



Introducción a la biología molecular. La PCR y la secuenciación de Sanger.

Apellidos, nombre	Cardona Serrate, Fernando (fcardona@btc.upv.es)
Departamento	Departamento de Biotecnología. Escuela Técnica Superior de Ingeniería Agronómica y del Medio Natural.
Centro	Universitat Politècnica de València

1 Resumen de las ideas clave

En este artículo vamos a estudiar la Reacción en Cadena de la Polimerasa (PCR) y las técnicas relacionadas. Para entender la PCR es necesario conocer la estructura básica de los ácidos nucleicos, así como el proceso de replicación del ácido desoxirribonucleico (ADN). La PCR es una técnica muy potente, versátil y rápida de amplificación de ácidos nucleicos, por lo que facilita enormemente el trabajo con estas moléculas. Una de las principales aplicaciones de la PCR es la secuenciación de ácidos nucleicos, comenzando por el método de secuenciación de Sanger. Además, para poder amplificar y secuenciar el ácido ribonucleico (ARN), normalmente es necesario una conversión previa a DNA, lo que se hace mediante la retrotranscripción.

2 Objetivos

Una vez que el estudiante lea con detenimiento este documento, será capaz de:

- Comprender la estructura básica de los ácidos nucleicos (ADN y ARN)
- Explicar el proceso de replicación del ADN
- Explicar el concepto de retrotranscripción (RT) y la RT-PCR
- Comprender en detalle la PCR
- Comprender en detalle la electroforesis de ADN
- Comprender el método de Sanger para secuenciar ADN y sus características

3 Introducción

Los ácidos nucleicos están formados por nucleótidos, que a su vez están formados por la unión de: una base nitrogenada, un azúcar simple y un grupo fosfato (imagen 1).

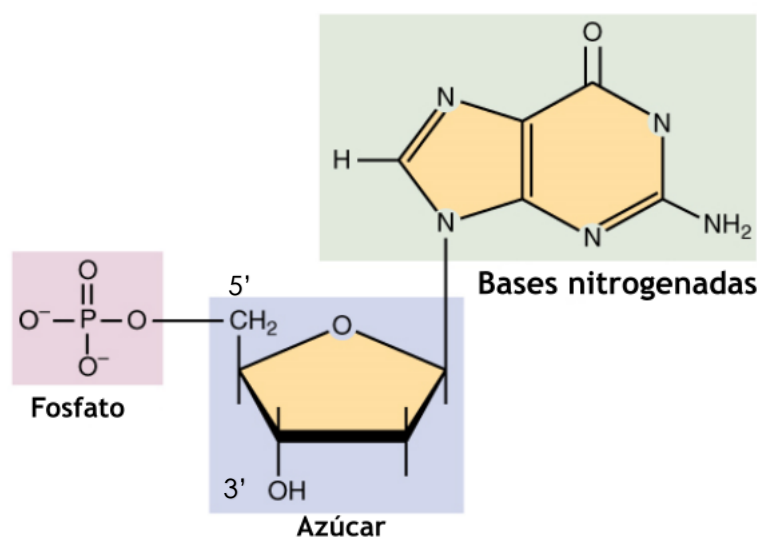


Imagen 1. Estructura de los nucleótidos. Fuente: Wikimedia Commons

La molécula de ADN (imagen 2) está compuesta por dos cadenas complementarias formando una doble hélice antiparalela de manera que se enrollan entre sí. Cada hebra tiene una estructura principal compuesta por grupos alternados de azúcar (desoxirribosa) y fosfato. Unida a cada azúcar hay una de cuatro bases: adenina (A), citosina (C), guanina (G) o timina (T). Las dos hebras se conectan por puentes de hidrógeno entre las bases: 2 enlaces para adenina con timina y 3 enlaces para citosina con guanina.

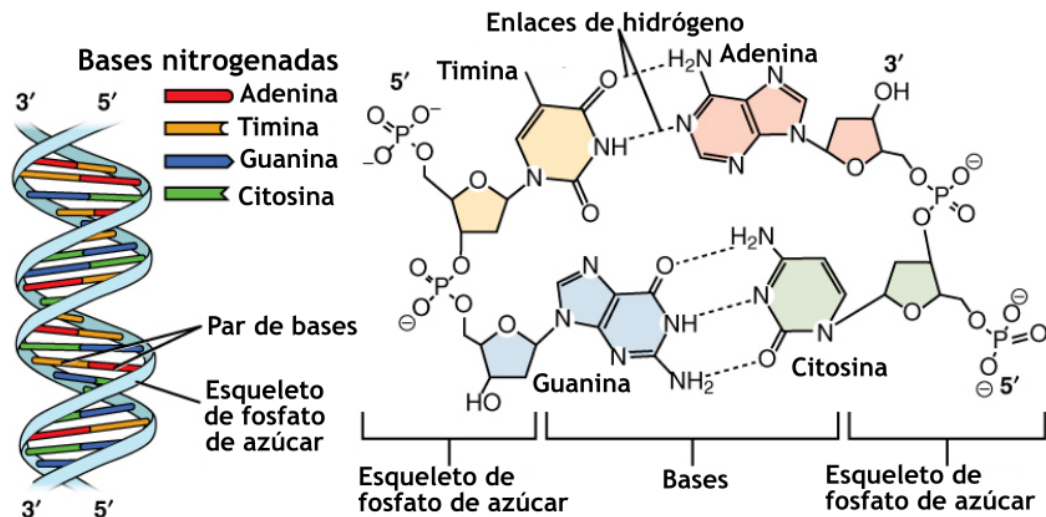


Imagen 2. Estructura de la doble hélice de ADN. Fuente: Wikimedia Commons

La reacción en cadena de la polimerasa (PCR) es una técnica ampliamente utilizada para obtener muchas copias de determinados fragmentos de ADN a partir de pequeñas cantidades de estos. Por otro lado, la PCR cuantitativa (qPCR) es una técnica ampliamente utilizada para cuantificar la cantidad de ADN presente en una muestra.

A diferencia del ADN, el ARN está formado por una única hebra que resulta de la transcripción de una de las hebras de ADN, y en lugar de timina (T) tiene la base uracilo (U). En el caso de los ARN mensajeros (ARNm), además tienen una cola de poli adeninas (cola poliA).

El dogma central de la biología molecular establece que, durante la expresión de un gen codificante de proteína, la información fluye de ADN \rightarrow ARNm \rightarrow proteína. El proceso de conversión de ADN a ARN se denomina transcripción, y el de ARN a proteína traducción. El proceso de traducción se hace convirtiendo tripletes de nucleótidos en aminoácidos, siguiendo el código genético. El código genético es el conjunto de reglas que define cómo se traduce una secuencia de nucleótidos en el ARN a una secuencia de aminoácidos en una proteína. Este código es común en todos los seres vivos, aunque hay pequeñas variaciones, y define la relación entre cada secuencia de tres nucleótidos, llamada codón, y cada aminoácido. Dado que existen 4 bases nitrogenadas en el ARN (A, G, C, U), existen 64 combinaciones posibles de tripletes, de las cuáles 61 codifican aminoácidos. Uno de ellos es el codón de inicio, AUG, que codifica además para el aminoácido metionina, y tres codifican sitios de parada (codones de stop UAA, UAG y UGA).

La retrotranscripción (RT) utiliza la actividad retrotranscriptasa para convertir ARN en ADN copia (cDNA), y puede hacerse utilizando poli timinas que hibridan con las poliA, o bien

utilizando primers específicos, o hexámeros que hibridan al azar en diferentes sitios del ARN (imagen 3). Esta técnica es ampliamente utilizada para determinar, por ejemplo, la expresión génica de un determinado gen a partir de su ARNm (RT-qPCR), o para secuenciar los transcritos que aparecen durante la transcripción (RT-PCR seguida de secuenciación).

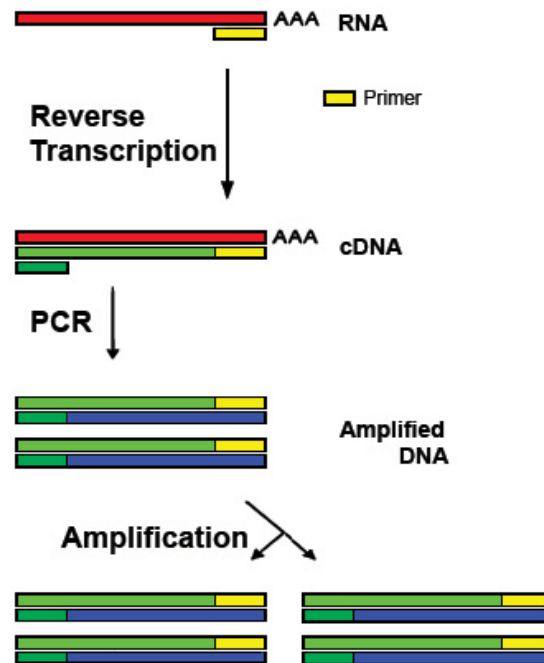


Imagen 3. La retrotranscripción. Fuente: Wikimedia Commons

La secuenciación de ADN ha sido revolucionaria en los campos de conocimiento de la biología molecular y la genética, ya que permite conocer la secuencia y estructura de los genes, así como las de otros elementos reguladores importantes en la regulación de la expresión génica. En el caso de humanos, es ampliamente conocido el proyecto genoma humano (1990-2001), que permitió conocer casi la totalidad de la secuencia del genoma humano. En este proyecto se utilizó el método de Sanger, conocido también primera generación de secuenciación [1].

4 Desarrollo

Durante la replicación de ADN, proceso por el cuál una célula duplica su genoma para dividirse. De esta manera, de una molécula de ADN única, se obtienen dos "réplicas". Esta duplicación del material genético se produce de acuerdo con un mecanismo semiconservador, lo que indica que las dobles cadenas generadas a partir del ADN molde original, al separarse, sirven de molde cada una para la síntesis de una nueva cadena complementaria de la cadena molde, de forma que cada nueva doble hélice contiene una de las cadenas del ADN original. Gracias a la complementación entre las bases que forman la secuencia de cada una de las cadenas, el ADN tiene la importante propiedad de reproducirse idénticamente, lo que permite que la información genética se transmita de una célula madre a las células hijas y es la base de la herencia del material genético.



La molécula de ADN se abre como una cremallera, por ruptura de los puentes de hidrógeno entre las bases complementarias en puntos determinados: los orígenes de replicación. Las proteínas iniciadoras reconocen secuencias de nucleótidos específicas en esos puntos y facilitan la fijación de otras proteínas que permitirán la separación de las dos hebras de ADN formándose una horquilla de replicación. La ADN polimerasa ADN dependiente añade nucleótidos a pequeñas secuencias de ADN denominados cebadores o *primers*, extendiéndolas copiando la hebra molde.

4.1 La reacción en cadena de la polimerasa

La reacción en cadena de la polimerasa (PCR) sirve para **hacer muchas copias (amplificar)** de un fragmento determinado de **ADN**. Con la PCR se amplifica (multiplica) ese trozo de ADN para que se pueda detectar. Sus aplicaciones en biología molecular, biomedicina y medicina son numerosas. La PCR puede usarse, como ejemplos, para identificar determinados cambios en un gen o cromosoma que ayudan a detectar y diagnosticar una enfermedad; para identificar trozos de ADN de determinadas bacterias, virus o microorganismos para diagnosticar una infección; o para identificar personas a partir de su ADN en aplicaciones forenses.

Fue desarrollada en 1986 por Kary Mullis [2], e Inicialmente la técnica era lenta, ya que las ADN polimerasas se desnaturalizaban al realizar los cambios de temperatura y era necesario agregar nuevas polimerasas en cada ciclo. Esto es debido a que las temperaturas de desnaturalización (95°C) durante la PCR suponen la inmediata desnaturalización de toda proteína. La verdadera revolución de la técnica, además de la aparición de los **termocicladores** que mantienen los tiempos y temperaturas necesarios en los ciclos de PCR, evitando el uso de baños termostáticos, fue el uso de **ADN polimerasas termoestables**, extraídas de microorganismos adaptados a vivir a altas temperaturas. Dichos microorganismos, generalmente arqueas, las más utilizadas son: *Thermus aquaticus* (Taq) y *Pyrococcus furiosus* (Pfu).

El ciclo del termociclador estándar, si bien puede variar dependiendo de la polimerasa y de la aplicación, podría ser:

Desnaturalización inicial: 95°/5'

1 -Desnaturalización: 95°C/30''

2 -Anillamiento de los cebadores (*annealing*): 50-65°C/30''

3 -Extensión: 72°C/1' por cada 1.000 nucleótidos (mínimo 45'')

} 35 ciclos

Extensión final: 72°C/10'

Las diferentes polimerasas tienen diferente procesividad y fidelidad de copia, siendo estas normalmente contrapuestas, es decir, a mayor procesividad menor fidelidad de copia y viceversa. También pueden emplearse mezclas de polimerasas muy procesivas (Taq) con otras de alta fidelidad de copia (Pfu). La **procesividad** es la cantidad de polimerización catalizada por una enzima cada vez que se une a un molde apropiado, o cebador-molde en el caso de las ADN polimerasas. Se mide en nucleótidos polimerizados por evento de unión. Las polimerasas de **alta fidelidad de copia** poseen la subunidad correctora de errores del

complejo enzimático de la ADN polimerasa, una exonucleasa 3' a 5', de forma que elimina los nucleótidos incorporados incorrectamente en 3', que es el extremo en el cual se produce la extensión. Cabe decir aquí que las polimerasas que no tienen esta subunidad, o incluso las que la tienen forzando las condiciones (aumentando a 30' la extensión final), incorporan adeninas de forma inespecífica al final de la cadena sintetizada de forma inespecífica (sin molde). Estas A "extras" son utilizadas en varias aplicaciones en biología molecular.

La adición de nucleótidos en la extensión se produce en **el grupo OH libre en 3'** (imagen 1 e imagen 4), y la amplificación **es exponencial**, de manera que una doble hélice da dos, que darán 4 en el siguiente ciclo (imagen 4), de manera que:

$$N = N_0 \cdot 2^c \cdot e$$

donde, N= no de copias de amplicón (fragmento amplificado)

N_0 = nº de copias de DNA molde

c = número de ciclos

e = eficacia de la reacción (e=1 en los primeros ciclos)

de forma que para $N_0=1$ y $c=35$ tenemos que $N= 68 \times 10^9$ copias por molécula

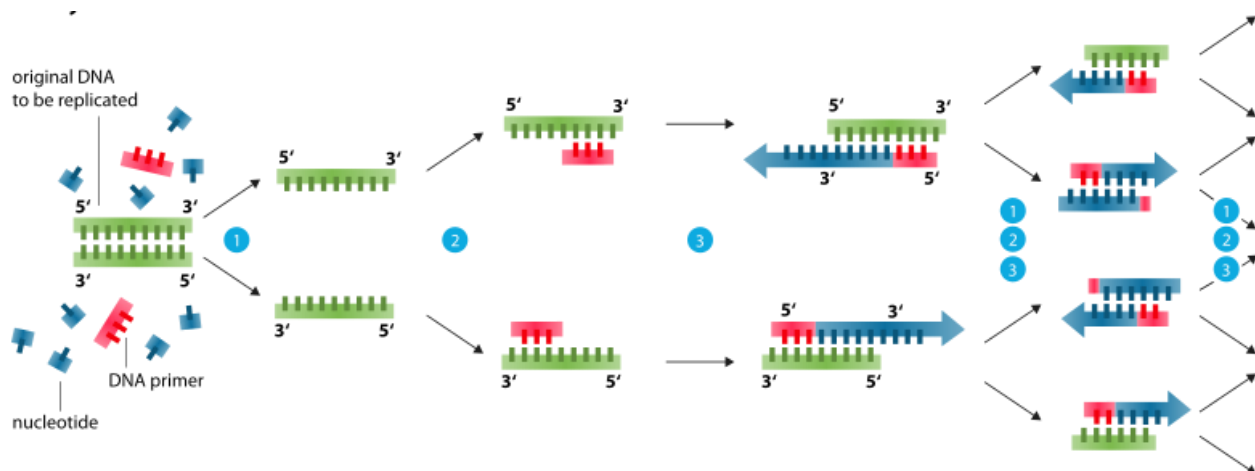


Imagen 4. Reacción en cadena de la polimerasa. Fuente: Wikimedia Commons

Como se observa en la imagen 4, el primer ciclo no da todavía el amplicón como tal, si no que la polimerasa sigue extendiendo desde el primer (directo) más allá del externo delimitado por el segundo primer, hasta que se agota su capacidad de extensión. Es en el segundo ciclo cuando esa hebra amplificada en el primer ciclo es delimitada por el otro externo por el primer complementario (reverso).

4.2 La electroforesis de ADN

La electroforesis de ADN es una técnica que permite separar las moléculas de ADN por tamaños. Dado que el ADN es un polianión debido a los grupos fosfato libres de los nucleótidos (imagen 1), al someter a un **campo eléctrico** a las moléculas situadas en una **matriz porosa** (por ejemplo, un gel de agarosa), las moléculas se van a mover hacia el ánodo (electrodo positivo) **proporcionalmente a su peso molecular** (número de nucleótidos). Así, las moléculas de ADN más grandes migrarán más lentamente en el gel y recorrerán menos distancia en el mismo tiempo que las pequeñas (imagen 5). Comparando su migración con la de un marcador de pesos moleculares (fragmentos de tamaño conocido), podemos conocer el tamaño de una mezcla de moléculas.

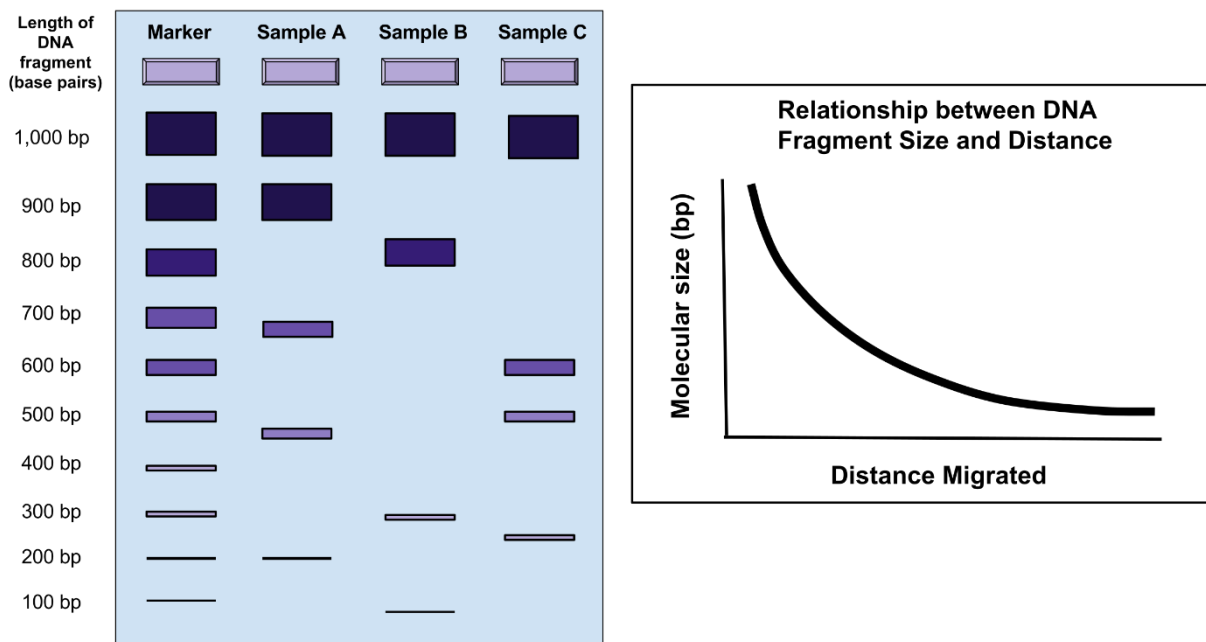


Imagen 5. Electroforesis de ADN y distancia desde el punto de aplicación. Fuente: Wikimedia Commons

4.3 La secuenciación de Sanger

La secuenciación de ADN por el método de Sanger [3] es el método de secuenciación de primera generación ampliamente utilizado. Con este método se secuenció el genoma humano en el Proyecto Genoma Humano [2]. Aunque actualmente no se utiliza para secuenciar genomas, sí es muy utilizada para secuenciar fragmentos relativamente pequeños de ADN, como pueden ser amplicones o plásmidos.

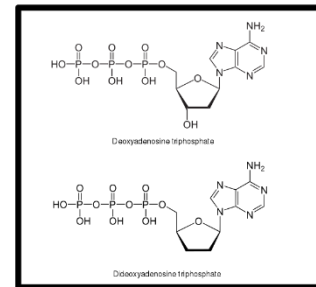
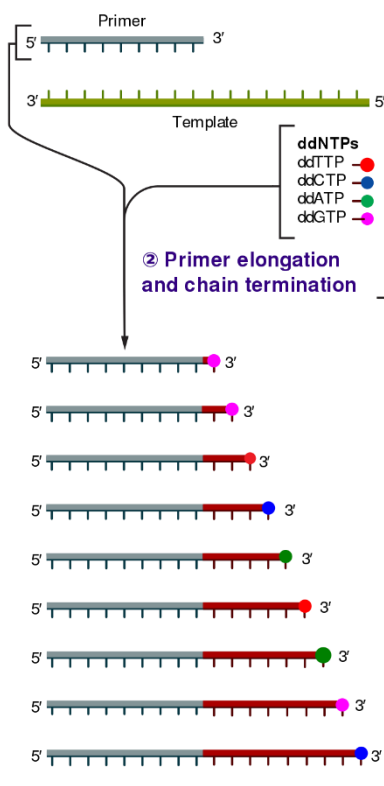
El método combina una PCR modificada con una electroforesis de ADN muy resolutive, capaz de diferenciar tamaños con diferencias de un solo nucleótido. La electroforesis se realiza en polímeros especiales en estado casi líquido y en capilares largos para aumentar la resolución.

La reacción de PCR modificada de Sanger (imagen 6) consiste en, utilizando un solo primer (se secuencian solo una hebra), añadir en la mezcla de reacción, además de los nucleótidos convencionales, unos nucleótidos modificados. Estos nucleótidos modificados están

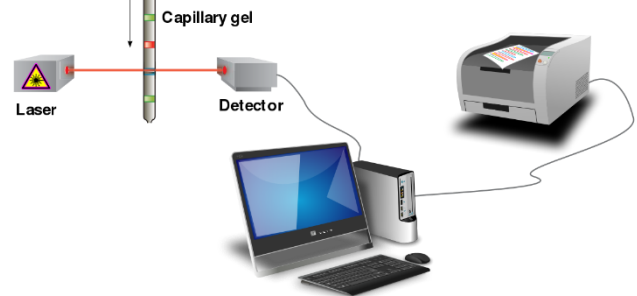
marcados con **colores** diferentes para cada base, y además carecen del OH libre en 3' que necesita la polimerasa para extender (nucleótidos **dideoxi**, imagen 6). La polimerasa añadirá nucleótidos hasta que por azar incorpore uno de estos nucleótidos modificados, momento en el que se para la reacción por no disponer de un OH libre para seguir extendiendo. De esta forma obtenemos **moléculas de diferente tamaño** (según el momento en el que se pare la copia) **marcadas terminalmente** con un color que indica que **nucleótido** ha sido el último en incorporarse (imagen 6).

① Reaction mixture

- ▶ Primer and DNA template ▶ DNA polymerase
- ▶ ddNTPs with flouochromes ▶ dNTPs (dATP, dCTP, dGTP, and dTTP)



③ Capillary gel electrophoresis separation of DNA fragments



④ Laser detection of flouochromes and computational sequence analysis

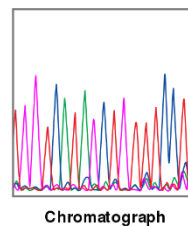


Imagen 6. Secuenciación por el método de Sanger

La electroforesis capilar de este ADN permite separar las moléculas por tamaños con diferencias de un solo nucleótido. La electroforesis se realiza en polímeros especiales en estado casi líquido y en capilares largos para aumentar la resolución, de forma que los fragmentos separados pasan por un detector láser que lee de qué color se trata. Dado que las moléculas pasarán por el detector láser, dependiendo de su tamaño, antes o después, obtenemos la secuencia de ADN. La molécula más pequeña (un solo nucleótido incorporado, se ha incorporado el dideoxi) pasará la primera, obteniéndose el primer nucleótido de la secuencia (G en la imagen 6), inmediatamente después la que ha incorporado 2 (uno normal y el 2º dideoxi), luego el 3º, etc., obteniéndose la secuencia de ADN.



Cabe decir que el límite de la técnica está en torno a 1.000 nucleótidos (suele incorporar antes un nucleótido dideoxi por azar), y que aproximadamente los 100 primeros no se leen bien por mala resolución electroforética. Al final de la secuencia puede observarse una A extra añadida por la polimerasa (ver apartado 4.1, principio de la página 5).

5 Cierre

A lo largo de este objeto de aprendizaje hemos introducido la estructura básica de los nucleótidos y los ácidos nucleicos, cómo se replica el ADN y el concepto de retrotranscripción. Posteriormente hemos visto en detalle la PCR, la electroforesis de ADN, y cómo la combinación de aplicaciones especiales de ambas técnicas permite la secuenciación de ADN por el método de Sanger.

6 Bibliografía

- 1- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*. **409**, 860–921 (2001). <https://doi.org/10.1038/35057062>
- 2- K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, H. Erlich. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol*. 1986:51 Pt 1:263-73. <https://doi.org/10.1101/sqb.1986.051.01.032>
- 3- F. Sanger, S. Nicklen, A. R. Coulson. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977 Dec; 74(12): 5463–5467. <https://doi.org/10.1073%2Fpnas.74.12.5463>