UNIVERSITAT POLITÈCNICA DE VALÈNCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Dpto. de Biotecnología

Estudio de alelos de riesgo de celiaquía e intolerancia a la lactosa en población celiaca española

Trabajo Fin de Máster

Máster Universitario en Biotecnología Biomédica

AUTOR/A: Albuixech Lluch, Eva

Tutor/a: Gadea Vacas, José

Cotutor/a externo: CHAVES MARTINEZ, FELIPE JAVIER

Director/a Experimental: GARCIA GARCIA, ANA BARBARA

CURSO ACADÉMICO: 2022/2023

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

DEPARTAMENTO DE BIOTECNOLOGÍA

Estudio de alelos de riesgo de celiaquía e intolerancia a la lactosa en población celiaca española

TRABAJO FIN DE MÁSTER EN BIOTECNOLOGÍA BIOMÉDICA

ALUMNO/A: Eva Albuixech Lluch
TUTOR/A: José Gadea Vacas
Curso Académico: 2022-2023

València, 2 de junio del 2023

TÍTULO

Estudio de alelos de riesgo de celiaquía en intolerancia a la lactosa en población celiaca española

RESUMEN

La enfermedad celíaca es una enteropatía del intestino delgado con una base genética y una prevalencia mundial del 1%. Genotipos concretos de los genes *HLA-DQA1* y *HLA-DQB1* confieren riesgo para esta enfermedad según la literatura. Por otro lado, la intolerancia a la lactosa es la incapacidad de digerir la lactosa y esta puede ser causada por lactasa no persistente, que consiste en la disminución de los niveles de lactasa después del destete. Esta afección también presenta una base genética ya que genotipos concretos en el gen *LCT* producen una malabsorción de la lactosa. El objetivo principal de este proyecto es analizar e identificar marcadores genéticos que confieren riesgo para la enfermedad celiaca y la lactasa no persistente en una población española de 221 individuos, llamada CeliacaVa, en la que se incluyen enfermos celiacos y sus familiares. Para ello se aplicó un panel de amplicones NGS para estudiar ciertas regiones de los genes *HLA-DQA1, HLA-DQB1 y LCT.* La secuenciación se llevó a cabo por el sistema MiSeq de Illumina y los resultados pasaron filtros de calidad. Las secuencias obtenidas se analizaron con la herramienta IGV y se obtuvieron los genotipos de los pacientes, además de sus frecuencias alélicas y genotípicas. Con estos datos se quiso asociar los polimorfismos del gen *LCT* con la enfermedad celiaca además de estudiar los genotipos de los genes *HLA* y asociarlos con el riesgo de celiaquía. Finalmente, se construyeron árboles familiares con los genotipos obtenidos. El análisis estadístico no encontró asociación entre la combinación de los distintos alelos de los genes *HLA* y la enfermedad celiaca. Se encontraron diferencias significativas entre las frecuencias de celiacos y de la población control para el genotipo GG del SNP rs182549. También se observaron diferencias significativas entre las frecuencias alélicas de los alelos DQA1*01, DQA1*05, DQB1*02, DQB1*05 y DQB1*06 entre la población celiaca y la población control. Con el objetivo de validar estos resultados, es necesario utilizar poblaciones de mayor tamaño que puedan tener más fuerza estadística.

Palabras clave: Celiaquía, Intolerancia a la lactosa, HLA, NGS, Alelos de riesgo.

Autora: Eva Albuixech Lluch

Localidad y fecha: València, junio de 2023

Tutor: José Gadea Vacas

Cotutor externo: Felipe Javier Chaves Martínez

Directora experimental: Ana Bárbara García García

TITTLE

Study of risk alleles for celiac disease and lactose intolerance in the Spanish celiac population

ABSTRACT

Celiac disease is a small bowel enteropathy with a genetic basis and a worldwide prevalence of 1%. Specific genotypes of the *HLA-DQA1* and *HLA-DQB1* genes confer risk for this disease according to the literature. On the other hand, lactose intolerance is the inability to digest lactose, and this can be caused by non-persistent lactase, which is a decrease in lactase levels after weaning. This condition also has a genetic basis as specific genotypes in the *LCT* gene result in lactose malabsorption. The main objective of this project is to analyze and identify genetic markers that confer risk for celiac disease and non-persistent lactase in a Spanish population of 221 individuals, called CeliacaVa, which includes celiac patients and their relatives. For this purpose, a panel of NGS amplicons was applied to study certain regions of the *HLA-DQA1, HLA-DQB1* and *LCT* genes. Sequencing was carried out by the Illumina MiSeq system and the results passed quality filters. The sequences obtained were analyzed with the IGV tool and the genotypes of the patients were obtained, as well as their allelic and genotypic frequencies. These data were used to associate *LCT* gene polymorphisms with celiac disease and to study the genotypes of the *HLA* genes and associate them with the risk of celiac disease. Finally, family trees were constructed with the genotypes obtained. Statistical analysis found no association between the combination of the different alleles of the *HLA* genes and celiac disease. Significant differences were found between the frequencies of celiacs and the control population for de GG genotype of the rs182549 SNP. Significant differences were found also between allelic frequencies in DQA1*01, DQA1*05, DQB1*02, DQB1*05 y DQB1*06 alleles between celiacs and the control population. In order to validate these results, it is necessary to use larger populations that may have more statistical power.

Key words: Celiac Disease, Lactose intolerance, HLA, NGS, Risk Allels.

Author: Eva Albuixech Lluch

Location and date: Valencia, June 2023

Tutor: José Gadea Vacas

External cotutor: Felipe Javier Chaves Martínez

Experimental director: Ana Bárbara García García

ÍNDICE

ÍNDICE DE LAS FIGURAS

ÍNDICE DE LAS TABLAS

ABREVIATURAS

APC: del inglés *Antigen presenting cells*, células presentadoras de antígenos

BAM: del inglés *Binary aligment map*

EC: abreviatura de enfermedad celiaca

EMA: Anticuerpos antiendomisio

ESPGHAN: siglas de European Society for Paediatric Gastroenterology, Hepatology and Nutrition

GWAS: del inglés *whole genome association study*, estudios de asociación de genoma completo

HLA: del inglés *human leucocyte antigen*, antígeno leucocitario humano.

IL: abreviatura de intolerancia a la lactosa

LNP: abreviatura de lactasa no persistente

MHC: del inglés *major histocompatibility complex,* complejo mayor de histocompatibilidad.

NCBI: siglas de National Center for Biotechnology Information

NGS: del inglés *next generation sequencing*

PCR: del inglés Polimerase Chain Reaction

SNP: del inglés *single nucleotide polymorphisms*

Ttg: del inglés *tissue transglutaminase*, transglutaminasa tisular

ULN: del inglés *upper limit of normal*

1. INTRODUCCIÓN

1.1. ENFERMEDAD CELIACA

La enfermedad celiaca (EC) es una enteropatía común del intestino delgado que aparece en individuos genéticamente predispuestos en la que la ingestión del gluten en la dieta causa un daño en la mucosa del intestino delgado produciendo síntomas principalmente gastrointestinales en el paciente e impidiendo la correcta absorción de nutrientes (Dunne *et al.*, 2020).

La EC afecta aproximadamente al 1% de la población mundial, pero se estima que esta proporción puede ser mucho mayor, y aunque en los últimos años su prevalencia ha aumentado, la tasa de diagnóstico ha aumentado más lentamente. La EC se diagnostica con mayor frecuencia en mujeres con una proporción mujer-hombre que oscila entre 2:1 y 3:1. La enfermedad puede manifestarse a cualquier edad, aunque es más común su aparición en dos picos de edad: el primero un poco después del destete del infante y su posterior incorporación de gluten en la dieta y el otro en la segunda o tercera década de vida (Catassi et al., 2022).

1.1.1. FISIOPATOLOGÍA DE LA ENFERMEDAD CELIACA

La EC es una enfermedad autoinmune con un componente genético clave (el antígeno leucocitario humano HLA-DQ2 y HLA-DQ8), un autoantígeno (la transglutaminasa tisular, (tTG)) y un desencadenante ambiental (el gluten) (Caio *et al.*, 2019). El gluten, entre otras proteínas, está formado por gliadinas que son proteínas complejas ricas en prolinas y glutaminas las cuales no son completamente digeribles por las enzimas intestinales por lo que los péptidos resultantes de la digestión parcial de la gliadina entran en la lámina propia del intestino delgado dónde son desaminados por la tTG (Serena et al., 2015). Cada tipo de HLA presenta un conjunto diferente de antígenos peptídicos, basados en la secuencia de aminoácidos presente en la región de unión al péptido del heterodímero. Las moléculas HLA-DQ2 y HLA-DQ8 de las células presentadoras de antígenos tienen una alta afinidad por los péptidos cargados negativamente como los generados en la desaminación de la gliadina, son los únicos capaces de presentar estos péptidos derivados del gluten, mientras que otros tipos de HLA no pueden. (Dieli-Crimi *et al.,* 2015; Brown et al., 2019). Se produce por lo tanto una respuesta inmunitaria adaptativa en la cual estos péptidos junto a la tTG y sin la tTG se presentarán a las células T CD4+ activándolas y produciendo citoquinas proinflamatorias y factores de crecimiento produciendo la muerte de células epiteliales. Los linfocitos CD4+ también activarán a las células B que producirán anticuerpos contra la tTG epitelial, la gliadina y la α-actina contribuyendo a la EC. Además, hay una sobreexpresión de IL-15 e IL-8 en los enterocitos de las células epiteliales que causan un

reclutamiento de neutrófilos produciéndose una respuesta inmune innata. Aunque no está totalmente clara la interacción entre estas dos respuestas inmunitarias parecen ser necesarias para la formación de la lesión patológica celiaca (Tian Ting *et al*., 2020; Caio *et al.,* 2019; Lebwohl, 2018). Hay otros mecanismos menos caracterizados que pueden desempeñar un papel en el desarrollo de las lesiones. Por ejemplo, los miofibroblastos intestinales cuando producen la remodelación tisular liberan tTG y α-actina que serán reconocidas por los anticuerpos empeorando la EC (Dunne *et al.*, 2020).

1.1.2. GENÉTICA DE LA ENFERMEDAD CELIACA

1.1.2.1. SISTEMA HLA

En la EC hay una influencia genética del sistema HLA, específicamente la región HLA de clase II. El sistema HLA (de las siglas en inglés Human Leucocyte Antigen) son un conjunto de genes que codifican las moléculas encargadas de presentar los antígenos a los linfocitos T, llamadas MHC (del inglés Major Histocompatibility Complex) o moléculas HLA. El sistema HLA se ubica en el brazo corto del cromosoma 6 y sus genes se agrupan en 3 regiones: clase I, II y III, pero son los genes clásicos de clase I y II los asociados con la pr esentación de antígenos. Las moléculas HLA se expresarán en la superficie de la célula presentadora de antígenos *(Figura 1).* El sistema HLA es poligénico además de muy polimórfico. Los polimorfismos (variaciones de aminoácidos) se localizan en los *pokets* de la hendidura de unión con el antígeno. Existen 35.821 alelos distintos del sistema HLA y 10.592 alelos distintos de los genes HLA de clase II (HLA nomenclature web, https://hla.alleles.org/nomenclature/index.html; Marsh et al., 2010). Los genes más relacionados con la celiaquía son *HLA-DQA1* y *HLA-DQB1* que juntos codifican las dos cadenas (α y β respectivamente) de las proteínas heterodímeras DQ que se expresan en la superficie de las células presentadoras de antígenos. Los dominios α y β se proyectan al medio extracelular en forma hélice y cadena plegada formando una hendidura dónde se unirán los antígenos peptídicos (Brown et al., 2019; Dieli-Crimi *et al.,* 2015). La cadena β tiene un peso molecular de aproximadamente 27kDa y el gen que codifica para esta consta de 6 exones. El exón 1 codifica el péptido líder, los exones 2 y 3 codifican los dominios extracelulares, el 4 el dominio transmembrana y el 5 la cola citoplasmática. Hay 556 alelos distintos para el gen *HLA-DQA1* y 2419 para el gen *HLA-DQB1* que codifican para 267 y 1490 proteínas distintas respectivamente (HLA nomenclature web; Marsh et al., 2010).

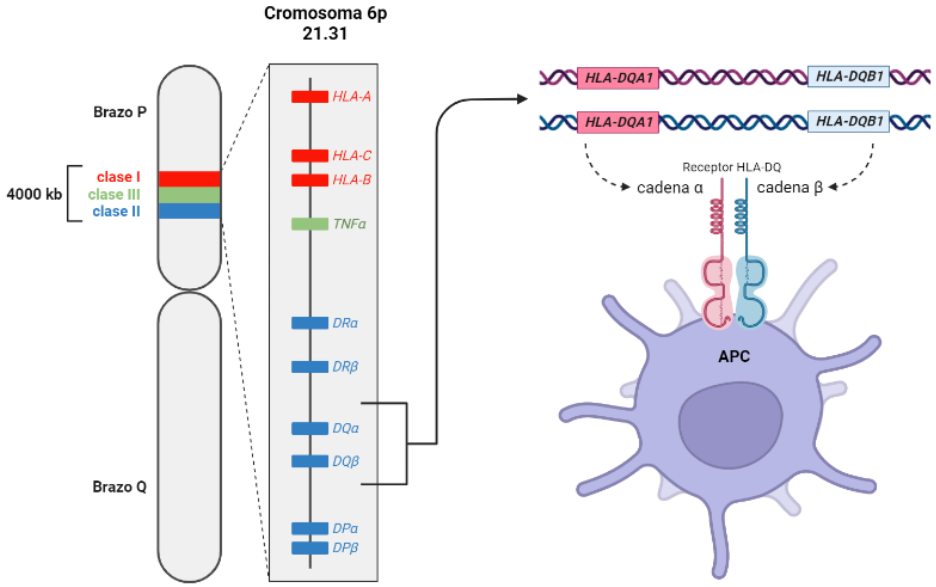

Figura 1.Estructura del sistema HLA y estructura proteica. Los genes que codifican las moléculas HLA se encuentran en el complejo MHC del cromosoma 6. Las moléculas HLA implicadas en la enfermedad celíaca se codifican en una región conocida como clase II en los loci DQ. Los loci HLA-DQA1 y HLA-DQB1 codifican para las cadenas α y β respectivamente, que se asocian como heterodímeros en la superficie de la célula presentadora de antígenos (APC). Figura adaptada de Brown et al., 2019. Creada con Biorender (https://www.biorender.com/).

1.1.2.2. NOMENCLATURA SISTEMA HLA

Cada alelo HLA tiene un nombre único que sigue una estructura concreta que consta de los siguientes apartados como se muestra en la *Figura 2* (HLA nomenclature web, 2020):

- Prefijo HLA
- Locus particular del HLA
- Grupo alélico
- Alelo HLA específico

Además de estos sufijos mencionados le pueden seguir a la nomenclatura otros adicionales para hacer distinción de mutaciones sinónimas, mutaciones fuera de la región codificante, alelos nulos o diferencias en la expresión en la superficie celular, pero en el presente trabajo se limitará la nomenclatura hasta el nivel alélico específico.

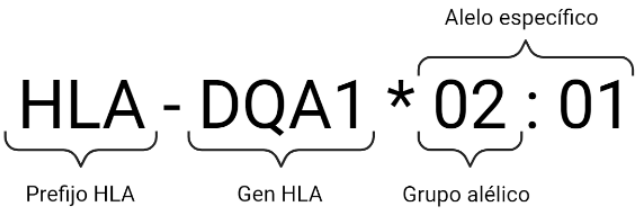

Figura 2. Estructura de la nomenclatura de los alelos HLA. Adaptada de la web HLA nomenclature.

1.1.2.3. GENOTIPOS DE RIESGO PARA LA ENFERMEDAD CELIACA

La EC, al igual que muchas enfermedades autoinmunes, tiene un fuerte componente hereditario y así lo demuestra su elevada recurrencia familiar (10-15%) y su elevada concordancia de la enfermedad entre gemelos monocigóticos (75-80%). Casi el 100% de los pacientes celiacos poseen variantes específicas de los genes *HLA-DQA1* y *HLA-DQB1*, concretamente los alelos que codifican para las moléculas HLA-DQ2 y HLA-DQ8 son las principales variantes genéticas que confieren riesgo de EC (Dieli-Crimi *et al.,* 2015; Lebwohl, 2018; Espino & Núñez, 2021). Alrededor del 90% de pacientes con EC son portadores de los alelos que codifican para la molécula HLA-DQ2.5: DQA1*05 y DQB1*02, que pueden heredarse en configuración cis o en trans. La mayoría de los pacientes celiacos que no manifiestan el HLA-DQ2.5 son portadores de los alelos DQA1*03 y DQB1*03:02 en configuración cis que codifican para la molécula HLA-DQ8. Casi todos los pacientes que no manifiestan HLA-DQ2.5 ni HLA-DQ8, presentan solo uno de los alelos que codifican para HLA-DQ2.5, DQB1*02 codificando HLA-DQ2.2 y en una minoría de casos presentan solo el alelo DQA1*05 codificando para HLADQ7.5. *(Figura 3)*. Finalmente, con una frecuencia extremadamente baja, algunos pacientes no portan ningún alelo de riesgo HLA conocido (Megiorni & Pizzuti, 2012; Dieli-Crimi *et al.,* 2015; Brown et al., 2019).

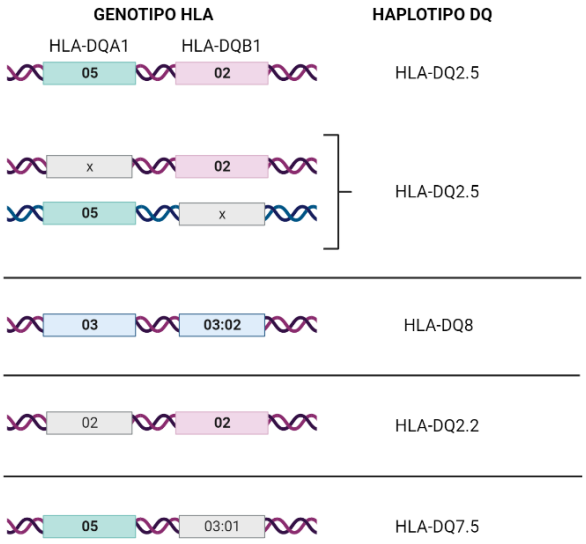

Figura 3. Alelos de los genes HLA-DQA1 y HLA-DQB1 y haplotipos DQ que confieren riesgo para la enfermedad celiaca. Los alelos marcados en negrita son los que confieren riesgo, los alelos no marcados acompañan a los alelos de riesgo para dar un haplotipo concreto. Adaptada de Espino & Núñez, 2021. Creada con Biorender.

El riesgo de desarrollar EC puede variar según los alelos heredados, los complejos de HLA-DQ formados y de la cantidad de complejos HLA-DQ que se expresan en la superficie de las células presentadoras de antígenos. Por ejemplo, los pacientes con dos copias de DQA1*05 y DQB1*02 y que, por lo tanto, son homocigotos para el haplotipo HLA-DQ2.5 lo que supone que pueden reconocer una mayor diversidad de péptidos del gluten por lo que se asocian con el mayor riesgo de desarrollar EC (Vader *et al.,* 2003; Pisapia *et al.,* 2016). En líneas generales los haplotipos que le seguirían en orden de conferir un riesgo de desarrollar EC son HLA-DQ8 (presente en los dos alelos), HLA-DQ2.2 y HLA-DQ7.5, a su vez la interacción entre haplotipos puede aumentar o disminuir el riesgo de desarrollar EC o no tener efectos aditivos (*Tabla 1*). Además, la expresión de la tTG puede aumentar debido a la lesión tisular y esta lesión puede ser propiciada por factores ambientales como infecciones víricas o microbianas. El riesgo también puede variar según el sexo y la edad (Dieli-Crimi *et al.,* 2015).

Tabla 1. Grupos de riesgo para la enfermedad celiaca según el haplotipo. Adaptada de Martínez-Ojinaga et al. (2018). DQX se refiere a los alelos que no confieren riesgo.

Riesgo	**Haplotipo DQ ***	**Descripcion HLA**
Muy alto	DQ2.5 / DQ2.5 DQ2.5 / DQ2.2	Portadores de DQ2.5 con dos copias de HLA-DQB1*02
Alto	DQ2.2 / DQ7.5 DQ2.5 / DQ8 DQ2.5 / DQ7.5 DQ2.5 / DQX DQ8 / DQ8	Portadores de DQ2.5 con solo una copia de HLA-DQB1*02 Individuos con DQ8 / DQ8
Intermedio	DQ8 / DQ7.5 DQ8 / DQ2.2 DQ8 / DQX DQ2.2 / DQ2.2 DQ 2.2 / DQX	Portadores heterocigotos de DQ8 (no portadores de DQ2.5) Portadores de DQ2.2 (no portadores de DQ2.5)
Bajo	DQ7.5 / DQ7.5 DQ7.5 / DQX	Portadores de DQ7.5 (no portadores de DQ2.5)

*La correspondencia de los haplotipos está plasmada en la *Figura 3*.

Además, estudios de GWAS (estudios de asociación de genoma completo) han relacionado variantes genéticas no relacionadas con los genes HLA (Dieli-Crimi *et al.,* 2015; Jansen et al., 2015). El factor genético es necesario, aunque no suficiente para el desarrollo de la enfermedad, algunos estudios relacionan el desarrollo de la EC con disbiosis intestinal, aunque aún se requiere más investigación en este campo (Caio *et al.*, 2019). También se han relacionado otros factores ambientales como el patrón de alimentación en el primer año de vida, la presencia o no de lactancia materna y el momento de introducción del gluten en la dieta. Si bien, recientes estudios han concluido que la lactancia materna no es determinante para el desarrollo de la EC (Auricchio y Troncone, 2021). En cuanto a la ingesta de gluten en los primeros años de vida, el

estudio TEDDY, que incluyó a 6605 niños predispuestos genéticamente para la EC concluyó que la ingesta diaria de gluten durante los primeros años de vida está asociada con el riesgo aumentado de desarrollar la EC en niños predispuestos genéticamente (Aronsson et al., 2019). Las infecciones gastrointestinales, principalmente por *Enterovirus*, también pueden ser un desencadenante (Auricchio y Troncone, 2021).

1.1.2. SÍNTOMAS, DIAGNÓSTICO Y TRATAMIENTO

Los principales síntomas de la EC son gastrointestinales como la diarrea crónica o intermitente, el estreñimiento, el dolor abdominal, el abdomen distendido o las náuseas y vómitos. También pueden presentarse síntomas extra gastrointestinales como pérdida de peso, problemas en el crecimiento, fatiga crónica, neuropatías, anemia o baja mineralización de los huesos y sus respectivas consecuencias (Kartpati et al., 2012; Leffler et al., 2015; SEGHNP, 2020). Según la ESPGHAN (de las siglas en inglés European Society for Paediatric Gastroenterology, Hepatology and Nutrition), las personas con los síntomas mencionados anteriormente deben ser consideradas para el diagnóstico de la EC, así como las personas con alguno de los siguientes factores de riesgo: familiares de primer grado con la enfermedad, personas con enfermedades autoinmunes como diabetes tipo 1 o enfermedad de la tiroides, personas con los síndromes de Down, Turner o Williams-Beuren o personas con deficiencia de IgA. Cuando hay sospecha de EC en un paciente se recomienda:

En primer lugar, pruebas serológicas como las siguientes:

- Anti-tTG-IgA: se valora en sangre la cantidad de anticuerpos anti-transglutaminasa tisular de clase IgA.
- Anti-Ttg-IgG: se procede a esta prueba cuando el paciente presenta un valor bajo de IgA respecto a su edad y se valora en sangre la cantidad de anticuerpos IgG anti-transglutaminasa, antipéptidos deaminados de la gliadina o antiendomisio.
- EMA-IgA: cuando la cantidad de anti-tTG-IgA ≥ 10xULN (de las siglas en inglés Upper Limit of Normal), es decir, 10 veces mayor a los valores normales, se procede a medir los anticuerpos del endomisio (EMA) en una muestra serológica nueva. En el caso de deficiencia en la IgA se medirá la EMA-IgG.

Si las pruebas anti-tTG-IgA o IgG y EMA-IgA o IgG son positivas y la anti-tTG-IgA/G ≥ 10XULN se considera que el paciente padece de la EC y no es necesario recurrir a la biopsia. Los anti-TTG no siempre están presentes en los pacientes diagnosticados con EC, sobre todo en aquellos que presentan un menor daño en la mucosa intestinal (Dore et al., 2017).

En segundo lugar, realizar una biopsia: se debe proceder a esta prueba cuando los niveles de anti-TTG-IgA <10XULN y la EMA-IgA/G es negativa. Los pacientes deben someterse a al menos 4 biopsias del duodeno distal y al menos 1 del bulbo mientras toman una dieta con gluten. En esta prueba se valorará el daño estructural de la mucosa (Husby *et al*., 2020).

El diagnóstico genético de la EC se basa en la detección de los haplotipos HLA-DQ2 y HLA-DQ8 del gen HLA. Sin embargo, la presencia de estos alelos no es suficiente para diagnosticar la enfermedad ya que las moléculas heterodímeras HLA-DQ2 y HLA-DQ8 están presentes en la mayoría de los pacientes con EC, pero también en el 30-40% de la población general. Por esta razón no se recomienda como test de cribado ya que los pacientes con compatibilidad genética para la EC no necesariamente desarrollarán la enfermedad. Aun así, esta prueba tiene un valor predictivo negativo casi del 100%, ya que puede descartar la enfermedad si no existe ninguna variante asociada (Espino & Núñez, 2021; Ludvigsson et al., 2014). A estas pruebas se suele recurrir cuando el paciente presenta algún factor de riesgo o un posible falso negativo en las pruebas serológicas. Para el tipaje del HLA también se puede recurrir a técnicas serológicas, aunque estas son limitadas ya que son dependientes del estado del paciente y de la viabilidad celular, en cambio las técnicas moleculares aportan mayores ventajas siendo más sensibles y específicas, detectando así un número mayor de polimorfismos (Husby *et al*., 2020).

Las técnicas moleculares utilizadas para la tipificación del HLA actualmente son la PCR (Polimerase Chain Reaction) en tiempo real y la NGS (Next Generation Sequencing). La PCR se utiliza para amplificar y detectar los alelos HLA-DQ2 y HLA-DQ8 en muestras de sangre periférica, mientras que la NGS permite analizar múltiples regiones del genoma simultáneamente y detectar variantes genéticas menos comunes asociadas a la EC. Recientemente, se han desarrollado diferentes kits comerciales para genotipar específicamente los alelos HLA-DQ2 y HLA-DQ8. También se utiliza la PCR-SSOP (sequence-specific oligonucleotide probes) en la que se usan oligos específicos para estos genes con un indicador, como la fluorescencia (Itoh, 2005). El tipaje del HLA puede ser una prueba útil en el cribado de personas con familiares celiacos de primer grado. Además, los estudios familiares han demostrado que la EC se produce casi exclusivamente en presencia de moléculas DQ de alto riesgo. Las investigaciones futuras deben dirigirse a identificar nuevos marcadores, que puedan mejorar la predicción de la enfermedad en personas portadoras de los alelos HLA de riesgo. (Megiorni y Pizzuti, 2012).

La carga económica y sanitaria que supone la EC es considerable. Las estrategias de prevención de esta enfermedad se basan en la identificación de los factores de riesgo a los que se enfrenta el paciente, como ser portador de un alelo HLA de riesgo, y la posterior prevención de los

posibles factores ambientales desencadenantes. Una de las estrategias de prevención podría centrarse en vacunar a la población frente a ciertos microorganismos o la manipulación de la microbiota de los pacientes. De hecho, algunos estudios han demostrado una menor incidencia de la EC en pacientes vacunados contra *Rotavirus.* Por el contrario, se han publicado estudios controvertidos sobre el uso de probióticos para la prevención de la EC. La identificación de más factores de riesgo puede resultar interesante para el desarrollo de nuevas estrategias de prevención (Auricchio y Troncone, 2021)

Actualmente, el único tratamiento eficaz disponible para la EC es una dieta libre de gluten de por vida con la que se elimina la presencia de autoanticuerpos, se reparan las vellosidades intestinales y se acaba con los síntomas intestinales y extraintestinales. Sin embargo, este tratamiento puede conllevar algunos inconvenientes como problemas psicológicos, posibles carencias de vitaminas y minerales, mayor riesgo cardiovascular o estreñimiento que pueden afectar en la calidad de vida del paciente. Por ello, en los últimos años se han intentado desarrollar terapias alternativas existiendo actualmente algunos ensayos clínicos en curso como, por ejemplo, el uso de proteasas del gluten provenientes de bacterias o el acetato de larazotida, un antagonista de la zonulina que ha demostrado eficacia en el control de los síntomas. También hay algunos ensayos clínicos como por ejemplo anticuerpos monoclonales IL-15 y una vacuna (Nexvax2) para desensibilizar a los pacientes con EC frente a la gliadina (Caio *et al.*, 2019).

1.2. INTOLERANCIA A LA LACTOSA

La lactosa, presente en la leche de los mamíferos, es un disacárido compuesto por los monosacáridos glucosa y galactosa. Para que estos puedan ser absorbibles por el tracto intestinal la lactosa ha de ser hidrolizada por una enzima llamada lactasa-florizina hidrolasa. Esta enzima está codificada por el gen de la lactosa (*LCT*). En gran parte de la población mundial, los niveles de esta enzima disminuyen drásticamente después del destete, impidiendo la digestión de la lactosa. Esto es conocido como lactasa no persistente (LNP), presenta una base genética y produce una malabsorción de la lactosa (Anguita et al., 2020; Szilagyi y Ishayek, 2018; Alkalay, 2021).

La malabsorción de la lactosa puede producir intolerancia a la lactosa (IL), que son el conjunto de síntomas que se producen después de la ingesta de leche. Entre estos síntomas se incluyen dolor abdominal, hinchazón abdominal, flatulencias, diarrea, nauseas o vómitos y son producidos por el gas derivado de la fermentación de la lactosa por las diferentes bacterias que habitan en el tracto gastrointestinal (Catanzaro et al., 2021). La sintomatología puede ser muy variable entre los afectados.

La prevalencia mundial de la IL es alrededor de un 57%, aunque se estima que puede ser de un 65%. Su prevalencia varía mucho entre los distintos continentes, e incluso en la población europea podemos encontrar prevalencias de 10% en los países del norte y de 50% en el área mediterránea (Anguita et al., 2020)

1.2.1. GENÉTICA INTOLERANCIA A LA LACTOSA

El gen *LCT* humano está localizado en el cromosoma 2q21, está formado por 17 exones y tiene un tamaño de aproximadamente 49 kb. Se han identificado polimorfismos en el gen LCT que parecen predisponer a la IL (Anguita-Ruiz et al., 2020).

En la población europea las variantes más frecuentes relacionadas con la IL son LCT-13910C>T y LCT-22018G>A. Se ha relacionado el polimorfismo LCT-13910C>T con la persistencia a la lactasa, siendo los genotipos C/T y T/T las que confieren persistencia de la lactasa y C/C se asocia a la LNP. En cuanto al polimorfismo LCT-22018G>A, los genotipos A/A y G/A confieren persistencia de la lactasa y G/G se asocia a la LNP (Mattar et al., 2012; Misselwitz et al., 2019). Sin embargo, estos SNPs son prevalentes en la población europea mientras que para otras poblaciones y etnias se relacionan otros SNPs con la IL (Anguita-Ruiz et al., 2020).

1.2.2. DIAGNÓSTICO Y TRATAMIENTO

Para el diagnóstico de la IL existen varios métodos. Uno de ellos es el test genético ya que, como hemos visto algunos polimorfismos confieren LNP (LCT- 13910C>T and LCT-22018G>A), pero la presencia de estos polimorfismos no implica IL. Sin embargo, resulta útil para distinguir IL primaria o secundaria y así acertar con el tratamiento. Otro método de diagnóstico es el test de hidrógeno espirado y es el más utilizado. Además, es barato, específico y fácil de realizar e interpretar. El test mide la cantidad de hidrógeno espirado después de la ingestión de 25-50 g de lactosa. Este hidrógeno espirado es producido por la fermentación de la lactosa, no digerida previamente, llevada a cabo por la microbiota intestinal. También se pueden realizar la prueba rápida de la lactasa, en la que se biopsia una muestra de la mucosa duodenal postbular y se incuba con lactosa para ver si hay o no actividad de la lactasa, aunque esta resulta demasiado invasiva para la población general. Finalmente, otro método es la prueba de la tolerancia a la lactosa en la que se mide en sangre la glucosa 30, 60 y 120 minutos después de la ingesta de lactosa (Catanzaro et al., 2021; Szilagyi y Ishayek, 2018).

Actualmente, existen tratamientos para la IL. El primero de ellos es evitar consumir productos con lactosa o elegir aquellos con una cantidad de lactosa reducida. Cada vez hay más productos sin lactosa en el mercado que además están suplementados con minerales y vitaminas propios

de los productos lácteos. Otro tratamiento es la lactasa oral que se debe tomar previa a la ingestión de lactosa. También se ha relacionado el consumo de probióticos con la mejora sintomática de la IL, aunque los resultados de los estudios son variables.

1.2.3. RELACIÓN CELIAQUÍA E INTOLERANCIA A LA LACTOSA

La LNP y la EC son dos trastornos con base genética muy prevalentes y que afectan a la mucosa del intestino delgado. Una de las consecuencias de la coexistencia de la IL y la EC puede ser la confusión entre la IL y una mala adaptación a la dieta libre de gluten. La IL también puede ser secundaria o adquirida, y sucede como consecuencia en personas con enfermedades gastrointestinales que dañan el borde del cepillo del intestino delgado y hacen que disminuyan los niveles de lactasa, como puede ser la EC. Generalmente, con la retirada del gluten, se restablece la absorción normal de la lactosa. (Basso et al., 2012; Usai-Satta et al., 2022; Alkalay, 2021).

Existen pocos estudios que relacionen la LNP e IL con la EC. En el estudio de Basso y sus compañeros (2012) concluyeron que los marcadores de LNP están igualmente representados en niños con EC que en la población control. También se ha encontrado malabsorción de la lactosa secundaria entre un 10 y 19% de niños celiacos en algunos estudios (Usai-Satta et al., 2022). En 2005 Ojetti y sus compañeros observaron una prevalencia de EC en pacientes con IL (positivos para el test de hidrógeno espirado para la detección de la IL) de un 24% respecto a una prevalencia del 2% en el grupo control.

2. OBJETIVOS

La EC y la LNP son dos enfermedades que afectan al tracto gastrointestinal y tienen una elevada prevalencia entre la población mundial. Además, ambas tienen un componente genético. El objetivo principal de este Trabajo de Fin de Grado es analizar e identificar marcadores genéticos que confieran riesgo para la EC y la LNP en una población española.

Para ello, los objetivos específicos de este trabajo son:

1. Aplicación de un panel de amplicones NGS diseñado por los miembros de la Unidad de Genómica y diabetes de INCLIVA para secuenciar 221 individuos pertenecientes a un estudio poblacional de enfermos celiacos y sus familiares para obtener sus genotipos y haplotipos de los genes *LCT, HLA-DQA1 y HLA-DQB1.*
2. Estudiar en una población española con EC la frecuencia de genotipos *HLA* y *LCT*, comparar con población general y evaluar la asociación de los polimorfismos del gen *LCT* con la EC.
3. Construir árboles familiares de una población española de personas celiacas y sus parientes con los genotipos del gen *HLA* obtenidos.

3. METODOLOGÍA

3.1. POBLACIONES 'CELIACAVA' Y CONTROLES

Las muestras analizadas provienen de un estudio llamado CeliacaVa. El estudio CeliacaVa recoge a 221 individuos, en los cuales 78 son celiacos y el resto son familiares.

Como población control para el análisis estadístico del genotipado de los genes *HLA-DQA1* y *HLA-DQB1* se utilizaron los datos de un estudio poblacional obtenido a través de la base de datos Allele Frequencies in Wordwide Populations (http://www.allelefrequencies.net/hla6003a.asp) (Gonzalez-Galarza et al., 2020). Esta población recoge un total de 1742 individuos sanos de Castilla y León, España.

Como población control para el análisis estadístico del genotipado del gen *LCT* se utilizaron los datos poblacionales de la página web ensembl (https://www.ensembl.org/index.html) (Cunningham et al., 2022), concretamente de la población IBS, formada por españoles.

3.2. EXTRACCIÓN MATERIAL GENÉTICO

Las muestras de la población CeliacaVa fueron extraídas previamente por la Unidad de Genómica y Diabetes mediante el sistema Maxwell usando el kit Maxwell 16 LEV Blood DNA kit. Se partió de 300 µl de sangre a la que se añadió proteinasa K, tampón de lisis y se incubó durante 20 minutos a 56ºC en el termobloque. A continuación, se traspasó a las columnas proporcionadas por la casa comercial junto al tampón de elución. Estas columnas incluyen *beads* magnéticas y etanol a distintas concentraciones. El sistema Maxwell realizó la extracción del DNA de forma automática.

3.3. REGIONES AMPLIFICADAS

Las regiones amplificadas en el proceso de secuenciación fueron los exones 2 y 3 del gen *HLA-DQA1*, los exones 2 y 3 del gen *HLA-DQB1.* También se amplificaron dos fragmentos de la región 5' UTR del gen *LCT.*

3.4. DISEÑO Y DISTRIBUCIÓN DE OLIGOS

Para el diseño in silico de oligos se utilizó el software abierto Primer3 (https://primer3.ut.ee/) (Untergasser et al., 2012). Una vez los oligos fueron diseñado se estudiaron las posibles interacciones entre ellos. Para poder evitar el solapamiento de oligos y por lo tanto la posible amplificación de fragmentos no deseados, aquellos que podían interactuar se separaron en dos reacciones distintas (A y B). Los oligos se distribuyeron como se indica en la *Tabla 2*.

Aunque normalmente para la amplificación de cada amplicón se diseñan dos oligos, para un amplicón se utilizaron 3, uno de ellos degenerado para asegurar la hibridación de los diferentes alelos.

Tabla 2. Oligos utilizados para la amplificación del material genético. Se incluyen su nombre y dirección del oligo, el gen para el que amplifican, el exón o zona que amplifican y la reacción en la que se utiliza el oligo (A o B).

Nombre oligo	Cromosoma	Gen	Exón	Reacción PCR	Dirección
DQA1-E2A-5	6	*DQA1*	Exón 2	A	Forward
DQA1-E2A-3					Reverse
DQA1-E2C-5S	6	*DQA1*	Exón 2	A	Forward
DQA1-E2C-5C					Forward
DQA1-E2C-3					Reverse
DQA1-E3A-5	6	*DQA1*	Exón 3	A	Forward
DQA1-E3A-3					Reverse
DQB1-E2Bnd2a-3	6	*DQB1*	Exón 2	A	Forward
DQB1-E2Bnd3a-5					Reverse
DQB1-E2Bnd2b-3	6	*DQB1*	Exón 2	A	Forward
DQB1-E2Bnd3b-5					Reverse
DQB1-E2C-3	6	*DQB1*	Exón 2	A	Forward
DQB1-E2C-5					Reverse
DQB1-E3-5	6	*DQB1*	Exón 3	A	Forward
DQB1-E3-3					Reverse
LAC-13900-5	2	*LCT*	5' UTR	A	Forward
LAC-13900-3					Reverse
DQA1-E2B-5	6	*DQA1*	Exón 2	B	Forward
DQA1-E2B-3					Reverse
DQA1-E3b-5	6	*DQA1*	Exón 3	B	Forward
DQA1-E3b-3					Reverse
DQA1-E3B-5	6	*DQA1*	Exón 3	B	Forward
DQA1-E3B-3					Reverse
DQB1-E2A-5n	6	*DQB1*	Exón 2	B	Forward
DQB1-E2Ab-3					Reverse
LAC.22000-5	2	*LCT*	5'UTR	B	Forward
LAC-22000-3					Reverse

3.5. MULTIPLEX PCR (PCR 1)

Una vez ya diseñados los oligos se procedió con la PCR selectiva, en la que se amplifican las regiones genómicas específicas que queremos estudiar. Las muestras se amplifican en dos reacciones PCR independientes, una utilizando los oligos para la reacción A y otra utilizando los oligos para la reacción B, ambas en las mismas condiciones.

Los reactivos y condiciones utilizados para la PCR1 se muestran en las *Tablas 3 y 4.*

Tabla 3. Condiciones de los reactivos para realizar la PCR1.

Reactivo	Volumen (µl)
Master Mix	7,50
Oligos específicos	2,00
Agua	1,50
DNA (5 ng/µL)	4,00
Volumen final	**15,00**

Tabla 4. Condiciones de la PCR1.

Programa	Temperatura	Tiempo	Ciclos
Activación inicial	95ºC	15 min	
Desnaturalización	98ºC	30 s	30x
Alineamiento	60ºC	30 s	
Extensión	72ºC	30 s	
Extensión final	72ºC	5 min	
Conservación	4ºC	∞	

3.6. BARCODING PCR (PCR 2)

Una vez ya amplificadas las regiones de interés se procedió a la segunda reacción de PCR que tiene como objetivo poder identificar las muestras una vez secuenciadas y para ello en la reacción se incluyen oligos con una secuencia de nucleótidos específica (*barcode*) que se incluirá en cada fragmento y servirá como 'etiqueta' (Mir et al., 2013). En esta PCR, junto a los *barcodes*, también se añaden adaptadores que serán útiles durante el proceso de secuenciación, para que los amplicones se puedan 'adherir' a la *Flow cell*. Para esta reacción se combina el producto obtenido de la PCR1 (multiplex A y multiplex B) de cada muestra y los reactivos.

Los reactivos y condiciones para la PCR2 se muestran en las *Tablas 5 y 6.*

Tabla 5. Condiciones de los reactivos para realizar la PCR2.

Reactivo	Volumen (µl)
Master Mix	7,50
Barcodes (5 µM)	2,00
Agua	3,50
Producto PCR1_reacciónA	1,00
Producto PCR1_reacciónB	1,00
Volumen final	**15,00**

Tabla 6. Condiciones de la PCR2.

Programa	Temperatura	Tiempo	Ciclos
Activación inicial	95ºC	15 min	
Desnaturalización	98ºC	20 s	38x
Alineamiento	60ºC	30 s	
Extensión	72ºC	1 min	
Extensión final	72ºC	10 min	
Conservación	4ºC	∞	

3.7. ELECTROFORESIS CAPILAR

Una vez realizada la PCR 2 se procedió con una electroforesis capilar utilizando el sistema Qiaxcel (Qiaxcel DNA Screening Kit, Qiagen®, Alemania) con el objetivo de comprobar la correcta amplificación de los fragmentos. Este sistema genera los resultados en forma de bandas y de picos con los que podemos ver el tamaño de nuestros fragmentos amplificados.

3.8. PURIFICACIÓN Y CUANTIFICACIÓN

A continuación, se purificó el pool para eliminar aquellos productos de PCR menores de 200pb, los oligos que han formado dímeros y los restos de reactivos. Para la purificación se utilizó el kit comercial Magsi-NGS PREP® kit (MagnaMedics Diagnostics B.V., The Netherlands). El fundamento de este kit consiste en unas perlas magnéticas que tienen afinidad con las cadenas de ADN. En primer lugar, se añaden estas perlas al producto de PCR (con una proporción de 0.6 respecto al volumen del producto de PCR) dando lugar a su unión por afinidad de carga. Se utiliza una placa magnética para que las perlas junto al DNA se una a un lado del pocillo. Después, se realizan varios lavados con etanol para retirar los productos no deseados y finalmente se eluye el DNA con agua ultrapura.

Para la cuantificación del material purificado se utilizó el kit QuantiFluor® dsDNA System (Promega, EE.UU.). Este kit consiste en un fluoróforo, con rangos de excitación y emisión de 504 nm y 531 nm respectivamente, que se unirá al DNA. Para interpolar los resultados en una recta se realizaron una serie de diluciones partiendo de una concentración de DNA conocida proveniente de fago Lambda. Como tampón y blanco se utilizó una solución de Tris-EDTA. La fluorescencia de las diluciones tras añadir el fluoróforo QuantiFluor se analizó mediante el espectrofluorómetro GLOMAX Multi-Detection System spectrofluorometer (Promega®, EE.UU.).

3.9. SECUENCIACIÓN

El secuenciador utilizado fue MiSeq® (Illumina®, EE.UU.).

La tecnología de secuenciación Illumina (Figura 4), perteneciente a la NGS, consiste en una hibridación y amplificación del material genético sobre una *flow cell* y una secuenciación de este material por síntesis y fluorescencia. En primer lugar, los fragmentos de DNA (amplicones) generados previamente con la librería se ligarán a adaptadores complementarios a los oligos ya posicionados en la *flow cell.* En nuestro caso, estos adaptadores se añaden durante el proceso de *barcoding* PCR. Una vez hibridados los fragmentos a la *flow cell* por complementación de bases, se añadirán nucleótidos y enzimas para realizar una PCR formándose puentes de doble cadena. A continuación, se producirá una desnaturalización para separar estos puentes quedando los *clusters* de una cadena. Este proceso se repetirá hasta obtener millones de *clusters* en la *flow cell* y se procederá a la secuenciación por síntesis. En cada ciclo de esta secuenciación se añadirán nucleótidos marcados con fluorescencia, unos terminadores reversibles (3'-OH-dNTPs) y DNA polimerasa. Cada nucleótido que se una será bloqueado por un nucleótido terminador reversible, una vez realizada la unión del nucleótido en la cadena se lavarán los nucleótidos sobrantes y se toma una imagen con un microscopio laser que identificará qué base ha sido añadida en cada *cluster* gracias a la fluorescencia. Después los fluoróforos serán eliminados y el grupo 3'-OH será regenerado para que pueda volver a tener función bloqueadora. A continuación, empezará otro ciclo y así sucesivamente hasta obtener una serie de imágenes con las cuales se obtienen las secuencias después de un procesado de estas (Godwin et al., 2016).

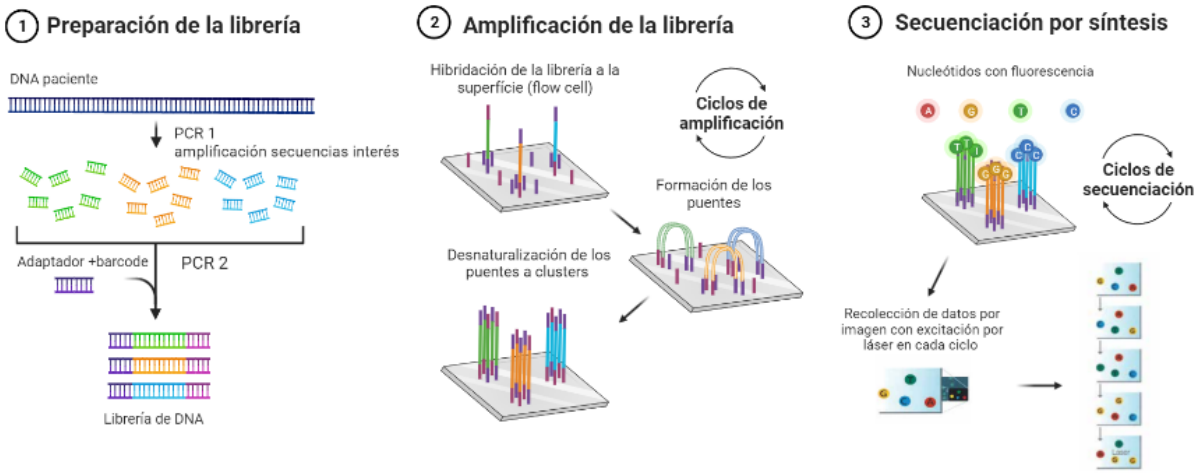

Figura 4. Diagrama Tecnología de secuenciación Illumina. Adaptada de Godwin et al. (2016). Creada con Biorender.

3.10. ANÁLISIS BIOINFORMÁTICO

En primer lugar, las lecturas obtenidas de la NGS se asignaron a archivos FastQ para cada muestra gracias a los *barcodes* utilizados durante la preparación de la librería (Mir et al., 2013). Este proceso, llamado multiplexeado, lo realiza el secuenciador MiSeq de forma automática por lo que partimos directamente de los archivos FastQ. Estos archivos contienen las secuencias de ADN de cada individuo, así como la calidad para cada base en la escala Phred. Esta escala es logarítmica y está representada por el valor Q que estima la probabilidad de que la base secuenciada sea incorrecta. Por ejemplo, un valor de 20 en la escala Phred (Q20) estima una probabilidad de error del 1% para la base secuenciada (Thomas y Hahn, 2019).

En nuestro caso se utilizó la herramienta Seqtk (v1.3-r106) (https://github.com/lh3/seqtk) para eliminar las secuencias cuyo valor en la escala Phred era Q<20 y/o tenían una longitud inferior a 80pb. Para la limpieza de secuencias adaptadoras, cebadores y bases de baja calidad en los extremos de las lecturas se utilizó el programa Cutadapt (v1.18) (Martin, 2011).

Una vez aplicados los filtros anteriores, el control de calidad utilizando FastQC (v.0.11.9) (Andrews et al., 2010) y se elaboró un informe final de las muestras con MultiQC (Ewels et al., 2016). FastQC realiza un control de calidad de los datos de secuenciación en bruto. De este modo, se puede comprobar el estado de los datos y tener en cuenta la presencia de problemas antes de realizar el análisis (Brown et al., 2017).

Los parámetros obtenidos con la herramienta FastQC están dirigidos al análisis de secuenciación de genoma. En nuestro caso, fue secuenciación de amplicones cortos (secuenciación dirigida), por lo que cabe esperar que algunos parámetros fallen durante el control de calidad. Los parámetros pueden fallar en este tipo de secuenciación, y no necesariamente significa que el proceso de secuenciación haya fallado, son: el contenido GC, el contenido de cada base para cada posición en la lectura, la distribución de la longitud de las secuencias, los niveles de duplicación de las secuencias y el número de secuencias sobrerrepresentadas.

El análisis bioinformático de las muestras fue realizado por el personal de bioinformática de la Unidad de Genómica y Diabetes de INCLIVA.

3.11. IDENTIFICACIÓN DE SNPS Y CLASIFICACIÓN DE HAPLOTIPOS

Los archivos obtenidos después del procesamiento y análisis bioinformático fueron archivos de tipo BAM (Binary Aligment Map) y se utilizaron para la visualización de las variantes genéticas de los genes *DQA1* y *DQB1* en los individuos a través del software Integrative Genomics Viewer (IGV) (Versión 2.16.0) (https://software.broadinstitute.org/software/igv/) (James et al., 2011).

Esta herramienta permite explorar visualmente los datos genómicos como los cambios de nucleótidos, deleciones e inserciones, el número de lecturas y la calidad de estas, así como otras muchas funciones.

Para el análisis de los SNPs del gen *LCT* se utilizó el programa Mutation Detector (Seqplexing, https://www.seqplexing.com/es).

Las secuencias de referencia para el análisis de los amplicones están recogidas a continuación y los SNPs (del inglés *Single Nucleotide Polymorphisms*) analizados para los genes *HLA-DQA1* y *HLA-DQB1* están marcados en azul en la secuencia.

Secuencias genes *HLA-DQA1* y *HLA-DQB1*

Secuencia de referencia para el haplotipo DQA1*01:02:01:02 según HLA Database:

Exón 2 DQA1*01:02:01:02

CCGCCTTCCTGCTTGTCATCTTCACTCATCAGCTGACCACGTTGCCTCTTGTGGTGTAAACTTGTACCAGTTTTACGG
TCCCTCTGGCCAGTACACCCATGAATTTGATGGAGATGAGCAGTTCTACGTGGACCTGGAGAGGAAGGAGACTGCCTG
GCGGTGGCCTGAGTTCAGCAAATTTGGAGGTTTTGACCCGCAGGGTGCACTGAGAAACATGGCTGTGGCAAAACACAA
CTTGAACATCATGATTAAACGCTACAACTCTACCGCTGCTACCAATGGTATGCGTCCACCATTCTGCCTCTCTTTACT
TAAGTTATCCCTCCATACCAGGGTTCATTATTTTCTTCCCAAGAGGTCCCCAGATCTTCTTAT

Exón 3 DQA1*01:02:01:02

CTTCAGGGCAGAGCTATTCACACTTCACACCAGTGCTGTTTCCTCACCACAGAGGTTCCTGAGGTCACAGTGTTTTCC
AAGTCTCCCGTGACACTGGGTCAGCCCAACACCCTCATTGTCTTGTGGACAACATCTTTCCTCCTGTGGTCAACATC
ACATGGCTGAGCAATGGGCAGTCAGTCACAGAAGGTGTTTCTGAGACCAGCTTCCTCTCCAAGAGTGATCATTCCTTC
TTCAAGATCAGTTACCTCACCTTCCTCCCTTCTGCTGATGAGATTTATGACTGCAAGGTGGAGCACTGGGGCCTGGAC
CAGCCTCTTCTGAAACACTGGGGTAAGGATGAGTTTCATCATTTTTTGATTCTTTCTTGTCTGTCAAGTTCAGAACTT
CCTGCCTTTTACTCCTATATCCAAAAACTTGTTTTCCACACTTCATGGGTTTCTTTTCTGTCTCTCTTTTTTTT

Secuencia de referencia para el haplotipo DQB1*06:02:01 según HLA Database:

Exón 2 DQB1*06:02:01

GGGGCGACGACGCTCACCTCTCCTCTGCAAGATCCCGCGGAACGCCACCTCGTAGTTGTGTCTGCACACCGTGTCCAA
CTCCGCCCGGGTCCCCTCCAGGACTTCCTTCTGGCTGTTCCAGTACTCGGCATCAGGCCGCCCCTGCGGCGTCACCGC
GCGGTACACCCCCACGTCGCTGTCGAAGCGCGCGTACTCCTCTCGGTTATAGATGTATCTGGTCACAAGACGCACGCG
CTCCGTCCCGTTGGTGAAGTAGCACATGCCCTTAAACTGGAACACGAAATCCTCTGCGGGAATCACCGGCCGGTCAG
TCAGGC

Exón 3 DQB1*06:02:01

GGGAGTCATTTCCAGCATCACCAGGATCTGGAAGGTCCAGTCACCGTTCCTAATGAGGGGGGTGGACACAACGCCGGC
TGTCTCCTCCTGATCATTCCGAAACCACCGGACTTTGATCTGCCCTGGATAGAAATCTGTCACCGAGCAGACCAGCAG
GTTGTGGTGGTTGAGGGCCTCTGTCCTGGATGGGGAGATGGTCA

Además, el cambio de nucleótido, posición y referencia en el NCBI (National Center for Biotechnology Information) para los SNPs analizados están recogidos en las *Tablas 7, 8 y 9*. La información de los genes *HLA-DQA1* y *HLA-DQB1* proviene de la base de datos HLA Database.

Tabla 7. SNPs del gen HLA-DQA1 analizados. En la tabla se incluye la referencia del NCBI, la posición cromosómica y el cambio de nucleótido para cada SNP.

HLA-DQA1

Referencia NCBI	Posición GRCh38.p13	Cambio de nucleótido	Referencia NCBI	Posición GRCh38.p13	Cambio de nucleótido
rs1129737	chr6:32641317	C>T	**rs3207985**	chr6:32641450	A>C
rs1129738	chr6:32641320	T>C	**rs28383449**	chr6:32641451	G>A
rs1129740	chr6:32641328	G>A	**rs1142335**	chr6:32641501	G>C, A
rs3205916	chr6:32641329	T>C	**rs9272742**	chr6:32641945	C>T
rs1071630	chr6:32641349	T>C	**rs9272744**	chr6:32641969	C>T
rs1129753	chr6:32641353	C>T	**rs1048124**	chr6:32642006	C>T
rs1129759	chr6:32641365	C>G	**rs1048134**	chr6:32642012	A>G
rs12722051	chr6:32641370	A>T	**rs707952**	chr632642029	C>T
rs1048023	chr6:32641373	C>G	**rs707951**	chr6:32642036	T>C
rs1048027	chr6:32641392	T>C	**rs41545514**	chr6:32642122	C>T
rs10093	chr6:32641396	C>G	**rs41561312**	chr6:32642187	G>T
rs1142323	chr6:32641415	A>G	**rs7990**	chr6:32642188	C>A
rs1142326	chr6:32641435	C>A, T	**rs2308889**	chr6:32642225	C>A
rs3207983	chr6:32641436	G>A	**rs2308891**	chr6:32642232	C>G, A
rs3207984	chr6:32641437	G>T			

Tabla 8. SNPs del gen HLA-DQB1 analizados. En la tabla se incluye la referencia del NCBI, la posición cromosómica y el cambio de nucleótido para cada SNP.

HLA-DQB1

Referencia NCBI	Posición GRCh38.p13	Cambio de nucleótido	Referencia NCBI	Posición GRCh38.p13	Cambio de nucleótido
rs1130397	chr6:32664810	A>G	**rs9274397**	chr6:32664941	T>A
rs1140320	chr6:32664812	A>G	**rs1049083**	chr6:32664947	C>T
rs1140319	chr6:32664815	C>G	**rs1049082**	chr6:32664967	C>T
rs1140318	chr6:32664816	C>T	**rs1063318**	chr6:32664968	G>A
rs17412833	chr6:32664821	A>T	**no rs**	chr6:32664971	T>A
rs9274379	chr6:32664822	A>G	**rs281874782**	chr6:32664972	A>T, C
rs9274380	chr6:32664824	G>T, C	**rs9274402**	chr6:32664976	C>T
rs1140317	chr6:32664828	C>A	**no rs**	chr6:32664991	G>A
rs1140316	chr6:32664831	C>G	**rs767838657**	chr6:32664992	T>C, G
rs281862113	chr6:32664841	T>C	**rs281862065**	chr6:32664993	A> T, G
rs1130391	chr6:32664847	C>T	**rs9274405**	chr6:32664998	G>C
rs1130389	chr6:32664850	C>G, T	**rs1049068**	chr6:32665000	C>T
rs1130392	chr6:32664851	G>C	**rs1049066**	chr6:32665004	A>T, C
rs9274384	chr6:32664858	A>C	**rs12722115**	chr6:32665005	G>A, C
no rs	chr6:32664860	T>G	**rs1140310**	chr6:32665006	A>C
no rs	chr6:32664861	C>A	**rs41540813**	chr6:32665013	C>A
rs3204379	chr6:32664865	C>T	**rs1049079**	chr6:32665018	C>T, G
rs9274386	chr6:32664868	G>T	**rs188170056**	chr6:32665023	G>A
rs9274387	chr6:32664869	G>T	**rs1130368**	chr6:32665041	T>G
rs1130390	chr6:32664870	T>C	**rs1130375**	chr6:32665043	C>G
no rs	chr6:32664872	C>T	**rs9274407**	chr6:32665055	A>T
no rs	chr6:32664873	C>T	**rs12722107**	chr6:32665056	A>G
rs9274390	chr6:32664882	C>T	**rs9274408**	chr6:32665073	G
rs281874783	chr6:32664883	T>G	**rs74186130**	chr6:32665082	C
no rs	chr6:32664886	C> -	**rs9274409**	chr6:32665087	G
no rs	chr6:32664887	T> -	**rs1049133**	chr6:32662070	G>A
no rs	chr6:32664888	T> -	**rs1049130**	chr6:32662082	G>A
rs41556215	chr6:32664892	G>C	**rs1049088**	chr6:32662091	G>A
rs1130382	chr6:32664895	G>A	**rs1049087**	chr6:32662112	G>A
no rs	chr6:32664910	A>G	**rs1063323**	chr6:32662114	C>T
no rs	chr6:32664911	T>C, G, A	**rs1049086**	chr6:32662127	A>G
no rs	chr6:32664912	C>T	**rs2647032**	chr6:32662128	T>C
rs1130381	chr6:32664914	G>A	**rs41544112**	chr6:32662143	C>T
rs1130380	chr6:32664917	C>A, G	**rs760326072**	chr6:32662144	G>A
rs1140313	chr6:32664923	T>A	**rs41542812**	chr6:32662154	C>G
rs9274395	chr6:32664926	G>A	**rs374741054**	chr6:32662157	C>G
rs1049074	chr6:32664934	C>T	**rs1049107**	chr6:32662159	C>T
rs3210148	chr6:32664937	G>C	**rs1063321**	chr6:32662178	C>T
rs1455718727	chr6:32664938	C>T	**rs1049100**	chr6:32662186	C>T
rs1049073	chr6:32664940	G>A			

Tabla 9. SNPs del gen LCT analizados. En la tabla se incluye la referencia del NCBI, la posición cromosómica, el cambio de nucleótido para cada SNP y la nomenclatura tradicional del SNP para los SNPs analizados del gen *LCT*.

Referencia NCBI	Posición GRCh38.p13	Cambio de nucleótido	Nomenclatura tradicional
rs41525747	Chr2:135851073	C>G	LCT-13907C>G
rs4988235	Chr2:135851073	C>T	LCT-13910C>T
rs41380347	Chr2:135851073	T>G	LCT-13915T>G
rs145946881	Chr2:135851073	G>C	LCT-14010G>C
rs182549	Chr2:135851073	G>A	LCT-22018G>A

De cada paciente se anotaron en un archivo Excel los SNPs analizados de los genes *HLA-DQA1*, *HLA-DQB1* y *LCT* concluyendo así sus alelos para HLA y SNPs en el gen *LCT*. Los distintos alelos posibles para cada gen están recogidos en las *Tablas 10 y 11*, aunque solo se anotaron los alelos según la nomenclatura hasta el nivel alelo-específico. Con los haplotipos obtenidos se diseñaron árboles familiares con la herramienta PowerPoint.

Tabla 10. Clasificación de alelos para el gen HLA-DQA1.

DQA1*01	DQA1*02	DQA1*03	DQA1*04	DQA1*05	DQA1*06
DQA1*01:01:02		DQA1*03:01:01	DQA1*04:01:02:01	DQA1*05:01:01:01	
DQA1*01:02:01:01		DQA1*03:02	DQA1*04:01:02:02	DQA1*05:01:01:02	
DQA1*01:02:01:02		DQA1*03:03:01	DQA1*04:02	DQA1*05:03	
DQA1*01:02:01:03				DQA1*05:05:01:01	
DQA1*01:02:01:04				DQA1*05:05:01:02	
DQA1*01:03:01:01				DQA1*05:05:01:03	
DQA1*01:03:01:02				DQA1*05:11	
DQA1*01:04:01:01					
DQA1*01:04:01:02					
DQA1*01:05:01					
DQA1*01:07Q					
DQA1*01:10					
DQA1*01:11					

Tabla 11. Clasificación de alelos para el gen HLA-DQB1.

DQB1*02	DQB1*03	DQB1*04	DQB1*05	DQB1*06
DQB1*02:01:01	DQB1*03:01:01:01	DQB1*04:01	DQB1*05:01:01:02	DQB1*06:01:01
DQB1*02:02:01:01	DQB1*03:01:01:02	DQB1*04:02	DQB1*05:01:01:03	DQB1*06:02:01
DQB1*02:53Q	DQB1*03:01:01:03	DQB1*04:03	DQB1*05:03:01:01	DQB1*06:02:25
DQB1*02:62	DQB1*03:02:01		DQB1*05:03:01:02	DQB1*06:03:01
	DQB1*03:03:02:01		DQB1*05:102	DQB1*06:03:20
	DQB1*03:03:02:02			DQB1*06:09:01
	DQB1*03:03:02:03			DQB1*06:125
	DQB1*03:03:02:04			DQB1*06:44
	DQB1*03:05:01			

3.12. ANÁLISIS ESTADÍSTICO

Todo el análisis estadístico se realizó con el software R Studio (versión 1.4.1106) (https://www.R-project.org/.).

En primer lugar, la población CeliacaVa se separó según la presencia de EC, se calcularon las frecuencias alélicas (Ecuación 1) para los distintos alelos en población celiaca y no celiaca.

$$Ecuación\ 1: Frecuencia\ alélica\ = \frac{Número\ de\ un\ alelo\ concreto}{2 \cdot número\ total\ de\ individuos\ genotipados}$$

$$Ecuación\ 2: Frecuencia\ genotípica\ = \frac{Número\ de\ un\ genotipo\ concreto}{Número\ total\ de\ individuos\ genotipados}$$

Para ver si existen diferencias significativas entre celiacos y población control, las frecuencias alélicas para los alelos HLA de los celiacos de la población CeliacaVa y de la población control de Castilla y León se compararon con el test Chi-cuadrado.

Para realizar el estudio estadístico de los haplotipos se realizó una regresión logística dónde la variable respuesta es la celiaquía y las covariables son los haplotipos de ambos genes. Además, se hizo una corrección por Benjamin-Hochberg.

Las frecuencias alélicas y genotípicas (Ecuación 2) se calcularon para los SNPs rs4988235 y rs182549 del gen *LCT* y para ver si existían diferencias significativas entre los individuos celiacos de la población CeliacaVa y los controles se realizó test chi-cuadrado.

Además, para cada SNP del gen *LCT* se obtuvieron cinco p.valores para la asociación con la celiaquía asumiendo, respectivamente, herencia dominante, recesiva, codominante, sobredominate (test de tabla de contingencia chi-cuadrado) y aditiva (prueba de Cochran-Armitage) en modelos separados utilizando el paquete estadístico DescTools (Andri Signorell et al., 2023).

4. RESULTADOS Y DISCUSIÓN

4.1. VALIDACIÓN DE LA LIBRERÍA

17 muestras y 3 controles negativos se validaron mediante electroforesis capilar antes de la secuenciación con el sistema Qiaxcel y se muestran en la *Figura 4*. Podemos ver en todas ellas bandas de entre 200 y 300 pb. También se pueden ver en las muestras, al igual que en los controles negativos, bandas más suaves entre 75 y 200 pb que corresponden a dímeros de oligos u oligos no hibridados.

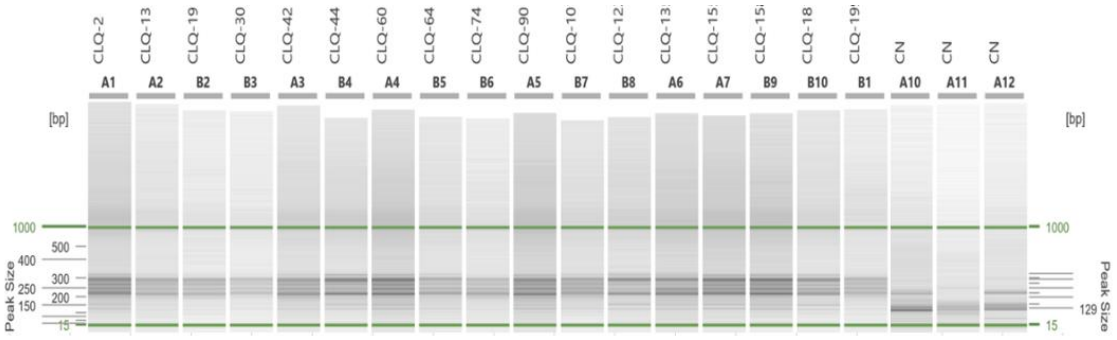

Figura 5. Validación de muestras de la población CeliacaVa mediante electroforesis capilar. Se validaron 17 muestras aleatorias nombradas como "CLQ-" y se validaron 3 controles negativos nombrados como "CN".

Dado que los resultados fueron los esperados, se agrupó en un mismo *pool* todas las muestras, 4µl de cada una. Este pool se validó mediante electroforesis y podemos observar el resultado en la columna A9 de la *Figura 6*. A continuación, se procedió a la purificación del pool mediante el kit Magsi-NGS PREP®. Este pool purificado se diluyó y el resultado de la electroforesis podemos observarlo en las columnas A11 (pool purificado diluido a 1/10) y A12 (pool purificado diluido a 1/20). Estas diluciones se realizaron para poder validar el pool final, ya que sin la dilución la concentración es demasiado alta para el sistema Qiaxcel. Como puede observarse en la *Figura 6*, las bandas de producto de PCR oscilan entre los 200 y 300 pb y no se observan productos PCR inespecíficos ni residuos de reactivos.

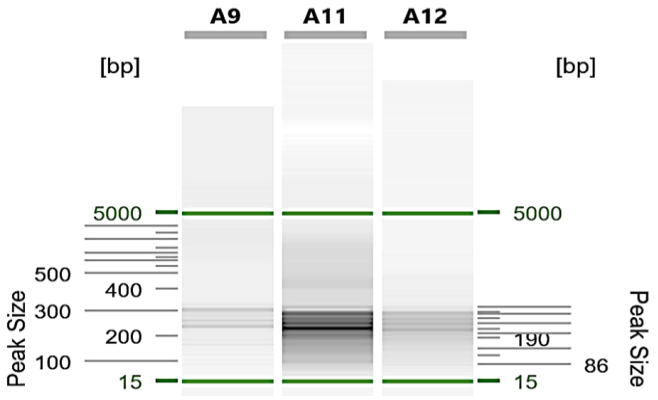

Figura 6. Electroforesis capilar del pool final. A9: pool final sin purificar. A11: pool final purificado diluido a 1/10. A12: pool final purificado diluido a 1/20.

4.2. CONTROL DE CALIDAD DE LA SECUENCIACIÓN

Una vez se realizó la secuenciación se aplicaron algunos filtros de calidad y además obtuvimos algunos datos y parámetros. El número medio de lecturas previo al procesado de los datos fue de 51019 ± 17377. Después del procesado de los datos en el que fueron retirados las secuencias adaptadoras, cebadores, bases de baja calidad en los extremos de las lecturas, lecturas con una calidad de Q<20 de longitud inferior a 80pb, la media de lecturas por individuo fue de 32452 ± 11161.

El porcentaje de lecturas mapeadas contra el genoma fue de 99,4% mientras que el porcentaje de lecturas mapeadas contra las regiones de interés fue de 63,4%. Esta diferencia entre los porcentajes podría ser indicativo de una amplificación inespecífica, pero 12 de 14 de los amplicones pertenecen a regiones HLA, las cuales son muy polimórficas y pueden dificultar el mapeo, aumentando el error.

La *Tabla 12* muestra un resumen de los resultados obtenidos con los parámetros de calidad de FastQC.

Tabla 12. Resultados obtenidos después de llevar a cabo el control de calidad utilizando FastQC. El número de archivos es el doble que el número de muestras analizadas.

	Análisis MultiQC pre-procesamiento Resultado		Análisis MultiQC post-procesamiento Resultado	
Parámetro de calidad	*PASS*	*FAIL*	*PASS*	*FAIL*
Calidad de cada base para la posición en la lectura	437	3	440	0
Contenido de cada base para la posición en la lectura	0	440	0	440
Contenido GC	0	440	0	440
Contenido de N para cada base	440	0	440	0
Niveles de duplicación de la secuencia	0	440	0	440
Secuencias sobrerrepresentadas	0	440	0	440
Contenido de adaptadores	0	440	440	0

Los parámetros de calidad que han fallado son debido a que se trata de lecturas obtenidas por secuenciación dirigida por amplicones (fragmentos sobrerrepresentados de longitud concreta) y no de genoma. Al no estar representado el genoma completo, tanto el contenido GC como el de cada base para la posición en la lectura no está equilibrado, y como solo se han secuenciado unas pocas regiones concretas, hay una sobrerrepresentación y duplicación de estas secuencias. También podemos observar en la *Tabla 12* que la calidad de cada base para la posición en la

lectura es correcta y 3 muestras han pasado finalmente el control de calidad después del procesado. Además, el contenido de N (base no identificada) es igual a 0 y se ha realizado correctamente la eliminación de adaptadores.

Después del procesamiento y filtrado de las muestras se obtuvo con la herramienta MultiQC un histograma (*Figura 7*) en el que se representa el valor Phred medio para cada posición de la base en la lectura. Podemos observar que los valores Phred están por encima de 30 excepto para una muestra que está entre 25 y 30. El filtro que se aplicó fue de Q20 por lo que todas las muestras tienen una calidad mayor a Q20 al pasar este filtro.

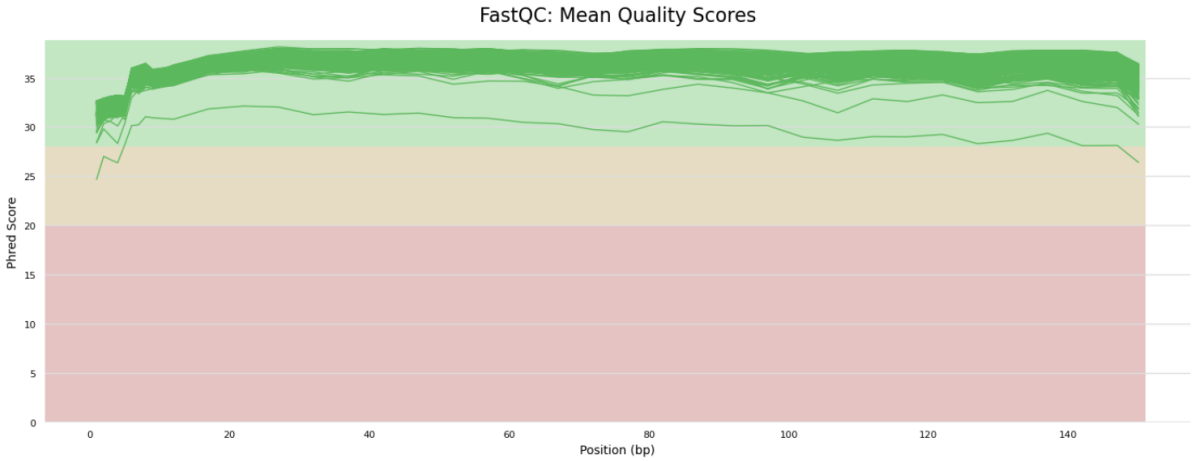

Figura 7. Histograma de la calidad de las secuencias. Representa el valor medio de calidad (valor Phred) para cada posición de la base en la lectura. Histograma realizado con la herramienta MultiQC.

La cobertura del proceso de secuenciación es representada por el número de lecturas para una región concreta del genoma. En nuestro caso la cobertura para cada amplicón se representa con barras en la *Figura 8*. La media de cobertura, representada por una línea roja discontinua, es de alrededor de 1800 lecturas. Podemos observar que algunos amplicones tienen más cobertura que otros y, aunque la media de todos supera las 500 lecturas y esto es suficiente para obtener resultados y analizarlos, sería conveniente equilibrar las coberturas de los amplicones en futuros procesos de secuenciación. Si bien, esto es complejo dada la variabilidad de las secuencias de oligos incluidos y de los propios amplicones.

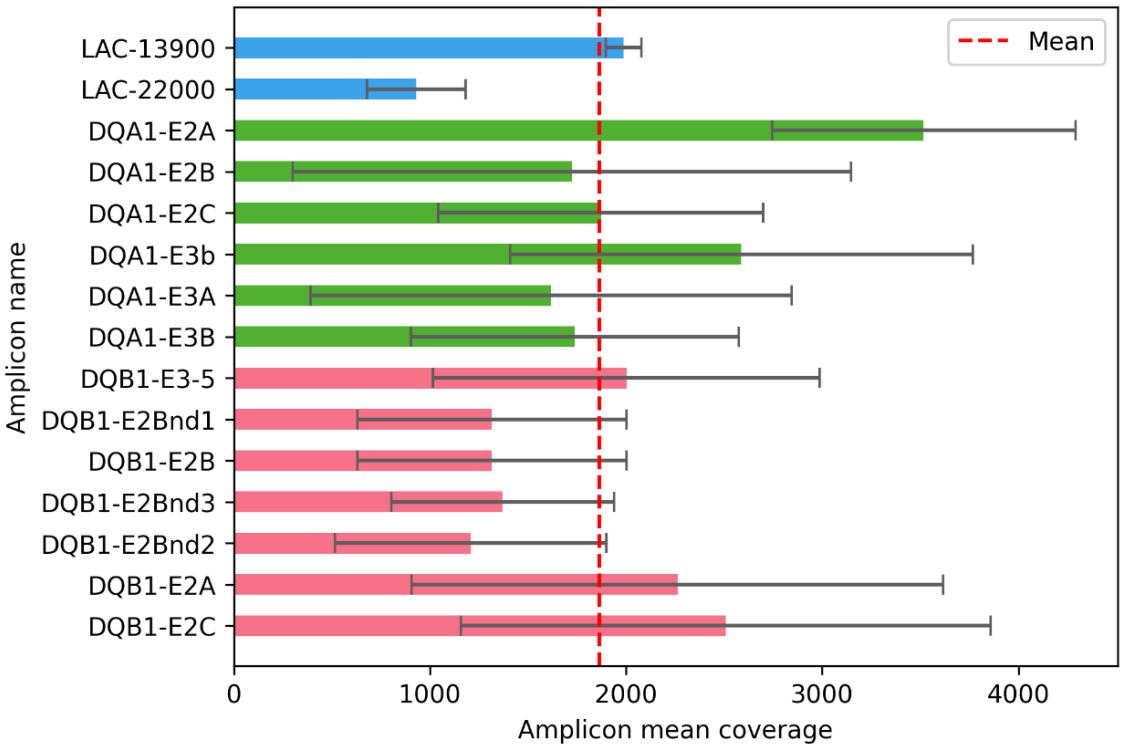

Figura 8. Diagrama de barras de la cobertura de los amplicones. Representa la media de cobertura para cada amplicón y su desviación típica. También se muestra la media de cobertura para todos los amplicones con la línea discontinua roja. Diagrama realizado con el software estadístico R Studio.

4.3. GENOTIPADO GENES *DQA1*, *DQB1* Y *LCT*

El genotipado de los genes *HLA-DQA1* y *HLA-DQB1* en la población CeliacaVa fue realizado mediante el software IGV y el número de homocigotos y heterocigotos para cada alelo de cada gen, la frecuencia para cada alelo en la población CeliacaVa y la frecuencia para cada alelo en los celiacos en la población CeliacaVa se muestran en las *Tablas 13 y 14*. Las frecuencias se realizaron con la población CeliacaVa total (221 pacientes) y con los pacientes diagnosticados de EC (78 pacientes).

Se llevó a cabo el test de Chi-cuadrado para ver si existían diferencias significativas entre las frecuencias alélicas de los celiacos de la población CeliacaVa y de la población control de Castilla y León. Los p.valores obtenidos también se muestran en las Tablas 13 y 14.

Tabla 13. Resultados genotipado del gen HLA-DQA1 en la población CeliacaVa y la población control. Los alelos de riesgo están marcados en negrita (DQA1*03 y DQA1*05).

	Población CeliacaVa				P. control Castilla y León			X^2 celiacos CeliacaVa y P.control
Alelos *DQA1*	**Hom totales**	**Het totales**	**Frecuencia alélica** no celiacos **(n^1)**	**Frecuencia alélica** en celiacos **(n^1)**	**Hom**	**Het**	**Frecuencia alélica (n^1)**	**p.valor**
DQA1*01 *	3	96	0,436 (68)	0,160 (25)	347	764	0,418 (1458)	0,0006 *
DQA1*02	2	80	0,314 (49)	0,199 (31)	63	427	0,159 (553)	0,504
DQA1*03	2	39	0,186 (29)	0,071 (11)	30	381	0,127 (441)	0,206
DQA1*04	0	9	0,038 (6)	0,013 (2)	3	71	0,022 (77)	-
DQA1*05 *	39	123	0,654 (102)	0,551 (86)	125	695	0,271 (945)	0,002*
DQA1*06	0	1	0,000 (0)	0,006 (1)	0	10	0,003 (10)	0,713

n^1: número de veces que aparece el alelo. Hom: homocigotos. Het: heterocigotos

Se observaron diferencias significativas entre las frecuencias alélicas de los alelos DQA1*01 y DQA1*05 entre las población celiaca y la población control ya que los p.valores fueron menores a 0,05.

Tabla 14. Resultados genotipado del gen HLA-DQB1 en la población CeliacaVa y la población control. SD: sin datos para ese alelo. Los alelos de riesgo están marcados en negrita (DQB1*02 y DQB1*03:02).

	Población CeliacaVa				P. control Castilla y León			X^2 celiacos y controles
Alelos *DQB1*	**Hom totales**	**Het totales**	**Frecuencia alélica** no celiacos **(n^1)**	**Frecuencia alélica** en celiacos **(n^1)**	**Hom**	**Het**	**Frecuencia alélica (n^1)**	**p.valor**
DQB1*02 *	47	135	0,763 (119)	0,622 (97)	144	643	0,267 (931)	0,0001*
DQB1*03	9	78	0,385 (60)	0,192 (30)	158	741	0,303(1057)	0,115
DQB1*03:01	6	56	0,269 (42)	0,141 (22)	SD	SD	SD	-
DQB1*03:02	1	20	0,083 (13)	0,045 (7)	SD	SD	SD	-
DQB1*03:03	0	4	0,019 (3)	0,006 (1)	SD	SD	SD	-
DQB1*03:05	0	1	0,006 (1)	0,000 (0)	SD	SD	SD	-
DQB1*04	0	10	0,038 (6)	0,019 (3)	1	93	0,027 (95)	-
DQB1*05 *	0	50	0,199 (31)	0,083 (13)	69	520	0,189 (658)	0,04*
DQB1*06 *	1	53	0,244(38)	0,083 (13)	97	549	0,213 (743)	0,01*

n^1: número de veces que aparece el alelo. Hom: homocigotos. Het: heterocigotos

Se observaron diferencias significativas entre las frecuencias alélicas de los alelos DQB1*02, DQB1*05 y DQB1*06 entre las población celiaca y la población control ya que los p.valores fueron menores a 0,05.

Además, todos los pacientes diagnosticados con EC de la población CeliacaVa tienen al menos un alelo de riesgo en su genotipo. 6 pacientes tienen un alelo de riesgo, 32 pacientes tienen dos

alelos de riesgo, 28 pacientes tienen tres alelos de riesgo y 12 pacientes tienen 4 alelos de riesgo en su genotipo, dos para cada gen.

Para comprobar si existe relación entre la celiaquía y las combinaciones de haplotipos se realizó una regresión y un ajuste de Benjamin-Hochberg. Los resultados se muestran en la *Tabla 15*.

Tabla 15. Resultados de la regresión logística. Se muestran las combinaciones de haplotipos que han resultado con un coeficiente diferente a 0, el p.valor y el p.valor ajustado (p.valor aj.) para cada combinación de haplotipos.

Términos de interacción	Coeficientes	p.valor	p.valor aj.	Términos de interacción	Coeficientes	p.valor	p.valor aj.
intercept	1,545	0,023	0,145	**DQA1*01/DQA1*05:05**	-2,638	0,016	0,145
DQB1*02/DQB1*03:01	0,400	0,437	0,546	**DQA1*02/DQA1*01**	-1,545	0,061	0,145
DQB1*02/DQB1*03:02	0,205	0,774	0,821	**DQA1*02/DQA1*02**	-1,545	0,040	0,145
DQB1*02/DQB1*03:03	-1,545	0,061	0,145	**DQA1*02/DQA1*03:01**	-1,417	0,062	0,145
DQB1*02/DQB1*04	-0,545	0,506	0,590	**DQA1*02/DQA1*03:03**	-1,945	0,046	0,145
DQB1*02/DQB1*05	1,405	0,103	0,200	**DQA1*02/DQA1*04**	-1,000	0,133	0,212
DQB1*02/DQB1*06	-0,545	0,257	0,345	**DQA1*02/DQA1*05:01**	-0,795	0,251	0,345
DQB1*03:01/DQB1*02	0,400	0,634	0,704	**DQA1*03:01:01/DQA1*01**	-2,388	0,062	0,145
DQB1*03:01/DQB1*03:01	-1,345	0,057	0,145	**DQA1*03:01/DQA1*01**	-1,888	0,097	0,194
DQB1*03:01/DQB1*03:02	-0,545	0,506	0,590	**DQA1*03:01/DQA1*02**	-1,750	0,040	0,145
DQB1*03:01/DQB1*03:03	-0,545	0,506	0,590	**DQA1*03:03/DQA1*02**	-1,545	0,061	0,145
DQB1*03:01/DQB1*06	1,092	0,125	0,212	**DQA1*03:03/DQA1*05:01**	-1,945	0,046	0,145
DQB1*03:02/DQB1*02	-0,545	0,506	0,590	**DQA1*03/DQA1*03**	-1,945	0,046	0,145
DQB1*03:02/DQB1*03:02	-1,545	0,061	0,145	**DQA1*03/DQA1*05:05**	0,800	0,121	0,212
DQB1*03:02/DQB1*05	0,342	0,724	0,792	**DQA1*04/DQA1*02**	-2,4205E-15	1,000	1,000
DQB1*03/DQB1*06	0,342	0,739	0,796	**DQA1*04/DQA1*05:01**	-1,000	0,133	0,212
DQB1*04/DQB1*02	-0,545	0,586	0,672	**DQA1*05:01/DQA1*01**	-2,638	0,016	0,145
DQB1*04/DQB1*03:01	-1,545	0,061	0,145	**DQA1*05:01/DQA1*02**	-1,000	0,133	0,212
DQB1*05/DQB1*02	2,092	0,032	0,145	**DQA1*05:01/DQA1*03**	-1,750	0,040	0,145
DQB1*05/DQB1*03:01	1,342	0,088	0,181	**DQA1*05:01/DQA1*03:01**	-1,000	0,133	0,212
DQB1*05/DQB1*03:02	0,842	0,386	0,492	**DQA1*05:01/DQA1*03:02**	-2,888	0,012	0,145
DQB1*05/DQB1*03:03	4,8527E-15	1,000	1,000	**DQA1*05:01/DQA1*03:03**	-1,545	0,061	0,145
DQB1*05/DQB1*03:05	1,092	0,262	0,345	**DQA1*05:01/DQA1*04**	-1,000	0,133	0,212
DQB1*05/DQB1*05	-1,545	0,061	0,145	**DQA1*05:01/DQA1*05:01**	-0,945	0,167	0,244
DQB1*05/DQB1*06	-1,545	0,061	0,145	**DQA1*05:01/DQA1*05:05**	-1,638	0,057	0,145
DQB1*06/DQB1*02	1,092	0,262	0,345	**DQA1*05:05/DQA1*01**	-2,638	0,009	0,145
DQB1*06/DQB1*03:01	-1,545	0,040	0,145	**DQA1*05:05/DQA1*02**	-1,279	0,139	0,216
DQB1*06/DQB1*03:03	1,092	0,217	0,309	**DQA1*05:05/DQA1*05:01**	-1,745	0,047	0,145
DQB1*06/DQB1*04	-1,545	0,061	0,145	**DQA1*05/DQA1*04**	-2,1971E-15	1,000	1,000
DQB1*06/DQB1*06	-1,545	0,061	0,145	**DQB1*02/DQB1*05 DQA1*01/DQA1*02**	-1,950	0,071	0,154
DQA1*01/DQA1*02	-1,000	0,084	0,177	**DQB1*02/DQB1*05 DQA1*01/DQA1*05:01**	-1,450	0,164	0,244
DQA1*01/DQA1*03:01	-2,638	0,016	0,145	**DQB1*02/DQB1*05 DQA1*02/DQA1*01**	-1,405	0,162	0,244
DQA1*01/DQA1*05:01	-1,000	0,067	0,150	**DQB1*02/DQB1*06 DQA1*02/DQA1*01**	0,745	0,279	0,362
DQB1*02/DQB1*03:02 DQA1*03:03/DQA1*02	-0,205	0,823	0,860	**DQB1*02/DQB1*06 DQA1*05:01/DQA1*01**	2,111	0,033	0,145
DQB1*05/DQB1*02 DQA1*02/DQA1*01	-1,759	0,115	0,212	**DQB1*02/DQB1*03:01 DQA1*03:03/DQA1*02**	-0,400	0,634	0,704

p.valor aj. significativo <0,05

Para la regresión logística, ningún p.valor ajustado ha sido menor a 0,05. Es decir, no hay una relación entre la variable respuesta (la EC) y las covariables (los haplotipos). Esto puede deberse a que el tamaño de la muestra de la población CeliacaVa no proporciona suficiente potencia estadística para poder observar diferencias.

El genotipado del gen *LCT* se realizó mediante el programa Mutation Detector. Todos los individuos analizados tuvieron genotipos CC, TT y GG para los SNPs rs41525747, rs41380347 y rs145946881 respectivamente. En la población control IBS de Ensembl las frecuencias alélicas y genotípicas fueron de 1,000 para los SNPs rs41380347 y rs145946881 y para el SNP rs41525747 la población control IBS no se recogía en Ensembl pero las frecuencias alélicas y genotípicas fueron de 1,000 en la población Non-Finnish European. Por lo tanto, estos tres SNPs no se incluyeron en los análisis estadísticos.

Las frecuencias genotípicas y alélicas de los SNPs rs4988235 y rs182549 se encuentran en las *Tablas 16 y 17* respectivamente. También se recoge el p.valor para el test chi cuadrado en el que se comprueba si hay diferencias significativas entre los celiacos de la población CeliacaVa y la población control.

Tabla 16. Frecuencias genotípicas de dos SNPs del gen LCT en la población CeliacaVa y en la población IBS de Ensembl. P.valor para el test estadístico chi cuadrado.

		Frecuencia genotípica (n[1])			p.valor
SNP	**Genotipo**	**No Celiacos P.CeliacaVa**	**Celiacos P. CeliacaVa**	**P. control Ensembl**	**X^2: celiacos y controles**
rs4988235	CC	0,487 (38)	0,397 (31)	0.318 (34)	0,348
	TC	0,756 (59)	0,449 (35)	0.449 (48)	0,998
	TT	0,372 (29)	0,154 (12)	0.234 (25)	0,198
rs182549	AA	0,382 (29)	0,382 (29)	0.234 (25)	0,060
	GA	0,763 (58)	0,461 (35)	0.449 (48)	0,904
	GG	0,487 (37)	0,158 (12)	0.318 (34)	0,020 *

n1: número de individuos con el genotipo
p.valor significativo <0,05

Tabla 17. Frecuencias alélicas de dos SNPs del gen LCT en la población CeliacaVa y en la población IBS de Ensembl. P.valor para el test estadístico chi cuadrado.

		Frecuencia alélica (n[1])			p.valor
SNP	**Alelo**	**No Celiacos P.CeliacaVa**	**Celiacos P. CeliacaVa**	**P. control Ensembl**	**X^2: celiacos y controles**
rs4988235	C	0,536 (135)	0,662 (97)	0.542 (116)	0,460
	T	0,464 (117)	0,378 (59)	0.458 (98)	0,383
rs182549	A	0,468 (116)	0,612 (93)	0.458 (98)	0,137
	G	0,532 (132)	0,388 (59)	0.542 (116)	0,111

n1: número de individuos con el alelo
p.valor significativo <0,05

Las diferencias en las frecuencias genotípicas no son significativas entre los celiacos de la población CeliacaVa y la población control excepto para el genotipo GG del SNP rs182549 en el que el p.valor es menor a 0,05.

Por otro lado, las diferencias en las frecuencias alélicas para estas poblaciones no son significativas ya que los p.valores para cada alelo y para cada genotipo son mayores de 0,05.

Los resultados de las frecuencias genotípicas difieren a los del estudio de Kuchay y colaboradores (2015) en el que las frecuencias en celiacos para en el SNP rs4988235 para el genotipo CC y CT son de 61,5 y 38,4 respectivamente mientras que para el genotipo TT es 0, frecuencias similares a su población control. Al igual que en el trabajo de Basso y colaboradores (2012) en el que las frecuencias genotípicas e SNP rs4988235 para el genotipo CC y CT son de 77,2 y 20,7 mientras que para el genotipo TT es 2,1. En el SNP rs182549 las frecuencias genotípicas GG y GA son 71,7 y 26,2 respectivamente, mientras que para el genotipo AA es de 2,2. Las frecuencias en la población control también son similares. Las diferencias entre las frecuencias genotípicas de los estudios y de nuestro trabajo pueden ser debidas al origen de las poblaciones estudiadas ya que, la del estudio de Kuchay se estudia una población del norte de India mientras que la del trabajo de Basso recoge una población italiana. Sin embargo, en ninguno se ven diferencias significativas para estos SNPs entre población celiaca y la población control, al igual que en nuestro estudio.

Tras asumir una herencia dominante, recesiva, codominante, sobredominante y aditiva en modelos separados, los correspondientes p.valores obtenidos se encuentran en la *Tabla 18*.

Tabla 18: Determinación del modelo de herencia que mejor se ajusta a la asociación de los SNP de LCT con la EC en función de los p.valores.

	modelo				
	codominante	dominante	recesivo	sobredominante	log_aditivo
SNP	**p.valor**				
rs4988235	0,261	0,275	0,203	0,851	0,656
rs182549	0,330	0,292	0,286	0,992	0,589

p.valor significativo <0,05

Según el estadístico aplicado, no existe relación entre la EC y los SNPs rs4988235 y rs182549 para ninguno de los modelos ya que los valores para cada modelo de herencia son mayores de 0,05.

4.4. ÁRBOLES FAMILIARES

Algunos de los árboles familiares diseñados con la herramienta PowerPoint se muestran en las *Figuras 8 y 9.* Los demás árboles se muestran en el anexo

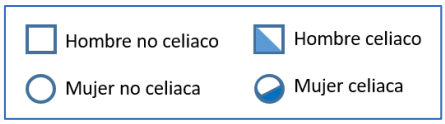

Figura 9. Leyenda para los árboles familiares de las Figuras 9 y 10.

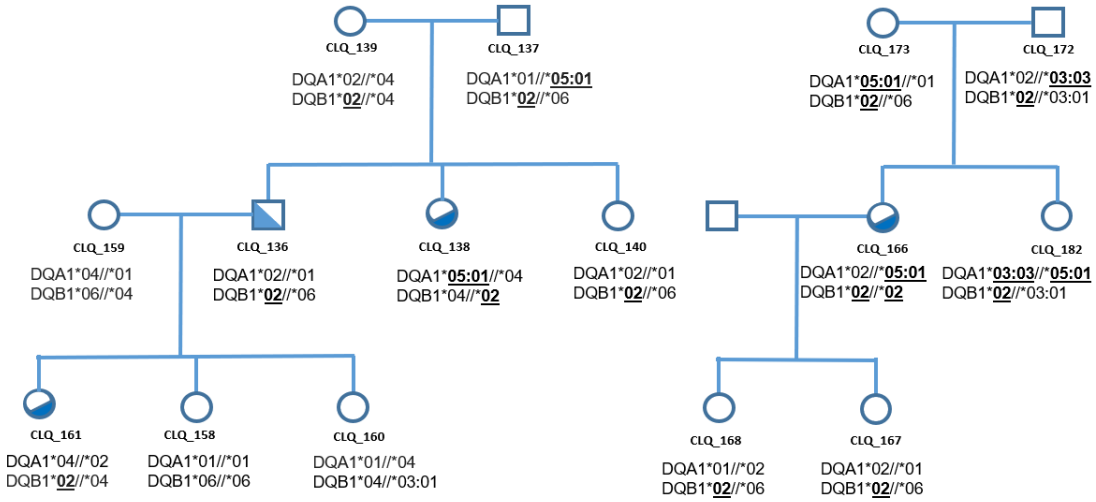

Figura 10. Árboles familiares con los genotipos DQA1 y DQB1 de dos familias de la población CeliacaVa. Los alelos de riesgo están marcados en negrita y subrayado.

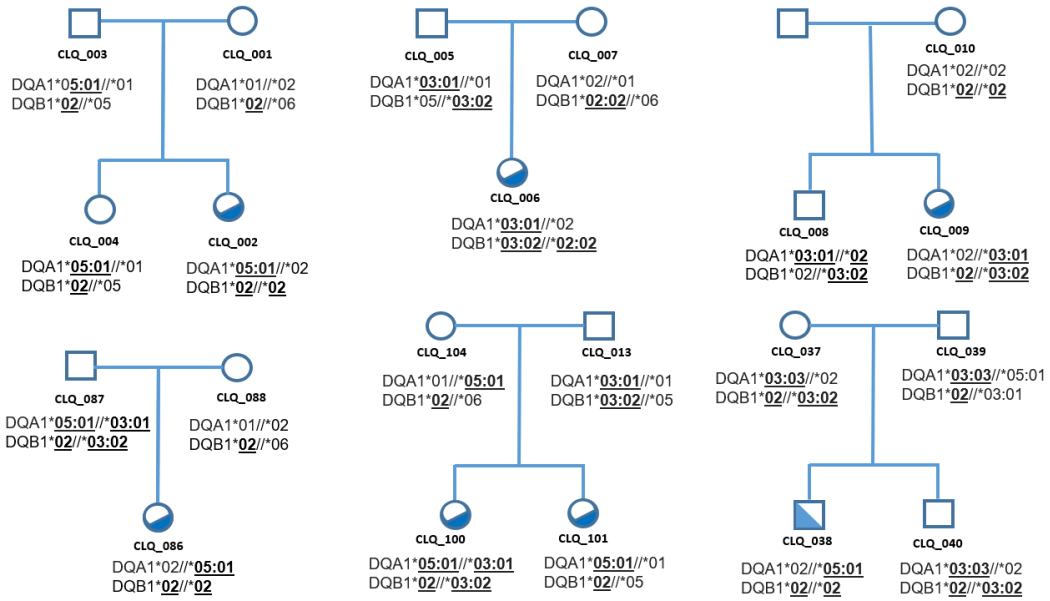

Figura 11. Árboles familiares con los genotipos DQA1 y DQB1 de seis familias de la población CeliacaVa. Los alelos de riesgo están marcados en negrita y subrayado.

4.5. LIMITACIONES DEL ESTUDIO

Una limitación del presente trabajo es el tamaño de la muestra de la población CeliacaVa ya que solo contaba con 221 individuos, 78 de los cuales eran celiacos. Además, estas muestras están relacionadas ya que forman parte de familias, por lo que sería conveniente aumentar el número de muestras no emparentadas.

La metodología utilizada no permite diferenciar si los alelos de riesgo se encuentran en cis o en trans para poder tipificar la molécula DQ, ya que no de todos los pacientes celiacos disponemos datos de sus padres.

También se debe continuar con la optimización del proceso de creación de librerías para poder equilibrar la cobertura de los amplicones, especialmente para cuando se trabaje con un número mayor de muestras, y así obtener mejores resultados.

Por otro lado, es necesario continuar con el trabajo bioinformático para mejorar *la pipeline* de análisis para los haplotipos de las regiones HLA, ya que, la herramienta IGV resulta útil pero ralentiza el trabajo.

5. CONCLUSIONES

1. El panel de amplicones NGS fue diseñado y aplicado con éxito, ya que los controles de calidad fueron superados. Gracias a este panel se pudieron secuenciar todos los individuos de una población española de enfermos celiacos y sus familiares, de los que se obtuvieron exitosamente regiones de interés de los genes *LCT, HLA-DQA1 y HLA-DQB1.*
2. Gracias a los datos obtenidos a partir de la secuenciación, se pudo estudiar la población CeliacaVa y compararla con población general. Se obtuvieron las frecuencias para los distintos haplotipos de los genes *HLA-DQA1 y HLA-DQB1,* pero no se encontraron diferencias significativas al compararlos con la población control. También se obtuvieron las frecuencias para cinco SNPs del gen *LCT* relacionados con la LNP. Se encontraron diferencias significativas entre las frecuencias de celiacos y de la población control para el genotipo GG del SNP rs182549. También se observaron diferencias significativas entre las frecuencias alélicas de los alelos DQA1*01, DQA1*05, DQB1*02, DQB1*05 y DQB1*06 entre la población celiaca y la población control. Aun así, estos resultados son preliminares y hay que tener en cuenta el pequeño tamaño de la muestra. Por lo tanto, es necesario continuar trabajando para validar estos resultados con otras poblaciones que puedan proporcionar una mayor fuerza estadística.
3. Los árboles familiares fueron completados con los genotipos del gen HLA.
4. En el futuro, será necesario equilibrar la amplificación de los diferentes amplicones para mejorar el rendimiento de las librerías, reducir los requerimientos en número de lecturas y reducir costes para trabajar en estudios con un número mayor de muestras. También sería necesario seguir trabajando para tipificar las moléculas DQ especialmente en los individuos celiacos.

6. BIBLIOGRAFIA

Alkalay M. J. (2021). Nutrition in Patients with Lactose Malabsorption, Celiac Disease, and Related Disorders. *Nutrients*, *14*(1), 2. https://doi.org/10.3390/nu14010002

Andrén Aronsson, C., Lee, H.-S., Hård Af Segerstad, E. M., Uusitalo, U., Yang, J., Koletzko, S., … TEDDY Study Group. (2019). Association of gluten intake during the first 5 years of life with incidence of celiac disease autoimmunity and celiac disease among children at increased risk. *JAMA: The Journal of the American Medical Association*, *322*(6), 514-523. doi:10.1001/jama.2019.10329

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Anguita-Ruiz, A., Aguilera, C. M., & Gil, Á. (2020). Genetics of lactose intolerance: An updated review and online interactive world maps of phenotype and genotype frequencies. *Nutrients*, *12*(9), 2689. doi:10.3390/nu12092689

Assa, A., Frenkel-Nir, Y., Tzur, D., Katz, L. H., & Shamir, R. (2017). Large population study shows that adolescents with celiac disease have an increased risk of multiple autoimmune and nonautoimmune comorbidities. Acta Paediatrica (Oslo, Norway: 1992), 106(6), 967-972. doi:10.1111/apa.13808

Auricchio, R., & Troncone, R. (2021). Can celiac disease be prevented? *Frontiers in Immunology*, *12*, 672148. doi:10.3389/fimmu.2021.672148

Basso, M. S., Luciano, R., Ferretti, F., Muraca, M., Panetta, F., Bracci, F., … Diamanti, A. (2012). Association between celiac disease and primary lactase deficiency. *European Journal of Clinical Nutrition*, *66*(12), 1364-1365. doi:10.1038/ejcn.2012.153

Brown, J., Pirrung, M., & McCue, L. A. (2017). FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics (Oxford, England)*, *33*(19), 3137–3139. https://doi.org/10.1093/bioinformatics/btx373

Brown, N. K., Guandalini, S., Semrad, C., & Kupfer, S. S. (2019). A clinician's guide to celiac disease HLA genetics. *The American Journal of Gastroenterology*, *114*(10), 1587-1592. doi:10.14309/ajg.0000000000000310

Caio, G., Volta, U., Sapone, A., Leffler, D. A., De Giorgio, R., Catassi, C., & Fasano, A. (2019). Celiac disease: a comprehensive current review. *BMC Medicine*, *17*(1), 142. doi:10.1186/s12916-019-1380-z

Catanzaro, R., Sciuto, M., & Marotta, F. (2021). Lactose intolerance: An update on its pathogenesis, diagnosis, and treatment. *Nutrition Research (New York, N.Y.)*, *89*, 23-34. doi:10.1016/j.nutres.2021.02.003

Catassi, C., Verdu, E. F., Bai, J. C., & Lionetti, E. (2022). Coeliac disease. *Lancet*, 399(10344), 2413-2426. doi:10.1016/s0140-6736(22)00794-2

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., … Flicek, P. (2022). Ensembl 2022. *Nucleic Acids Research*, 50(D1), D988-D995. doi:10.1093/nar/gkab1049

Dieli-Crimi, R., Cénit, M. C., & Núñez, C. (2015). The genetics of celiac disease: A comprehensive review of clinical implications. *Journal of Autoimmunity*, *64*, 26–41. doi:10.1016/j.jaut.2015.07.003

Dore, M. P., Pes, G. M., Dettori, I., Villanacci, V., Manca, A., & Realdi, G. (2017). Clinical and genetic profile of patients with seronegative coeliac disease: the natural history and response to gluten-free diet. *BMJ Open Gastroenterology*, *4*(1), e000159. doi:10.1136/bmjgast-2017-000159

Dunne, M. R., Byrne, G., Chirdo, F. G., & Feighery, C. (2020). Coeliac disease pathogenesis: The uncertainties of a well-known immune mediated disorder. *Frontiers in Immunology*, *11*, 1374. doi:10.3389/fimmu.2020.01374

Enfermedad celiaca. (s. f.). Recuperado 2 de marzo de 2023, de Seghnp.org website: https://www.seghnp.org/familias/enfermedad-celiaca

Espino, L., & Núñez, C. (2021). The HLA complex and coeliac disease. *International Review of Cell and Molecular Biology*, *358*, 47–83. doi:10.1016/bs.ircmb.2020.09.009

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics (Oxford, England)*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Flores Monar, G. V., Islam, H., Puttagunta, S. M., Islam, R., Kundu, S., Jha, S. B., … Sange, I. (2022). Association between type 1 diabetes mellitus and celiac disease: Autoimmune disorders with a shared genetic background. *Cureus*, *14*(3), e22912. doi:10.7759/cureus.22912

Gonzalez-Galarza, F. F., McCabe, A., Santos, E. J. M. D., Jones, J., Takeshita, L., Ortega-Rivera, N. D., Cid-Pavon, G. M. D., Ramsbottom, K., Ghattaoraya, G., Alfirevic, A., Middleton, D., & Jones, A. R. (2020). Allele frequency net database (AFND) 2020 update: gold-standard data classification,

open access genotype data and new query tools. *Nucleic acids research*, *48*(D1), D783–D788. https://doi.org/10.1093/nar/gkz1029

Greenbaum, C. J., Eisenbarth, G., Atkinson, M., Yu, L., Babu, S., Schatz, D., … DPT-1 study group. (2005). High frequency of abnormal glucose tolerance in DQA1*0102/DQB1*0602 relatives identified as part of the Diabetes Prevention Trial?Type 1 Diabetes. *Diabetologia*, *48*(1), 68-74. doi:10.1007/s00125-004-1608-z

Husby, S., Koletzko, S., Korponay-Szabó, I., Kurppa, K., Mearin, M. L., Ribes-Koninckx, C., … Wessels, M. (2020). European society paediatric gastroenterology, hepatology and nutrition guidelines for diagnosing coeliac disease 2020. *Journal of Pediatric Gastroenterology and Nutrition*, *70*(1), 141–156. doi:10.1097/MPG.0000000000002497

Itoh, Y., Mizuki, N., Shimada, T., Azuma, F., Itakura, M., Kashiwase, K., Kikkawa, E., Kulski, J. K., Satake, M., & Inoko, H. (2005). High-throughput DNA typing of HLA-A, -B, -C, and -DRB1 loci by a PCR-SSOP-Luminex method in the Japanese population. *Immunogenetics*, *57*(10), 717–729. https://doi.org/10.1007/s00251-005-0048-3

James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. Integrative Genomics Viewer. Nature Biotechnology 29, 24–26 (2011)

Jansen, H., Willenborg, C., Schlesinger, S., Ferrario, P. G., König, I. R., Erdmann, J., Samani, N. J., Lieb, W., & Schunkert, H. (2015). Genetic variants associated with celiac disease and the risk for coronary artery disease. Molecular genetics and genomics : MGG, 290(5), 1911–1917. https://doi.org/10.1007/s00438-015-1045-3

Kuchay, R. A., Thapa, B. R., Mahmood, A., Anwar, M., & Mahmood, S. (2015). Lactase genetic polymorphisms and coeliac disease in children: a cohort study. *Annals of human biology*, *42*(1), 101–104. https://doi.org/10.3109/03014460.2014.944216

Kylökäs, A., Kaukinen, K., Huhtala, H., Collin, P., Mäki, M., & Kurppa, K. (2016). Type 1 and type 2 diabetes in celiac disease: prevalence and effect on clinical and histological presentation. *BMC Gastroenterology*, *16*(1), 76. doi:10.1186/s12876-016-0488-2

Lebwohl, B., Sanders, D. S., & Green, P. H. R. (2018). Coeliac disease. *Lancet*, *391*(10115), 70–81. doi:10.1016/s0140-6736(17)31796-8

Leffler, D. A., Green, P. H. R., & Fasano, A. (2015). Extraintestinal manifestations of coeliac disease. *Nature Reviews. Gastroenterology & Hepatology*, *12*(10), 561-571. doi:10.1038/nrgastro.2015.131

Li, H. (s. f.). *seqtk: Toolkit for processing sequences in FASTA/Q formats*.

Ma, Z.-J., Sun, P., Guo, G., Zhang, R., & Chen, L.-M. (2013). Association of the HLA-DQA1 and HLA-DQB1 alleles in type 2 diabetes mellitus and diabetic nephropathy in the Han ethnicity of China. *Journal of Diabetes Research*, *2013*, 452537. doi:10.1155/2013/452537

Marsh, S. G. E., Albert, E. D., Bodmer, W. F., Bontrop, R. E., Dupont, B., Erlich, H. A., … Trowsdale, J. (2010). Nomenclature for factors of the HLA system, 2010. *Tissue Antigens*, *75*(4), 291-455. doi:10.1111/j.1399-0039.2010.01466.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10. doi:10.14806/ej.17.1.200

Martínez-Ojinaga, E., Molina, M., Polanco, I., Urcelay, E., & Núñez, C. (2018). HLA-DQ distribution and risk assessment of celiac disease in a Spanish center. *Revista espanola de enfermedades digestivas: organo oficial de la Sociedad Espanola de Patologia Digestiva*, *110*. doi:10.17235/reed.2018.5399/2017

Mattar, R., de Campos Mazo, D. F., & Carrilho, F. J. (2012). Lactose intolerance: diagnosis, genetic, and clinical factors. *Clinical and Experimental Gastroenterology*, *5*, 113-121. doi:10.2147/CEG.S32368

Megiorni, F., & Pizzuti, A. (2012). HLA-DQA1 and HLA-DQB1 in Celiac disease predisposition: practical implications of the HLA molecular typing. *Journal of Biomedical Science*, *19*(1), 88. doi:10.1186/1423-0127-19-88

Mir, K., Neuhaus, K., Bossert, M., & Schober, S. (2013). Short barcodes for next generation sequencing. *PloS one*, *8*(12), e82933. https://doi.org/10.1371/journal.pone.0082933

Misselwitz, B., Butter, M., Verbeke, K., & Fox, M. R. (2019). Update on lactose malabsorption and intolerance: pathogenesis, diagnosis and clinical management. *Gut*, *68*(11), 2080-2091. doi:10.1136/gutjnl-2019-318404

Ojetti, V., Nucera, G., Migneco, A., Gabrielli, M., Lauritano, C., Danese, S., … Gasbarrini, A. (2005). High prevalence of celiac disease in patients with lactose intolerance. Digestion, 71(2), 106-110. doi:10.1159/000084526

Pisapia, L., Camarca, A., Picascia, S., Bassi, V., Barba, P., Del Pozzo, G., & Gianfrani, C. (2016). HLA-DQ2.5 genes associated with celiac disease risk are preferentially expressed with respect to non-predisposing HLA genes: Implication for anti-gluten T cell response. *Journal of Autoimmunity*, *70*, 63-72. doi:10.1016/j.jaut.2016.03.016

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. *IPD-IMGT/HLA Database Nucleic Acids Research* (2020) https://hla.alleles.org/nomenclature/index.html

Signorell, A. (2023). DescTools: Tools for descriptive statistics. R package version 0.99.48.

Szilagyi, A., & Ishayek, N. (2018). Lactose intolerance, dairy avoidance, and treatment options. *Nutrients*, *10*(12), 1994. doi:10.3390/nu10121994

Thomas, G. W. C., & Hahn, M. W. (2019). Referee: Reference Assembly Quality Scores. *Genome biology and evolution*, *11*(5), 1483–1486. https://doi.org/10.1093/gbe/evz088

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic acids research*, *40*(15), e115. https://doi.org/10.1093/nar/gks596

Usai-Satta, P., Lai, M., & Oppia, F. (2022). Lactose malabsorption and presumed related disorders: A review of current evidence. *Nutrients*, *14*(3), 584. doi:10.3390/nu14030584

Vader, W., Stepniak, D., Kooy, Y., Mearin, L., Thompson, A., van Rood, J. J., … Koning, F. (2003). The HLA-DQ2 gene dose effect in celiac disease is directly related to the magnitude and breadth of gluten-specific T cell responses. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(21), 12390–12395. doi:10.1073/pnas.2135229100

7. ANEXO

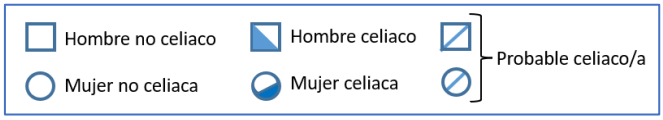

Figura 12. Leyenda para los árboles familiares de las Figuras 13, 14, 15, 16, 17, 18, 19, 20, 21.

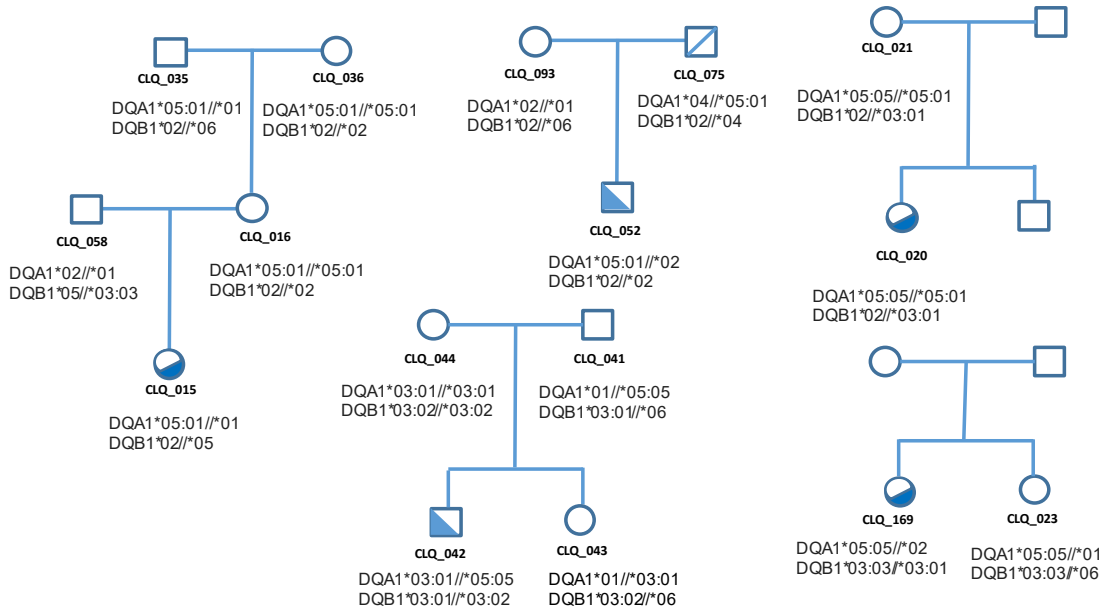

Figura 13.Árboles familiares con los genotipos DQA1 y DQB1 de cinco familias de la población CeliacaVa.

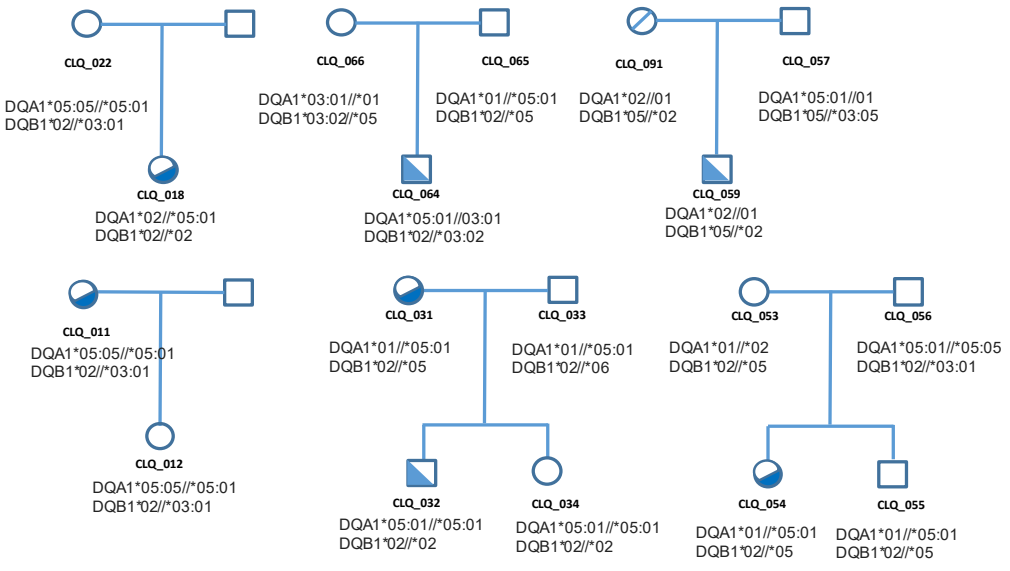

Figura 14. Árboles familiares con los genotipos DQA1 y DQB1 de seis familias de la población CeliacaVa.

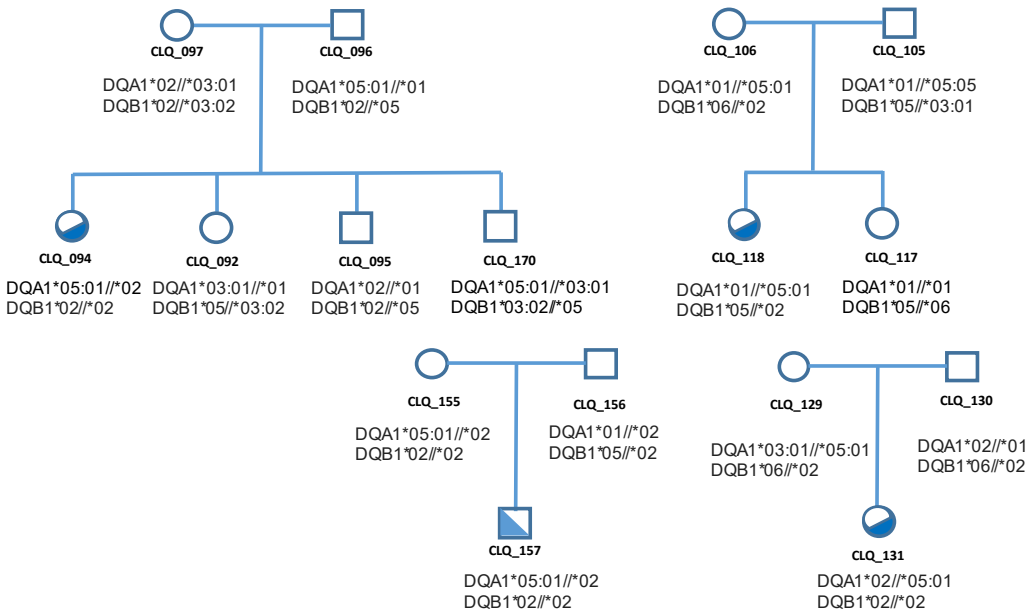

Figura 15. Árboles familiares con los genotipos DQA1 y DQB1 de cuatro familias de la población CeliacaVa.

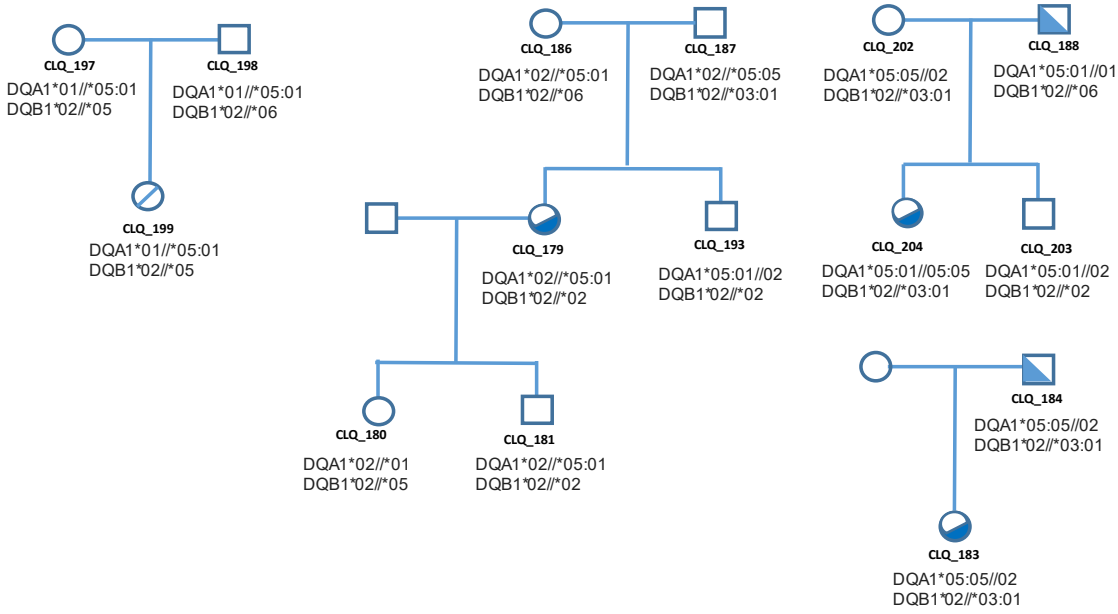

Figura 16. Árboles familiares con los genotipos DQA1 y DQB1 de cuatro familias de la población CeliacaVa.

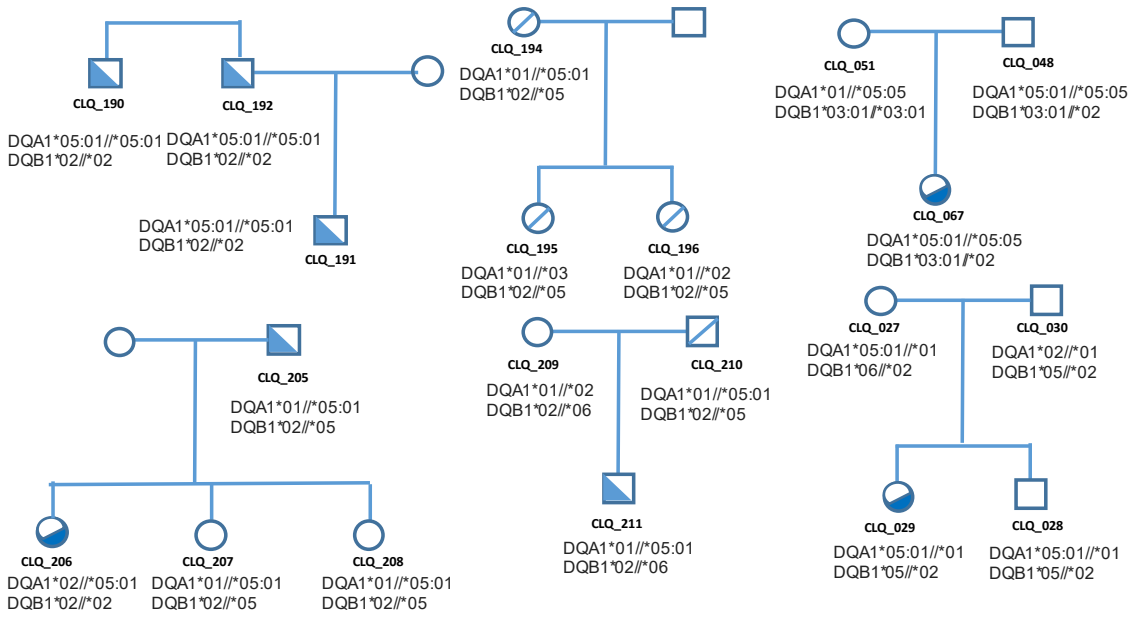

Figura 17. Árboles familiares con los genotipos DQA1 y DQB1 de seis familias de la población CeliacaVa.

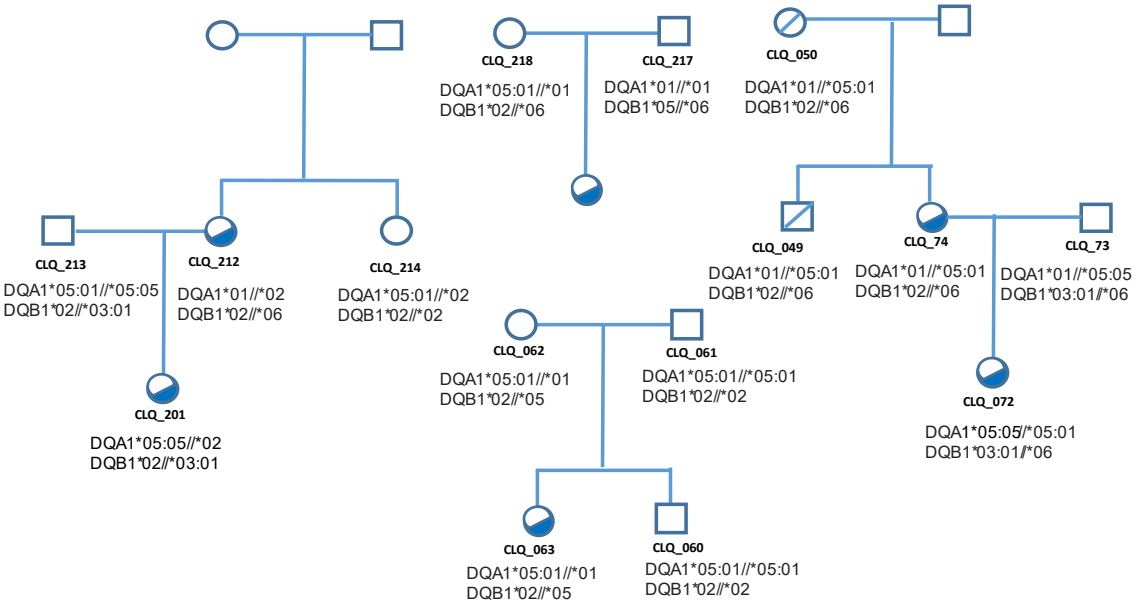

Figura 18. Árboles familiares con los genotipos DQA1 y DQB1 de cuatro familias de la población CeliacaVa.

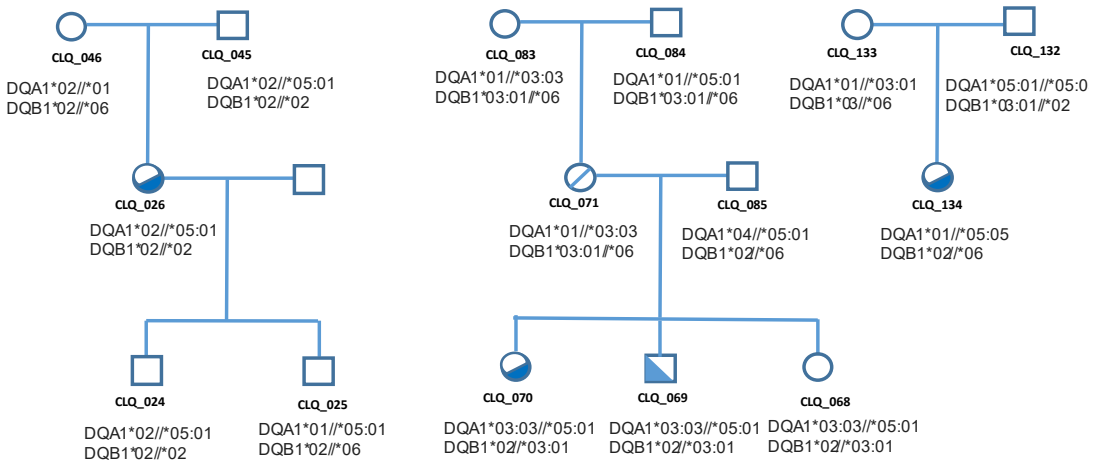

Figura 19. Árboles familiares con los genotipos DQA1 y DQB1 de tres familias de la población CeliacaVa.

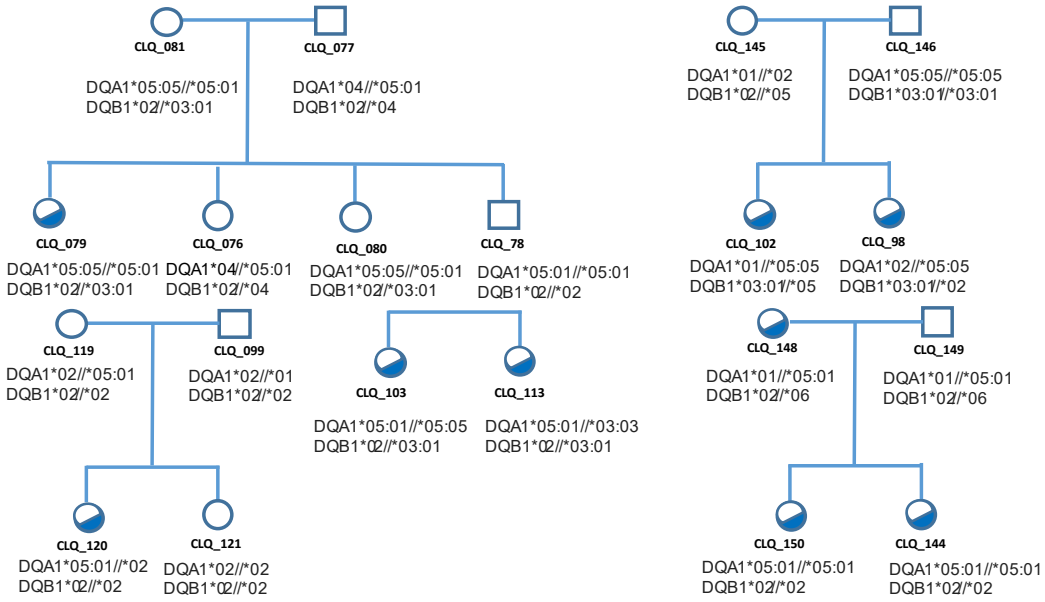

Figura 20. Árboles familiares con los genotipos DQA1 y DQB1 de cinco familias de la población CeliacaVa.

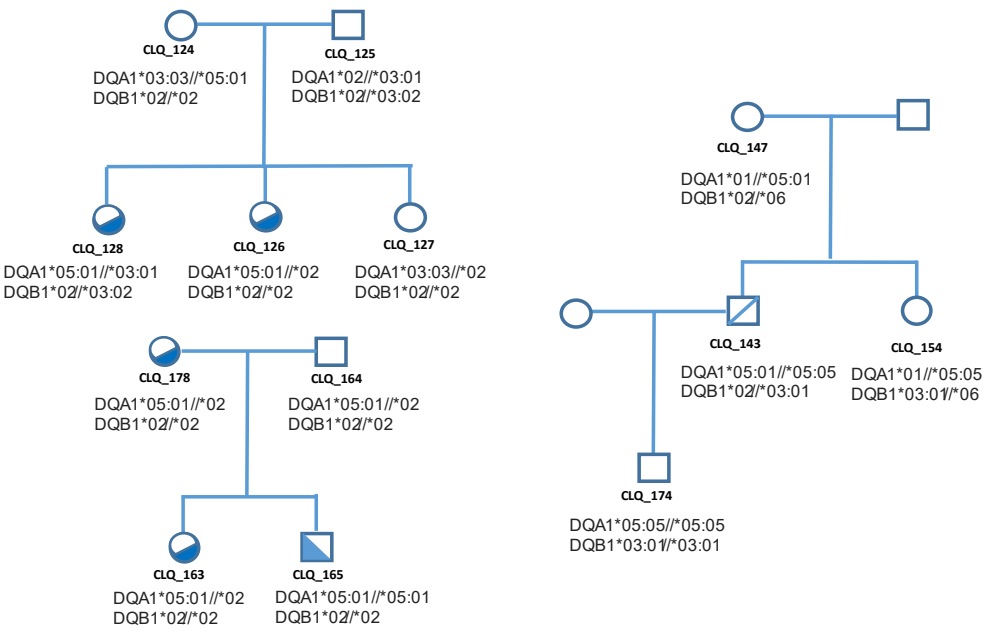

Figura 21. Árboles familiares con los genotipos DQA1 y DQB1 de tres familias de la población CeliacaVa.

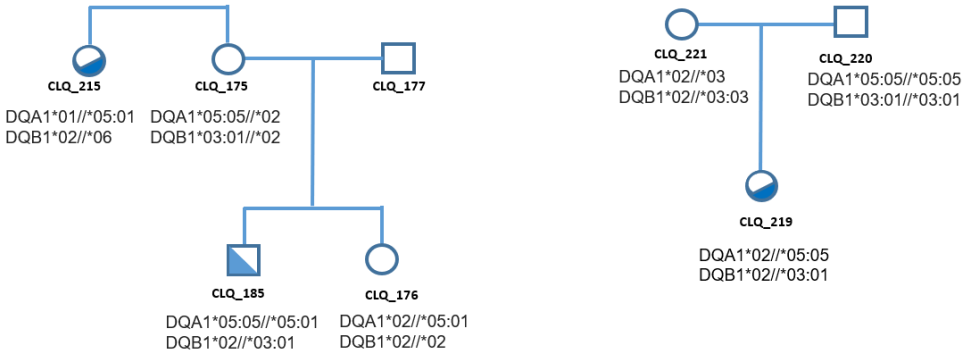

Figura 22. Árboles familiares con los genotipos DAQA1 y DQB1 de dos familias de la población CeliacaVa.