/40, pero al mismo tiempo se encuentran cercanamente relacionadas en los genes RD6 y *Avr3a*. Por su parte el análisis de máxima probabilidad sirvió para

determinar la evolución molecular de los genes estudiados y ayudó a determinar el tipo de selección bajo el que se encuentra cada gen.

#### 4.1 Diversidad alélica y evolución molecular encontrada en la familia de efectores RXLR de especies del género *Phytophthora* que atacan a cultivos andinos

##### 4.1.1 Gen RD2

RD2 fue el único gen que pudo ser obtenido de las especies *P. andina*, *P. infestans* y *P. mirabilis*. Se mostró que este gen se encuentra sumamente conservado en la especie *P. infestans* (Figura 3.4). La secuencia 2P16A perteneciente a *P. mirabilis* resultó idéntica a las 27 secuencias analizadas de *P. infestans*. Se pudo identificar un alelo único para *P. andina*, a través de la secuencia 2P1A, dicho alelo fue diferenciado de entre 29 secuencias (27 de *P. infestans* + una de *P. mirabilis*). Estos hallazgos son congruentes con los obtenidos por Garzón (2007), en donde se señala la existencia de 2 copias polimórficas para *P. infestans*, 3 para *P. mirabilis* y 2 para *P. andina* (varias de las cuales se encontraron en las muestras ecuatorianas).

A pesar de haber encontrado el motivo RXLR en todas las secuencias (lo que ubica a este gen dentro de la familia de elicitores RXLR), hasta el momento no se ha podido encontrar información sobre la existencia de un gen de resistencia específico en alguna especie de solanacea que interactúe directamente con el gen RD2 para producir respuesta de hipersensibilidad. Según las observaciones realizadas en el departamento de patología vegetal de OSU, este gen produce respuesta de hipersensibilidad en plantas de *Nicotiana benthamiana* sin ningún gen de resistencia presente. Al parecer se trata de un elicitador general que no necesita de una proteína específica de reconocimiento para producir HR (Sophien Kamoun, comentarios personales).

El análisis de estructura secundaria muestra cierta diferencia entre el alelo de *P. andina* 2P1A y los alelos conservados de *P. infestans* (Tabla 3.8 y Figura 3.12). De igual forma existe una leve diferencia en el punto isoeléctrico del alelo 2P1A con relación al resto de secuencias estudiadas (Tabla 3.7), esto sugiere una posible diferenciación estructural entre los alelos. Las consecuencias de estas posibles diferencias podrían ser evaluadas a través de análisis de funcionalidad, ya que se han visto casos en los que la función de una proteína tiene un cambio significativo por el cambio estructural (Nespoulous *et al.*, 1992) o por el contrario las diferencias no son notorias a nivel funcional (Orsomando *et al.*, 2003). Según lo observado por Oliva (2006) (datos no publicados), el alelo 2P1A elicitó plantas de *Nicotiana benthamiana* de igual forma como lo hace el alelo de *P. infestans* encontrado, por lo que se deduce que los cambios a nivel estructural no tienen ingerencia en el reconocimiento producido en la planta, cualquiera que este sea.

Al realizar el análisis de neutralidad no se logró determinar sitios bajo selección positiva con probabilidades posteriores importantes, ni tampoco se encontraron diferencias significativas entre los modelos neutrales y los modelos alternativos al realizar la prueba de LTR, por lo que se sugiere que este gen no se encuentra bajo selección positiva sino más bien que se encuentra bajo selección relajada, cercana a la neutralidad (Win *et al.*, 2007). Garzón *et al.*, (2007) ratifica la ausencia de selección positiva en este gen. Esto contradice la idea de que se trata de un gen de avirulencia (se ha encontrado selección positiva en más del 75% de las proteínas RXLR [Win *et al.*, 2007]), por lo menos en un sentido estricto, ya que hacen falta análisis funcionales que demuestren la acción de este gen en el hospedero. Teóricamente los genes de avirulencia se encuentran bajo selección positiva, ya que la interacción co-evolutiva con las proteínas R obliga a este tipo de moléculas a cambiar para evadir el reconocimiento de la planta, en lo que se denomina modelo de carrera armamentista (Thomas y Stephen, 1999; Sthali y Bishop, 2000). La ausencia de selección positiva observada en este gen, posiblemente debida a la falta de presión evolutiva por parte de un gen R

específico en los hospederos, es concordante con las ideas del modelo de carrera armamentista descrito anteriormente.

La diversidad encontrada en el gen RD2 apoya los estudios realizados para diferenciar a las especies *P. infestans* y *P. andina* (Adler *et al.*, 2004; Oliva *et al.*, 2007). Por otro lado, la falta de selección positiva encontrada, apoya la teoría de que el gen RD2 se trata de un elicitor general (no requiere de un gen de resistencia específico para provocar HR).

#### 4.1.2 Gen *Avr3a*

Bos *et al.*, (2006) describe dos clases de alelos en este gen: AVR3a<sup>KI</sup> (K<sup>80</sup>, I<sup>103</sup>) (avirulento) y AVR3a<sup>EM</sup> (E<sup>80</sup>, M<sup>103</sup>) (virulento), en donde el aminoácido 103 puede ser indistintamente I (isoleucina) o M (metionina), esto quiere decir que puede existir el alelo AVR3a<sup>KM</sup> (K<sup>80</sup>, M<sup>103</sup>) y AVR3a<sup>EI</sup> (E<sup>80</sup>, I<sup>103</sup>). Tanto el alelo KM y EI disminuyen su capacidad de avirulencia o virulencia respectivamente, con respecto a KI y EM, pero el polimorfismo en la posición 80 determina esta cualidad realmente (KI ó KM avirulento y EM ó EI virulento). En las secuencias analizadas (tanto ecuatorianas como secuencias OSU) los alelos KI, EI y EM fueron encontrados. El sitio 139 (contando los 21 aminoácidos correspondientes al péptido señal, que no fue amplificado en las secuencias utilizadas en este estudio) fue identificado como polimórfico en el alineamiento a nivel de aminoácidos (Figura 3.6). Dicho sitio no fue reportado por Bos *et al.*, (2006) aunque al realizar pruebas de funcionalidad, estos autores encontraron falta de actividad elicitora al cortar la proteína AVR3a en el sitio 131, a pesar de encontrarse las posiciones 80 y 103 presentes. El haber hallado al sitio 139 como polimórfico puede dar una pista para explicar dicha observación y es posible que este sitio tenga ingerencia en la acción elicitora de AVR3a.

A nivel de estructura primaria se presenta una diferencia importante en el punto isoeléctrico de la secuencia 3P6A (*P. infestans*) con relación al resto de

alelos AVR3a, al parecer producido por el aminoácido metionina en el sitio 103, dado que es el único cambio existente que no es reportado en ninguna otra secuencia AVR3a.

El resultado del análisis de máxima probabilidad (ML) determinó que tanto los modelos nulos como los modelos alternativos no presentan diferencias de acuerdo a la prueba de LRT, lo que indicaría que este gen se encuentra bajo selección neutra y/o purificadora. De acuerdo a Yang y Bielawski (2000), el método ML necesita información tanto de las sustituciones sinónimas como de no sinónimas, y dado que no se registran sustituciones sinónimas entre las secuencias analizadas, se presume que el método ML falló al calcular los índices de probabilidad para determinar la LRT. Al excluir las secuencias que no presentan diferencias entre sí (3P12A y 3P12B), el resultado con ML fue distinto ya que en el modelo nulo M1, se obtuvieron 3 sitios bajo selección positiva al igual que en los modelos alternativos M2, M3 y M8 (sitios 80, 103 y 139) (datos no mostrados en la sección resultados). Este fenómeno resulta extraño ya que el modelo M1 no debería detectar sitios bajo selección positiva, sin embargo la falta de sitios sinónimos entre las secuencias *Avr3a* pudo haber llevado a dicho resultado. De igual forma cuando la selección positiva ocurre en intervalos de tiempo muy cortos y afecta a pocos aminoácidos cruciales solamente, el método ML no resulta muy exitoso para el análisis, fenómeno que no es descartado para explicar el comportamiento de este gen (Yang y Bielawski, 2000). El simple hecho de no encontrar sitios sinónimos en las secuencias ya es una prueba de que el gen *Avr3a* se encuentra bajo selección positiva a través de por lo menos 3 sitios (80, 103 y 139), sitios que indican regiones genómicas de importancia funcional, lo que resalta la importancia del sitio 139 identificado en el alineamiento a nivel de aminoácidos (Nielsen y Yang, 1998).

La variabilidad encontrada en el gen *Avr3a* no determina ninguna diferencia con la variabilidad encontrada a nivel mundial, sin embargo, el sitio polimórfico 139 encontrado en el alineamiento, cuya importancia ratifica el método de máxima

probabilidad, constituye un hallazgo importante que puede significar un avance en el entendimiento de la patogenicidad que conlleva esta secuencia.

#### 4.1.3 Gen RD6

De este gen se pudo recuperar únicamente la secuencia 6P4C perteneciente a *P. andina*, donde se determinó su semejanza con la secuencia RD6\_OSU\_CU3 de *P. infestans* a nivel de aminoácidos. Las secuencias RD6\_OSU\_IPIO y RD6\_OSU\_88069 (*P. infestans*) mantienen una notable semejanza, con solo 3 sitios de divergencia, pero con 19 sitios divergentes con respecto a las secuencias 6P4C y RD6\_OSU\_CU3. La secuencia RD6\_OSU\_IPIO resulta 99% igual al gen IPIO1 (Pieterse *et al.*, 1994) después de buscar esta secuencia en BLASTP. Por su parte la secuencia RD6\_OSU\_88069 resultó 100% igual al gen IPIO2 (Pieterse *et al.*, 1994) tras hacer idéntica búsqueda. Dado que no existen más estudios acerca de este gen en *P. andina*, no se puede concluir que exista un alelo único para esta especie ya que el alelo encontrado es prácticamente idéntico a uno de los alelos de *P. infestans*, lo que ratifica más bien el parentesco entre estas dos especies. Por otro lado, por lo menos 5 alelos han sido reportados por van der Vossen (2007), de los cuales 2, reportados como IPIO3 e IPIO4 se piensa son virulentos, dadas las pruebas de funcionalidad. Con respecto a estos hallazgos, los autores sugieren a los genes *Rpi-sto1* y *Rpi-blb1* (reportados por van der Vossen *et al.*, (2003)) como los posibles genes *R* en solanaceas, responsables de la detección del gen RD6. Estas pruebas confirman la acción del gen RD6 (IPIO) como un elicitador de tipo RXLR.

A nivel de estructura primaria, las secuencias RD6 no presentan diferencias notables, tanto en peso como en punto isoeléctrico, no así en la predicción de estructura secundaria, donde se nota claramente la diferencia entre el alelo 6P4C de los alelos RD6\_OSU\_IPIO (IPIO1) y RD6\_OSU\_88069 (IPIO2) (Figura 3.16). El alto número de sustituciones radicales entre los alelos antes mencionados y el porcentaje de estructuras predichas, justifican de cierta forma el cambio estructural

del alelo 6P4C con respecto a las secuencias OSU\_IPIO y 88069 (IPIO2), pero las implicaciones a nivel de función y reconocimiento solo podrán ser dilucidadas a través de análisis funcionales.

El análisis realizado a través del programa codeml del paquete bioinformático PAML, determinó que el gen RD6 se encuentra bajo selección positiva, dado que los modelos alternativos (M2, M3, M8)(para detectar selección positiva) resultaron estadísticamente diferentes a los modelos nulos (M0, M1, M7), una vez realizada la prueba LRT y comparada a la distribución chi-cuadrado. Este resultado indica que los modelos alternativos se encuentran describiendo a los datos de mejor manera que los modelos nulos. En los tres modelos alternativos M2, M3, M8 se encontraron los mismos sitios bajo fuerte selección diversificativa (o positiva) con probabilidades posteriores >95%, estos fueron 101 R (arginina) y 103 L (leucina). Adicionalmente cinco sitios con probabilidades posteriores de entre el 55 y el 90% fueron encontrados. Estos datos son equivalentes a los encontrados por Garzón *et al.*, (2007), que identificó selección positiva para este gen con sitios probables similares a los mencionados anteriormente. Estos resultados sugieren que el gen RD6 se encuentra determinado bajo el modelo de carrera armamentista (Thomas y Stephen, 1999; Sthal y Bishop, 2000) y se encontraría formando parte activa de la batería de moléculas utilizadas por los patógenos *P. infestnas* y *P. andina* para realizar el ataque a sus hospederos.

El gen RD6 (IPIO) constituye un firme candidato para ingresar en la familia de elicitores RXLR, tanto por las pruebas de funcionalidad que se le han realizado como por el tipo de selección que presenta. La secuencia 6P4C, distinta a IPIO1 e IPIO2, podría tratarse del alelo virulento IPIO3, IPIO4 u otro alelo desconocido, ya que no se ha podido determinar alguna semejanza con secuencias publicadas hasta el momento.

#### 4.1.4 Gen RD39/40

La diversidad alélica encontrada en este gen revela un proceso de selección especialmente activo funcionando sobre la molécula. Esta diversidad revela la importancia de este gen y su actividad biológica dentro de los organismos analizados (Liu *et al.*, 2005).

Los alelos encontrados para *P. infestans* en los aislados ecuatorianos fueron encontrados también en las secuencias enviadas desde Ohio, al igual que una de las secuencias de *P. andina* (40P5A), lo que sugiere que estas dos especies se encuentran cercanamente relacionadas. El haber encontrado un alelo único para *P. andina* es un hallazgo importante, dada la diversidad alélica observada en 114 secuencias analizadas. Lo que indica por su parte una especialización de cada especie a su medio, resultado que respalda los estudios realizados por Ordoñez *et al.*, (2000), Wattier *et al.*, (2003), Adler *et al.*, (2004) y Oliva *et al.*, (2007).

Con respecto al plegamiento secundario de estas proteínas, se observan 3 tipos de estructuras diferentes entre sí, en donde destaca la estructura perteneciente a la secuencia 40P5C de *P. andina*, por tener un patrón de plegamientos notablemente distinto al de los patrones seguidos por las secuencias de *P. infestans* (Figura 3.17). De igual forma este fenómeno se ve reflejado en el número de sustituciones radicales y conservativas entre cada una de las secuencias (Tabla 3.11). Este resultado resalta la importancia de la secuencia 40P5C encontrada en *P. andina* como un alelo diferenciado, exclusivo para esta especie.

Después de analizar los datos por medio del método ML, se nota al gen RD39/40 bajo selección positiva en por lo menos 8 sitios (con probabilidades importantes), el mayor número de sitios encontrados en todo el estudio, lo cual resulta interesante ya que la proteína RD39/40 es la más pequeñas de las cuatro analizadas. Esto denota la importancia funcional de este gen y la fuerte presión co-evolutiva a la que se encuentra sometido. Estos resultados son proporcionales

a los obtenidos por Win *et al.*, (2007), en donde se destaca la presión selectiva (positiva) a la que están sometidas las proteínas de la familia RXLR en distintas especies.

Datos obtenidos de pruebas de funcionalidad indican que el gen RD39/40 produce respuesta de hipersensibilidad en presencia del gen *Rpi-blb3* (Kamoun, comentario personal), lo que junto a las pruebas que se aportan en este estudio ubican a este gen como un fuerte candidato para ser considerado como un elicitor de la familia de proteínas RXLR.

#### 4.2 Transformación bacteriana, detección de clones positivos y secuenciación

Los resultados obtenidos de la transformación bacteriana fueron sumamente alentadores. El protocolo de transformación presentó resultados altamente eficientes, dada la cantidad de colonias obtenidas luego de dicho proceso (4200 colonias transformadas exitosamente). El reporte enviado sobre las construcciones realizadas en la Universidad Estatal de Ohio (OSU) (datos no mostrados), revelan que la ligación inserto-plásmido fue exitosa, 12 clones de 20 muestreados presentaron construcciones positivas. Varias razones han sido propuestas para explicar el bajo porcentaje de clones positivos logrados en CIP-Quito (30 clones de 4260 transformados presentaron construcciones positivas, el 0.7%): la extracción del plásmido de las células originalmente transformadas en Ohio, la liofilización realizada para el traslado del material genético hacia Ecuador, la resuspensión de los plásmidos en el laboratorio de CIP-Quito y el mismo proceso de transformación en células *E. coli*. Sin embargo el plásmido pGR106 ha sido utilizado en diferentes investigaciones como vector de transformación con muy buenos resultados (Guo *et al.*, 2007; Gaulin *et al.*, 2006; Bos *et al.*, 2006).

Un fenómeno interesante fue encontrado en la detección por PCR de clones positivos. Como se puede ver en la figura 3.1, las transformaciones 6P4B, 40P5B, 40P8B, 40P8C y 40P11D presentan una doble banda en el gel de

electroforesis, una de las cuales corresponde al control negativo (plásmido solamente) y la otra al plásmido + inserto. Este fenómeno pudo ser provocado por contaminación, ocurrida el momento de respaldar los clones transformados (probablemente se tomaron dos colonias, una positiva y otra negativa). Dado que en muchas de las cajas de cultivo originales, la densidad de colonias era muy alta y en algunos casos las colonias se encontraban unidas, esta teoría resulta plausible para explicar el patrón de bandas inusual encontrado en varios de los carriles de la figura 3.1.

De 30 clones enviados como positivos hacia Ohio (OSU), solamente 21 lograron ser secuenciados, los 9 restantes fueron reportados como falsos positivos ya que no se pudieron procesar efectivamente. De las 21 secuencias enviadas desde Ohio (OSU), 3 resultaron incongruentes frente al resto de secuencias, por lo que fueron retiradas del análisis bioinformático (Tabla 3.2). Varios errores en la secuenciación se aducen a la posible contaminación explicada en la sección anterior, a posibles falsos positivos no detectables a través de PCR, a la pérdida de inserto en el plásmido, que pudo ser producida en el viaje hacia Ohio o a graves errores encontrados en el centro MCIC de Ohio, instituto encargado de realizar la secuenciación para este proyecto (errores como: contaminación de muestra e identificación equivocada de las secuencias fueron comprobados).

## **CAPÍTULO V: CONCLUSIONES**

1. La falta de selección positiva encontrada en el gen RD2 apoya la teoría de que este gen no está sometido a presión co-evolutiva con un gen de resistencia en el hospedero, lo que explicaría su acción elicitora en plantas sin genes de resistencia específicos. Por otro lado la diversidad encontrada en RD2 apoya los estudios realizados para diferenciar a *P. infestans* y *P. andina*.
2. No existen diferencias notables entre la variabilidad reportada a nivel mundial y los hallazgos obtenidos en este estudio con respecto al gen *Avr3a*. Se ha

logrado determinar 3 de los 4 alelos reportados en aislados ecuatorianos, sin embargo un nuevo sitio polimórfico no descrito ha sido observado. Las implicaciones de este sitio en la virulencia o avirulencia de la proteína son desconocidos.

3. La secuencia 6P4C no presenta semejanza con los alelos reportados IPIO1 e IPIO2, por lo que se deduce que se trata del alelo IPIO3, IPIO4, o algún otro alelo no determinado. Dadas las pruebas de funcionalidad realizadas por otros autores y la selección positiva detectada en el presente trabajo, este gen se presenta como un candidato importante para ser considerado un gen de avirulencia dentro de la familia de proteínas RXLR.
4. La alta diversidad alélica y la fuerte selección positiva a la que se encuentra sometido el gen RD39/40 supone una función de importancia dentro de los organismos que han sido analizados. El alelo exclusivo encontrado en *P. andina*, sugiere que este gen tiene relación con la especialización del organismo a su ambiente, dado que las especies estudiadas *P. infestans* y *P. andina* no comparten hospederos comunes.
5. Existen diferencias entre *P. infestans* y *P. andina* en dos de los cuatro genes estudiados (RD2 y RD39/40), lo que afirma la primera hipótesis planteada en este trabajo.
6. Existen diferencias intra-específicas en tres de los cuatro genes analizados (*Avr3a*, RD6 y RD39/40), lo cual es congruente con el sistema de hipótesis planteado.
7. El gen RD2 fue identificado como neutro, a diferencia de los genes *Avr3a*, RD6 y RD39/40 en donde la selección natural fue determinada. Este hecho ratifica la hipótesis planteada con respecto a este tema y es una prueba clara de la

existencia de distintas fuerzas evolutivas actuando al mismo tiempo sobre diferentes genes.

8. Solamente en alelos del gen RD6 y RD39/40 se lograron identificar diferencias importantes en cuanto a la estructura secundaria de las proteínas. Esta observación ratifica la hipótesis planteada al respecto pero al mismo tiempo no determina una relación clara entre la estructura de la proteína y su acción elicitora.

## **CAPÍTULO VI: RECOMENDACIONES**

Son necesarios estudios de funcionalidad que ratifiquen o descarten a los genes RD2, RD6 y RD39/40 como genes de avirulencia, así como estudios que determinen los correspondientes genes *R* que se encuentran en las solanaceas, en procura de una pronta aplicación de estos conocimientos como herramienta en el control de plagas.

Es importante analizar una mayor cantidad de genes de avirulencia de una muestra significativa de aislados, lo que ayudaría a caracterizar la diversidad genética de los patógenos que infectan a la familia de las solanaceas en los andes, de lo cual se conoce muy poco hasta el momento.

Se deberían realizar estudios similares que involucren a plagas de importancia dentro de los cultivos del Ecuador.

## **CAPÍTULO VII: REFERENCIAS BIBLIOGRÁFICAS**

Adler N., Erselius L., Chacón M., Flier W., Ordoñez M., Kroon L., Forbes G. (2004). Genetic diversity of *Phytophthora infestans sensu lato* in Ecuador provides new insight into the origin of this important plant pathogen. *Phytopathology*, 94, 154-162.

Agrios G. (1988). Plant Patology (3th ed.). San Diego: Academic Press.

Andrade H., Cuesta X., Oyarzún P. (1998). Mejoramiento participativo en Ecuador aplicado al incremento de la severidad de *Phytophthora infestans* en el INIAP. Participatory Research & Gender Análisis. Extraído el 10 de octubre, 2007, del sitio Web de PRGA: <http://www.prgaprogram.org/cds/fmp/NADINE-PDF/ANDRADE.pdf>

Arahana V. (2006). Sílabo de Laboratorio 1er Semestre. Quito: U.S.F.Q.

Armstrong M., Whisson S., Pritchard L., Bos J., Venter E., Avrova A., Anne P., Rehmany A., Bohme U., Brooks K., Cherevach I., Hamlin N., White, B., Fraser A., Lord A., Quail M., Churcher C., Hall N., Berriman M., Huang S., Kamoun S., Beynon J., Birch P. (2005). An ancestral oomycete locus contains late blight avirulence gene *Avr3a*, encoding a protein that is recognized in the host cytoplasm. PNAS. 102 (21), 7776-7771.

Backeljau, T., Bruyn, L., De Wolxt, H., Jordaens, K., Van Dongen, S., Winnepeninckx, B. (1996). Multiple UPGMA and Neighbor-joining Trees and the Performance of Some Computer Packages. Molecular Biology and Evolution 13(2), 309-313.

Berman H., Westbrook J., Feng Z., Gilliland G., Bhat T., Weissig H., Shindyalov I., Bourne P. (2000). The Protein Data Bank, Nucleic Acid Research, 28(1), 235-242.

Birch P., Rehmany A., Pritchard L., Kamoun S., Beynon J. (2006). Trafficking arms: oomycete effectors enter host plant cells. Trends in Microbiology. 14(1), 8-11.

Bos J., Kanneganti, T., Young C., Cakir C., Huitema E., Win J., Armstrong M., Birch P.R.J., Kamoun S. (2006) The C-terminal half of *Phytophthora infestans* RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in *Nicotiana benthamiana*. *Plant Journal*, 48,165-176.

Centro Internacional de la Papa. (1999). Laboratory manual for *Phytophthora infestans* work at CIP-Quito. Quito: CIP-Quito.

Centro Internacional de la Papa. (1996). *Papa. 25 Aniversario CIP.* Lima: Autor.

Dangl, J., McDowell, J. (2006). Two modes of pathogen recognition by plants, *PNAS*, 103(23), 8575-8576.

Drenth A., (1994). Molecular genetic evidence for a new sexually reproducing population of *Phytophthora infestans* in Europe. (Ph.D. thesis). Wageningen Agricultural University, Netherlands.

Erwin D., y Ribeiro K. (1996). *Phytophthora Diseases Worldwide.* Minnesota: APS press.

FAO-CIP. (1995). Potatoes in 1990s Situation and prospects for the world potato economy. Roma: Autor.

Flier W. (2001). *Variation in Phytophthora infestans-sources and implications.* (Ph.D. thesis), Netherlands: Wageningen University.

Flier W., Adler N., Kroon L., Forbes G. (2004). *Phytophthora andina* a new heterothallic pathogen of Solanaceous hosts in the Andean highlands of Ecuador.  
Enviado

Flor H. (1971) Current status of gen-for-gen concept. Annual Review of Phytopatology, 9. 275-296.

Garzón C. (2007). Evolution of RXLR effector proteins in the *Phytophthora infestans* complex (Dissertation, American Phytopathological society meeting, 2007)

Gaulin E., Drame N., Lafitte C., Torto-Alalibo T., Martinez Y., Ameline-Torregrosa C., Khatib M., Mazarguil H., Villalba-Mateos F., Kamoun S., Mazars C., Dumas B., Bottin A., Esquerre-Tugaye M., Rickauer M. (2006). Cellulose-binding domains of a *Phytophthora* cell wall protein are novel pathogen-associated molecular patterns, Plant Cell, 18, 1766-1777.

Goodman R., Novacky A. (1994). The Hypersensitive reaction in Plants to Pathogens. Minnesota: APS press.

Guo Y., German T., Schultz R. (2007). A cryptic promoter in potato virus X vector interrupted plasmid construction, BMC Molecular Biology, 8 (17), 1-6.

Hawkes J. G. (1990). The Potato, evolution, biodiversity and genetic resources. London: Belhaven Press.

Henfling J. (1990). El Tizón Tardío de la Papa (Rep No 4). Lima: International Potato Center.

Hillis D., Moritz C. (Eds). (1990). Molecular Systematics. Massachusetts: Sinauer.

INEC, MAG, SICA. (2002). III Censo Nacional Agropecuario Ecuador. Quito: Autor

INEC, MAG, SICA. (2002). III Censo Nacional Agropecuario Pichincha. Quito: Autor

Kamoun S. (1993). Molecular Genetics of Pathogenic Oomycetes. *Eukaryotic Cell*, 2(2), 191-199.

Kamoun S. (2006). A catalogue of the effector secretome of plant pathogenic oomycetes, *Annual Review of Phytopathology*, 44, 41-60.

Kroon L., Bakker F., van den Bosch G., Bonants P., Flier W. (2004). Phylogenetic analysis of *Phytophthora* species based on mitochondrial and nuclear DNA sequences, *Fungal Genetics and Biology*, 41, 766–782.

[Lamour K., Win J., Kamoun, S.](#) (2007). Oomycete genomics: new insights and future directions. *FEMS Microbiology Letters*, 274, 1-8.

Lehninger A., Nelson D., Cox M. (1982). Principles of Biochemistry (2da ed.). New York: Worth Publishers.

Li W., Graur D. (1991). *Fundamentals of Molecular Evolution* Massachusetts: Sinauer.

Liu Z., Bos J., Armstrong M., Whisson S., da Cunha L., Torto-Alalibo T., Win J., Avrova A., Wright F., Birch P., Kamoun S. (2004). Patterns of Diversifying Selection in the Phytotoxin-like scr74 Gene Family on *Phytophthora infestans*. *Molecular Biology and Evolution*. 22(3), 659-672.

Martin F., Tooley P. (2003). Phylogenetic relationships among *Phytophthora* species inferred from sequence analysis of mitochondrially encoded cytochrome oxidase I and II genes, *Mycologia*, 95(2), 269–284

Mayoral C., Noroña M., Baroni M.R., Giani R., Zalazar F. (2005). Evaluation of a nested-PCR assay for *Streptococcus pneumoniae* detection in pediatric patients

with community-acquired pneumonia, *Revista Argentina de Microbiología*, 37, 184-188.

McDonald B. (2004). Population Genetics of Plant Pathogens, *The Plant Health Instructor*. DOI:10.1094/PHI-A-2004-0524-01.

Nei M. (1987). *Molecular Evolutionary Genetics*. New York: Columbia Univ. Press.

Nespoulous C., Huet J. C., Pernollet J. C. (1992). Structure-function relationships of and elicitors, signal proteins involved in the *plant-Phytophthora* interaction, *Planta*, 186, 551-557.

Nielsen R., Yang Z. (1998). Likelihood models for detecting positively selected amino acid sites and application to the HIV-1 envelope gene, *Genetics*, 148, 929–936.

Oliva R., Chacón M., Cooke D., Lees A., Forbes G. (2007). Is *Phytophthora infestans* a good taxonomist? Host recognition and co-evolution in the *Phytophthora/Solanum* interaction, *Acta Horticulturae*, 745, 465-471.

Ordóñez M. (1993). *Identificación de Genes Mayores de Resistencia a Phytophthora infestans en Patata (Solanum tuberosum)*. Tesis de Licenciatura en Ciencias Biológicas. Quito: PUCE.

Ordóñez M., Hohl H., Velasco J., Ramon M., Oyarzun P., Smart C., Fry W., Forbes, G., Erselius L. (2000). A novel population of *Phytophthora*, similar to *P. infestans*, attacks wild *Solanum* species in Ecuador. *Phytopathology*, 90, 197-202.

Orsomando G., Lorenzi M., Ferrari E., de Chiara C., Spisni A., Ruggieri S. (2003). PcF protein from *Phytophthora cactorum* and its recombinant homologue elicit

phenylalanine ammonia lyase activation in tomato, *Cellular and Molecular Life Sciences*, **60**, 1470–1476.

Oyarzun P., Taibe J., Forbes G. (2001). *Phytophthora infestans* activity and particularities in Ecuador. CGIAR. Extraído el 4 de febrero, 2007, de [http://research.cip.cgiar.org/typo3/GILB/fileadmin/Pdfs\\_Archive/Proceedings\\_GILB/Spanish/1.3\\_P.i.\\_su\\_actividad\\_y\\_particularidades\\_en\\_el\\_Ecuador\\_perfil\\_de\\_pais\\_OK.pdf](http://research.cip.cgiar.org/typo3/GILB/fileadmin/Pdfs_Archive/Proceedings_GILB/Spanish/1.3_P.i._su_actividad_y_particularidades_en_el_Ecuador_perfil_de_pais_OK.pdf).

Pieterse C., van West P., Verbakel H., Brass P., van den Berg-Velthuis G., Govers F. (1994). Structure and genomic organization of the *ipiB* and *ipiO* gene clusters of *Phytophthora infestans*, *Gene*, **138**, 67-77.

Rzhetsky A., Nei M. (1993). Theoretical Foundation of the Minimum-Evolution Method of Phylogenetic Inference. *Molecular Biology and Evolution*. **10**(5), 1073-1095.

Sthal E., Bishop J. (2000). Plant–pathogen arms races at the molecular level, *Current Opinion in Plant Biology*, **3**, 299–304.

Taylor C. (1996). Inside Avirulence. *The Plant Cell*, **8**, 1091-1093.

Thomas F., Stephen R. (1999). Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. Proceedings of the Nacional *Academy of Sciences*. **96**, 6820–6823.

van der Vossen E., Sikkema A., te Lintel B., Gos J., Stevens P., Muskens M., Wouters D., Pereira A., Stiekema W., Allefs S. (2003). An ancient *R* gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato. *The plant journal*, **36**, 867–882

van der Vossen E. (2007). Durable late blight resistance in potato: dream or reality (Dissertation, International Potato Center, 2007)

Vega M. (1999). *Estudio de la Especialización de la Patogenicidad del Linaje US-1 de Phytophthora infestans (MONT.) de Bary en Papa y Tomate con Aislados Africanos y andinos*. Tesis de Licenciatura en Ciencias Biológicas. Quito: PUCE.

Wattier R., Gathercole L., Assinder S., Gliddon C., Deahl K., Shaw D., Mills D. (2003). Sequence variation of intergenic mitochondrial DNA spacers (mtDNA-IGS) of *Phytophthora infestans* (Oomycetes) and related species, *Molecular Ecology Notes*, 3, 136-138.

Win J., Morgan W., Bos J., Krasileva K., Cano L., Chaparro-Garcia A., Ammar R., Staskawicz B., Kamoun S. (2007). Adaptive Evolution Has Targeted the C-Terminal Domain of the RXLR Effectors of Plant Pathogenic Oomycetes, *The Plant Cell*, 19, 1-21.

Yang Z., Bielawski J. (2000). Statistical methods for detecting molecular adaptation, *Tree*, 15(12), 496-503.

Yang Z., Nielsen R., Goldman N., Krabbe A. (2000). Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites, *Genetics*, 155, 431–449.

Zolan M. (1995). Chromosome-Length Polymorphism in Fungi, *Microbiological Reviews*, 59, 686–698.