

genes

Cyrce Salinas

March 22, 2005

Bruno Vázquez Espinoza - A00838521

Cyrce Danae Salinas Rojas - A01666121

Genes y Genomas

Parte 1

En un organismo con un genoma de ADN doble hélice, si $G=C$ y $A=T$, entonces, ¿ $G+C = A+T$? ¿Por qué? En un genoma de ADN de doble hélice, Guanina (G) siempre se empareja con Citosina ©, y Adenina (A) siempre se empareja con Timina (T). Por lo tanto, G+C forman un par de bases complementarias, al igual que A+T. Sin embargo, G+C no es igual a A+T en términos de cantidad total de bases. La suma de G+C no es igual a la suma de A+T debido a las diferencias en la estructura de las bases nitrogenadas y los enlaces de hidrógeno entre ellas. En resumen, aunque los pares de bases son complementarios, sus cantidades totales no son iguales.

Si un organismo tiene 38% de A, ¿cuál es el porcentaje de G? Si un organismo tiene un 38% de adenina (A), según las reglas de apareamiento de bases nitrogenadas en el ADN, también tendrá un 38% de timina (T). Esto significa que la suma de A+T es igual al 76%. Como el ADN es una doble hélice, la suma de G+C y A+T debe ser igual al 100%. Dado que $G=C$, el porcentaje de guanina (G) también será del 12%.

¿Cuáles son las consecuencias de un alto contenido de GC? Mayor estabilidad: Los enlaces de hidrógeno entre G y C son más fuertes que los de A y T, lo que hace que el ADN con mayor contenido de GC sea más estable y resistente a la desnaturalización. Mayor temperatura de fusión: La T_m , que es la temperatura a la que se separa la doble hélice del ADN, es más alta en ADN con alto contenido de GC. Mayor densidad de información genética: Las secuencias con GC alto pueden codificar más información en una longitud dada debido a la mayor estabilidad de la doble hélice.

¿A qué refiere el superenrollamiento del ADN? El ADN superenrollado es una molécula de ADN bicatenario que forma una superhélice. En lugar de seguir una curva plana, la doble hélice se enrolla sobre sí misma. Este fenómeno es crucial para la compactación y regulación del ADN.

¿Cómo se desnaturaliza el ADN? Es decir, ¿qué métodos se usan para separan las hebras? La desnaturalización del ADN implica separar las dos hebras de la doble hélice. Se puede lograr mediante calentamiento, cambios de pH o mediante agentes químicos.

¿Qué es la Tm del ADN de doble cadena? La Tm es la temperatura a la que se desnaturaliza la mitad del ADN de una mezcla. Depende del contenido de GC y afecta la estabilidad y la hibridación del ADN.

¿Por qué es importante saber la Tm de un segmento de ADN de doble cadena? (desde el punto de vista biotecnológico) La Tm es crucial para diseñar cebadores en PCR hibridación y otras técnicas. Permite ajustar las condiciones para obtener resultados específicos.

¿Qué factores afectan la Tm? Menciona algunos factores que afectan la Tm. El contenido de GC, la longitud de la secuencia, la concentración de sal y otros factores influyen en la Tm.

Ejercicio: Observa la siguiente secuencia y encuentra la secuencia complementaria, la reversa y la reversa complementaria: 5'-ATGCTTGACGCTCAAACCATCGC-3'

- Secuencia complementaria: 3'-TACGAAGTTCGAGTTTGGTAGCG-5'
- Reversa: 5'-CGCTAGGACTCGAATTTGGATCG-3'
- Reversa complementaria: 3'-GCTATCCTGACGCTAAACCTAGC-5'

Parte2

Analicen la siguiente secuencia de una hebra del gene hmp (flavohemoglobina) de Escherichia coli, una bacteria mesófila, no patógena, enterobacteria y respondan: ¿cuál será la secuencia de aminoácidos de la flavohemoglobina? [20 pts]

LQHLMRAKAFVMNVFLLHAQIHstopLFRGKLHEANRAAENULLELHCRIAERTFQFAQIHQTFAIKLTFS AIGFARLTN

```
library(stringi)
library(seqinr)

DNASequences <- c('ATGCTTGACGCTCAAACCATCGCTACAGTAAAAGCCACCATCCCTTTACTGGTGGAAACGGGGCCAAAGTTAACCGCCATTCTT')

complementary_strand <- function(sequence) {
  complement <- chartr("ATGC", "TACG", sequence)
  return(complement)
}

RNAm <- stri_reverse(complementary_strand(DNASequences))
print(RNAm)
```

```
## [1] "TTACAGCACCTTATGCGGGCCAAAGCATTGTAATGAATGTTTTCTGCTTCACGCCAGATCCACTAACTGTTTCGCGGTAAACTGCATGAAGC"
```

```
DNAtoRNA <- function(sequence) {
  RNA <- gsub("T", "U", sequence)
  return(RNA)
}

RNA <- DNAtoRNA(RNAm)
print(RNA)
```

```
## [1] "UUACAGCACCUUAUGCGGGCCAAAGCAUUCGUAAUGAAUGUUUCCUGCUUCACGCCAGAUCCACUAAUGUUUCCUGCGGUAAAUGCAUGAAGC"
```

```

codon <- list(ATA = "I", AUC = "I", AUU = "I", AUG = "M", ACA = "T",
  ACC = "T", ACG = "T", ACU = "T", AAC = "N", AAU = "N", AAA = "K",
  AAG = "K", AGC = "S", AGT = "S", AGA = "R", AGG = "R", CTA = "L",
  CUC = "L", CUG = "L", CUU = "L", CCA = "P", CCC = "P", CCG = "P",
  CCU = "P", CAC = "H", CAU = "H", CAA = "Q", CAG = "Q", CGA = "R",
  CGC = "R", CGG = "R", CGU = "R", GUA = "V", GUC = "V", GUG = "V",
  GUU = "V", GCA = "A", GCC = "A", GCG = "A", GCU = "A", GAC = "D",
  GAU = "D", GAA = "E", GAG = "E", GGA = "G", GGC = "G", GGG = "G",
  GGU = "G", UCA = "S", UCC = "S", UCG = "S", UCU = "S", UUC = "F",
  UUU = "F", UUA = "L", UUG = "L", UAC = "Y", UAU = "Y", UAA = "stop",
  UAG = "stop", UGC = "C", UGU = "C", UGA = "stop", UGG = "W")

codons <- stri_sub(RNA, seq(1, nchar(RNA), by = 3), length = 3)
amino_acids <- sapply(codons, function(x) codon[[x]])
amino_acids_string <- paste(amino_acids, collapse = "")

print(amino_acids_string)

```

```
## [1] "LQHLMRAKAFVMNVFLLHQAQIHstopLFRGKLHEANRAAENULLELHCRIAERTFQFAQIHQTFAIKLTFSAGFARLTNULLPGVRGKARQstop"
```

Parte 3

¿Qué significa el formato FASTA de una secuencia? Describe sus características más relevantes que lo separan de otros formatos. El formato FASTA es un formato basado en texto utilizado para representar secuencias de nucleótidos o aminoácidos mediante códigos que emplean letras. Sus características más relevantes son: 1. Línea de definición: Comienza con un símbolo de mayor que (“>”), seguido de un identificador único de secuencia (SeqID). 2. Información de la fuente: Proporciona detalles sobre el organismo del cual se obtuvo la secuencia. 3. Título de la secuencia: Opcional, brinda una breve descripción de la secuencia.

¿Qué bancos de datos existen para obtener secuencias de ADN de genomas de organismos? NCBI, EMBL, DDBJ, ENSEMBL, KEGG y OMIM. Algunas siendo originarias de países potencia como Estados Unidos, Japon e incluso Europa.

Busca la secuencia de ADN de la Alcohol Deshidrogenasa de Escherichia coli (gen adhE) y obtengan la secuencia de nucleótidos. La secuencia de nucleótidos de la Alcohol Deshidrogenasa de Escherichia coli (gen adhE) es la siguiente:

```

library(stringr)
library(sequinr)

#sequence.fasta es un archivo proveniente de la base de datos ncbi
fasta_file <- "sequence.fasta"
fasta_lines <- readLines(fasta_file)

# Remove the header line
sequence <- paste(fasta_lines[-1], collapse = "")

# Optionally remove non-nucleotide characters
sequence <- gsub("[^ACTGactg]", "", sequence)

```

```
# Now 'sequence' contains your cleaned DNA sequence
print(sequence)
```

```
## [1] "ATGGCTGTTACTAATGTCGCTGAACTTAACGCACTCGTAGAGCGTGTAAGAAAAAGCCCAGCGTGAATATGCCAGTTTCACTCAAGAGCAAGTAGA"
```

```
nucleotide_sequence <- function(sequence) {
  countA <- 0
  countT <- 0
  countG <- 0
  countC <- 0

  for (i in sequence) {
    if (toupper(i) == "A") {
      countA <- countA + 1
    } else if (toupper(i) == "T") {
      countT <- countT + 1
    } else if (toupper(i) == "G") {
      countG <- countG + 1
    } else if (toupper(i) == "C") {
      countC <- countC + 1
    }
  }

  cat("A:", countA, "T:", countT, "G:", countG, "C:", countC, "\n")
}

# Split the sequence into smaller chunks
chunk_size <- 10
chunks <- strsplit(sequence, "")
for (chunk in chunks) {
  nucleotide_sequence(chunk)
}
```

```
## A: 680 T: 641 G: 665 C: 690
```

Referencias:

- Mira, P. J. (2021, 7 febrero). Desnaturalización: Definición, tipos funciones y ejemplos - El Gen Curioso. El Gen Curioso. <https://www.elgencurioso.com/diccionario/desnaturalizacion/>
- cómo se desnatura el adn. (s. f.). <https://aleph.org.mx/como-se-desnaturaliza-el-adn>
- Tm (temperatura de fusión). (s. f.). <https://www.genome.gov/es/genetics-glossary/Tm-Temperatura-de-fusion>
- Revisión y reparación del ADN (artículo) | Khan Academy. (s. f.). Khan Academy. <https://es.khanacademy.org/science/biology/dna-as-the-genetic-material/dna-replication/a/dna-proofreading-and-repair>
- ¿Qué es el ADN superenrollado? (s. f.). <https://www.genome.gov/es/genetics-glossary/ADN-superenrollado>
- Lololo. (2024, 10 febrero). Bacterias gram positivas: el impacto del alto contenido de GC - Coinref.es. Coinref.es. <https://coinref.es/bacterias-gram-positivas-con-alto-contenido-de-gc/>
- Escherichia coli str. K-12 substr. MG1655, complete genome - Nucleotide - NCBI. (n.d.). <https://www.ncbi.nlm.nih.gov/nuccore>
- R: read FASTA formatted files. (n.d.). <https://search.r-project.org/CRAN/refmans/seqinr/html/read.fasta.html>