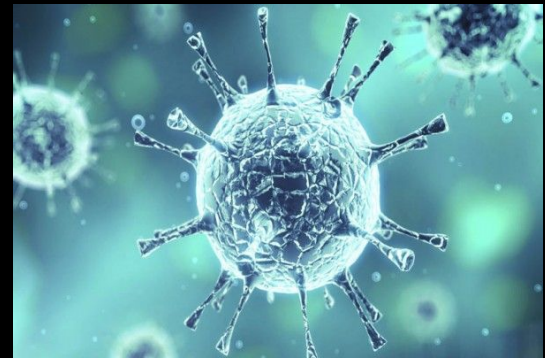


Generación sintética de ADN y virus

—
Guillermo Perez
Sergio Herrero



Generación sintética de ADN

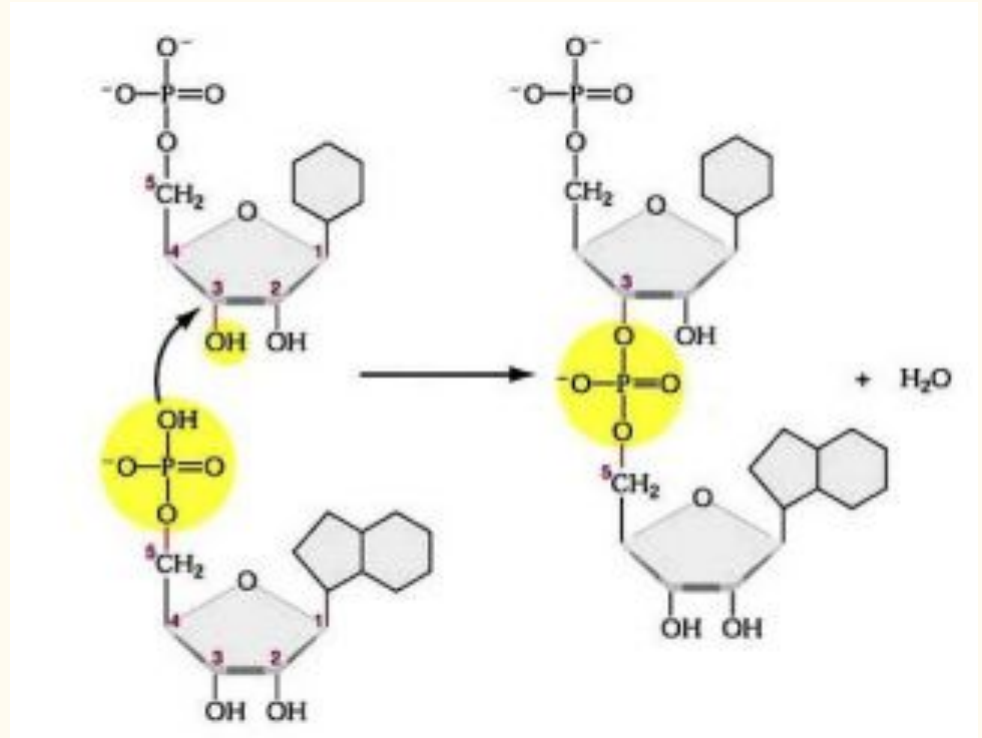
—

Definición

- Creación de genes dadas varias moléculas simples.
- Campo reciente de la biología sintética.
- Dos fases principales: síntesis de oligonucleótidos y ensamblaje de ADN.

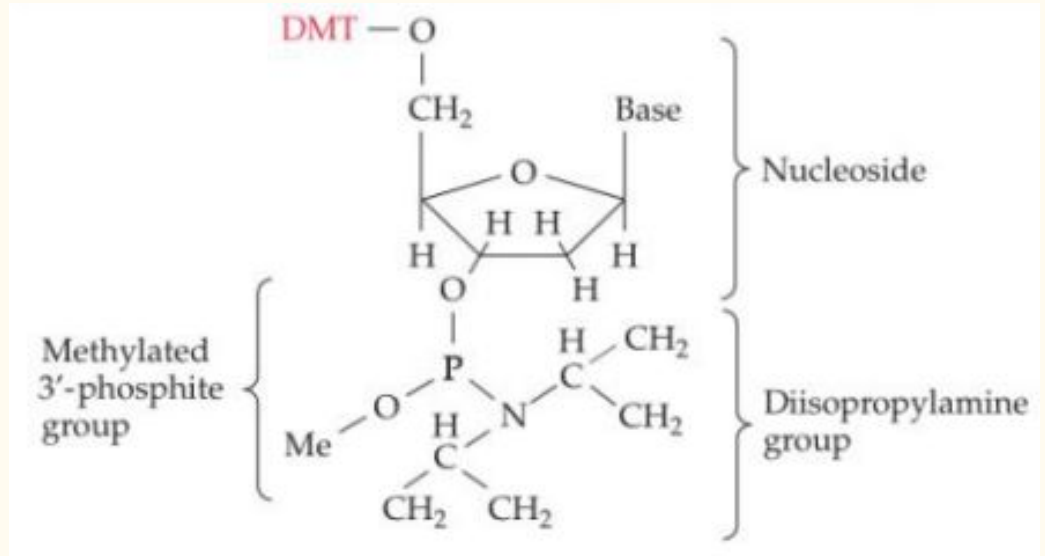
Síntesis de oligonucleótidos (I)

- Oligonucleótido: secuencia corta de ADN o ARN.
- En la naturaleza la síntesis de ADN es de 5' a 3'



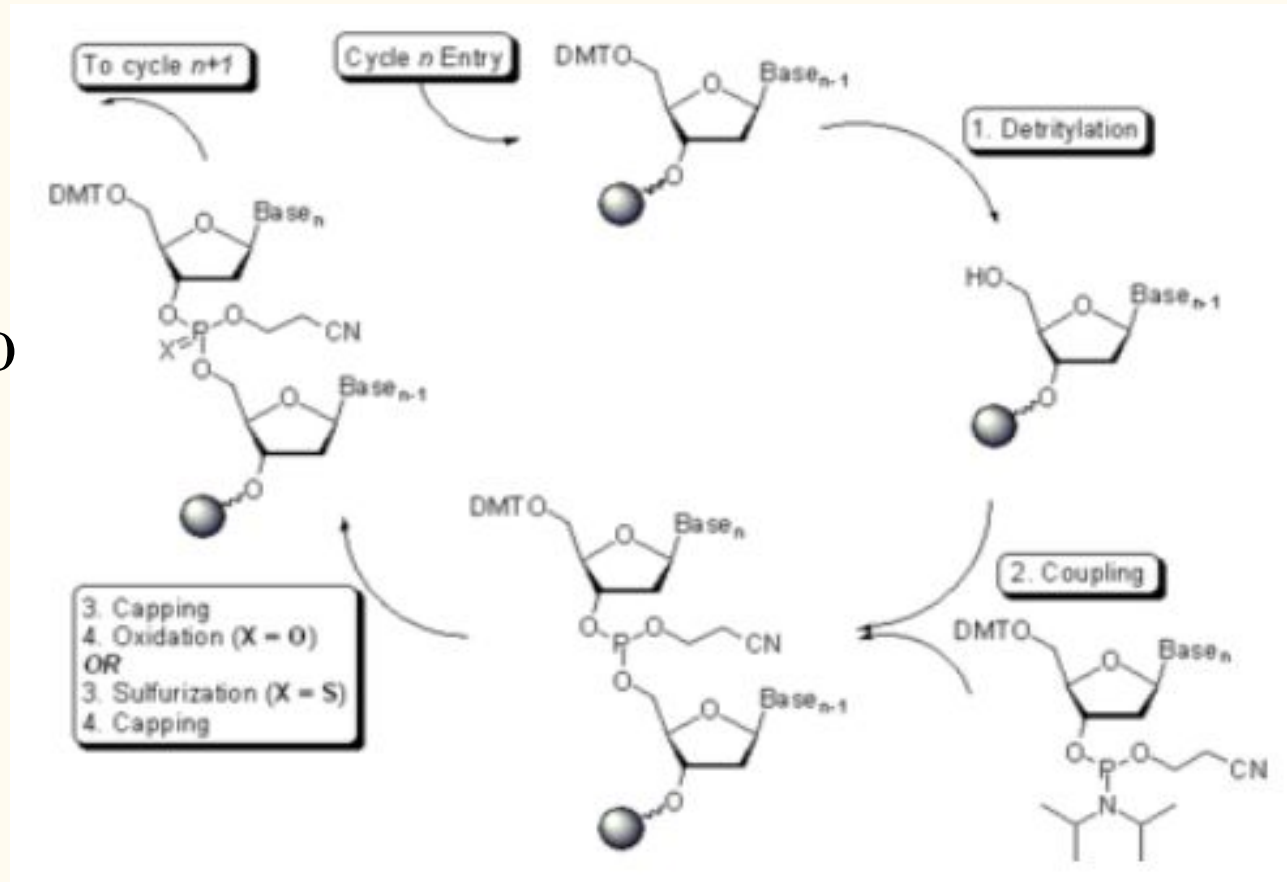
Síntesis de oligonucleótidos (II)

- Síntesis implementada usando el método de la fosforamidita



Síntesis de oligonucleótidos (III)

- Detritilación (desbloqueo)
- Acoplamiento
- Oxidación



Ensamblaje de ADN (I)

Unión de los diferentes oligonucleótidos

- Basado en enzimas: Golden Gate, BioBricks, MoClo...
- Basado en homología de secuencia (in vitro)
- Basado en homología de secuencia (in vivo)

Usos del ADN sintético

—

- Medio de almacenamiento de información.
- Optimización del genoma. Proyecto “Synthetic Yeast”.
- Síntesis de genomas virales para la creación de vacunas.

Herramienta bioinformática: EuGene

—

EuGene (I)

Software de rediseño genético sintético.

The screenshot displays the Eugene software interface, titled "Eugene - Gene Redesign and Analysis for Heterologous Expression". The interface is divided into several panels:

- Gene Optimization Panel:** Located on the left, it features a "Host Species" dropdown set to "Aquil...". Below this are several optimization options with checkboxes: GC Content, Codon Correlation Effect, Site Removal, Codon Context, Repeats Removal, Codon Usage (checked), Hidden Stop Codons (checked), Unmodified tRNAs, and RNA Secondary Structure. Under "Codon Usage", there are radio buttons for "Maximize codon usage", "Minimize codon usage", and "Harmonize codon usage", with a "Use:" section containing "RSCU" (selected) and "CAI". A "Keep Rare Codons" checkbox is also present. At the bottom of this panel are buttons for "Gene Diagnosis", "Save Redesign Protocol", "Load Redesign Protocol", and a "Redesign Gene" button with a "Fast" checkbox.
- Central Panel:** Displays a sequence editor for "tufA1" from "Aquifex aeolicus VF5". The sequence is shown in a table format with codons highlighted in yellow and their corresponding amino acids below them: AUG (M), GCA (A), AAG (K), GAG (E), AAA (K), UUU (F), and G (I).
- Protein Viewer Panel:** Located on the top right, it shows "No protein to display" and a "Jmol" button.
- Gene Information Panel:** Located on the bottom right, it provides detailed information about the gene. It includes sections for "[Species and Gene]", "[General information]", "[Colour details]", and "Obtain Gene". The "General information" section lists: Genome name (Aquifex aeolicus VF5), Gene name (tufA1), Number of codons (486), GC content (49.59%), Average RSCU (1.31), Codon Pair Bias (-0.007), Effective Number of Codons (42.655), and CAI (0.721). The "Colour details" section shows a "Codon Usage color scale" from "zero RSCU" (red) to "max RSCU" (green). The "Obtain Gene" section has a text input field for "Insert Gene Transcript:" and a "Search" button.
- Progress Panel:** Located at the bottom left, it shows a status message: "Fetching from EBI Error connecting!".

EuGene (II)

File Edit View Gene Pool Help

Gene Optimization

Host Species: Aquifex

- ☒ GC Content
- ☒ Codon Correlation Effect
- ☒ Site Renoval
- ☒ Codon Context
- ☒ Repeats Renoval
- ☒ Codon Usage

☒ Maximize codon usage
☐ Minimize codon usage
☐ Harmonize codon usage

Use: ☒ RSCU ☐ CAI

☐ Keep Rare Codons

☒ Hidden Stop Codons

☒ Unmodified tRNAs

☒ RNA Secondary Structure

Gene Diagnostics | Redesign Gene

Save Redesign Protocol | Load Redesign Protocol

Fast

Progress Panel

Redesigning Done

Fetching from EMBL Error connecting!

Protein Viewer

No protein to display

jmol

Gene Information Panel

[Species and Gene]

Genome name: Aquifex aeolicus VF5
Gene name: tufA1

[General information]

Number of codons: 406
GC content: 45.73%
Average RSCU: 1.433
Codon Pair Bias: -0.134
Effective Number of Codons: 28.289
CAI: 0.82

[Optimizations Report]

Codon Usage: 87.2% (11K)
Hidden Stop Codons: 97.1% (59.6K)

[Colour details]

Codon Usage color scale

Zero RSCU | max RSCU

Obtain Gene

Insert Gene Transcript:

Search

Project 1

[Aquifex aeolicus VF5] tufA1 (Optimized by: Multiple Redesigns)

tufA1

1 5 10 15 20

AUG GCU AAG GAA AAC UUU GAA AGC ACU AAG GAA CAC GUU AAG GUA GGA ACG UUA GGA G

M A K E K F E R T K E H V N V G T I G I

[Aquifex aeolicus VF5] elongation factor EF-Tu (tufA1)

tufA1

1 5 10 15 20

AUG GCU AAG GAA AAC UUU GAA AGC ACU AAG GAA CAC GUU AAG GUA GGA ACG UUA GGA G

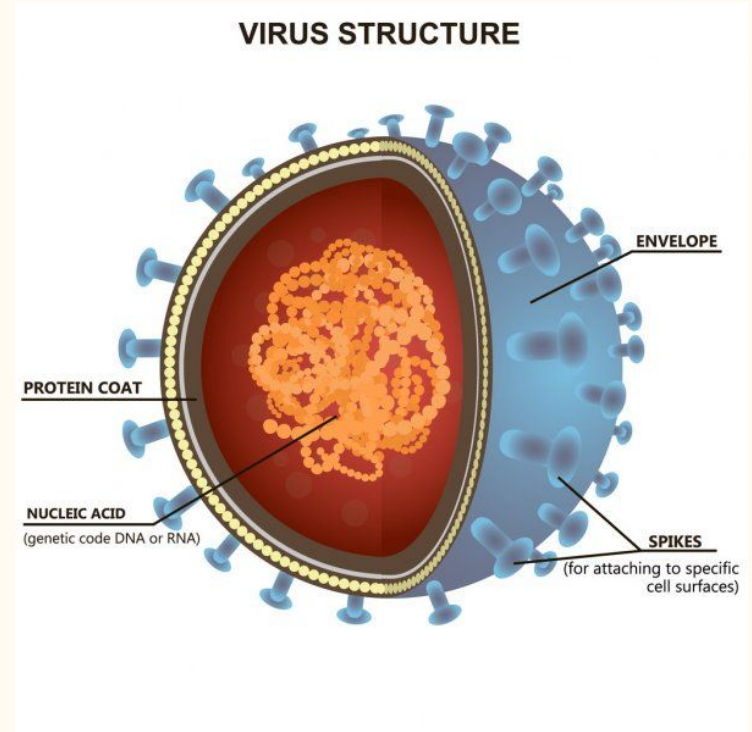
M A K E K F E R T K E H V N V G T I G I

Generación sintética de virus

—

Generación sintética de virus

- Virología sintética
- Constitución de los virus
 - ADN o ARN
 - Cápside
 - Envoltura de lípidos y glicoproteínas (algunos)



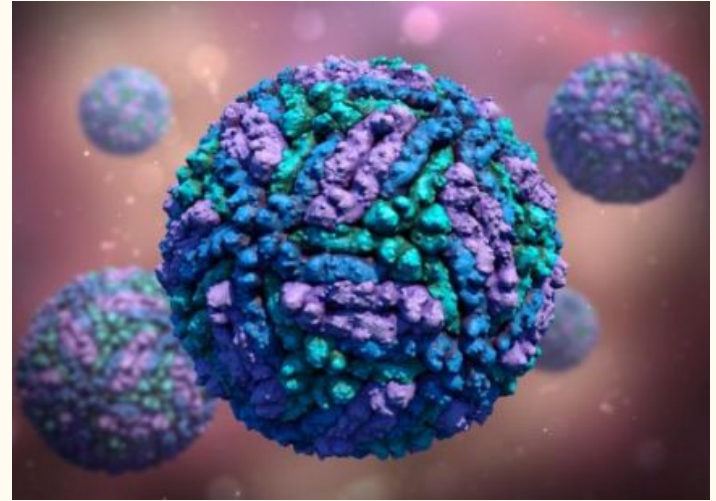
Casos prácticos



- Poliovirus
- Gripe española
- Coronavirus
- **Virus del Dengue**

Virus del dengue

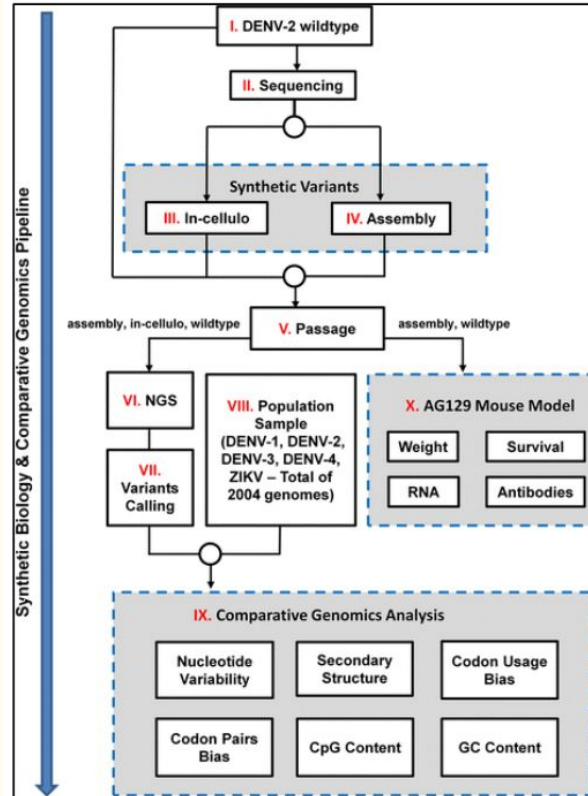
- Flavivirus ARN
- Transmitido por mosquitos
- 4 serotipos diferentes
- Genoma de 11.000 pb



Sintetización del virus del Dengue

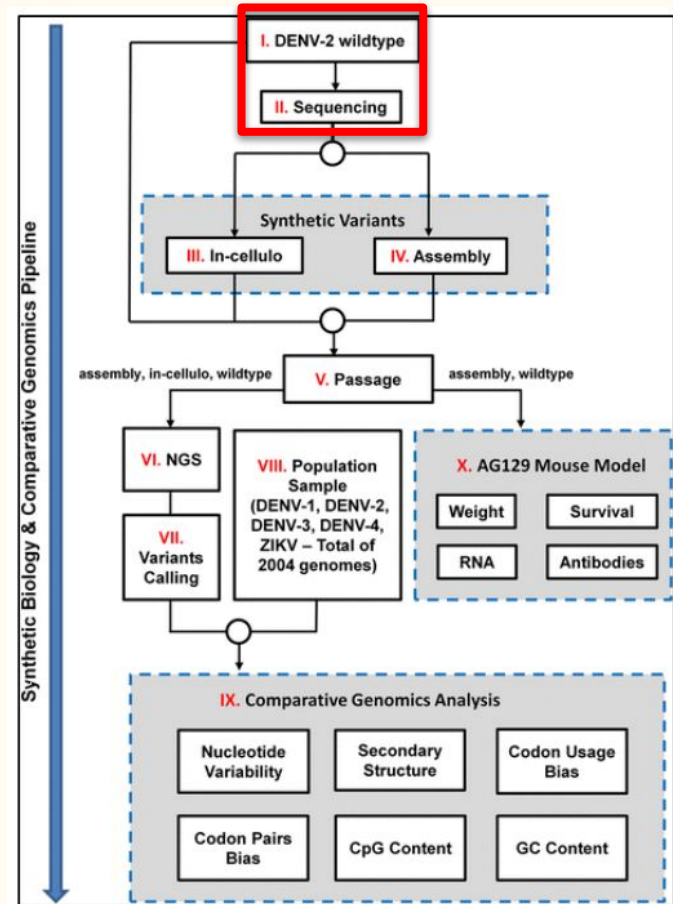
- DENV-2 (Virus del dengue, serotipo 2)
- 2 variantes sintéticas
- Uso de pipeline multidisciplinar

Pipeline



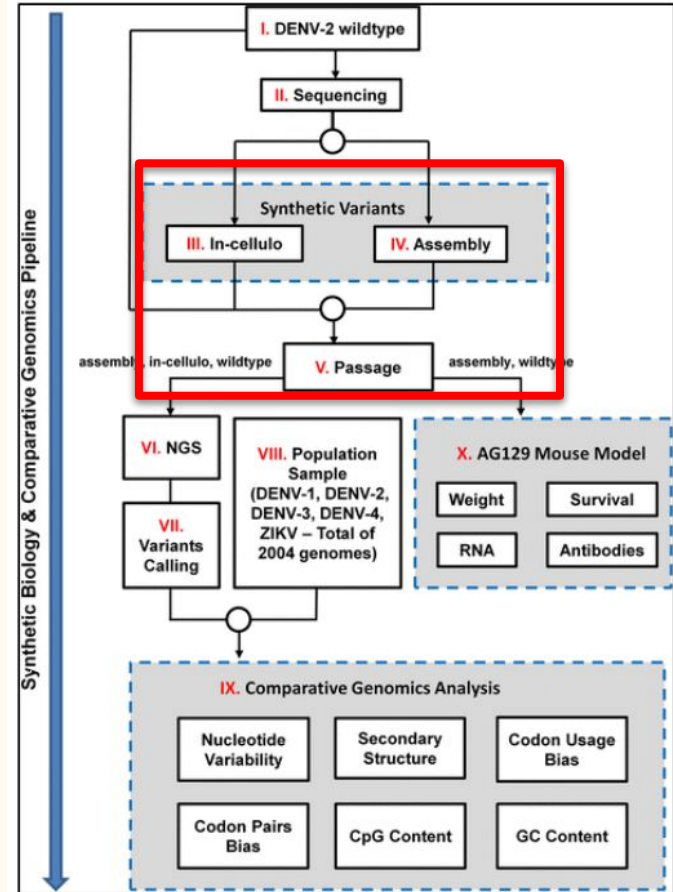
Secuenciación

- Obtener secuencia de origen animal
- Secuenciar y duplicar con PCR



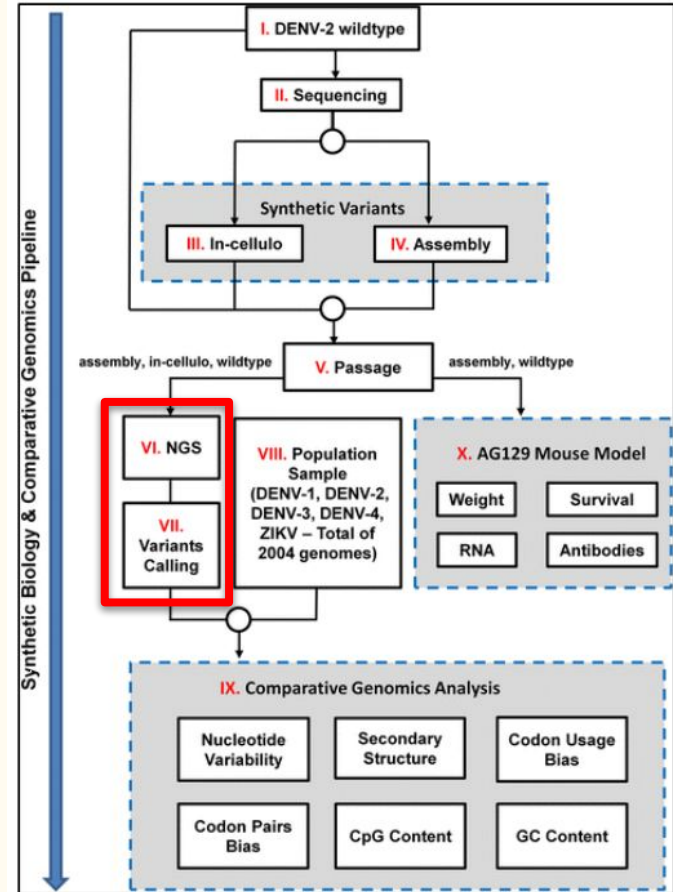
Variantes sintéticas

- 2 diferentes:
 - amplicones infecciosos sub-genómicos
 - Ensamblado ADN por PCR



Next Sequencing generation y variantes(I)

- Nuevas secuencias alineadas con genoma de referencia
- Selección de los mejores alineamientos
- Comparaciones entre ellas

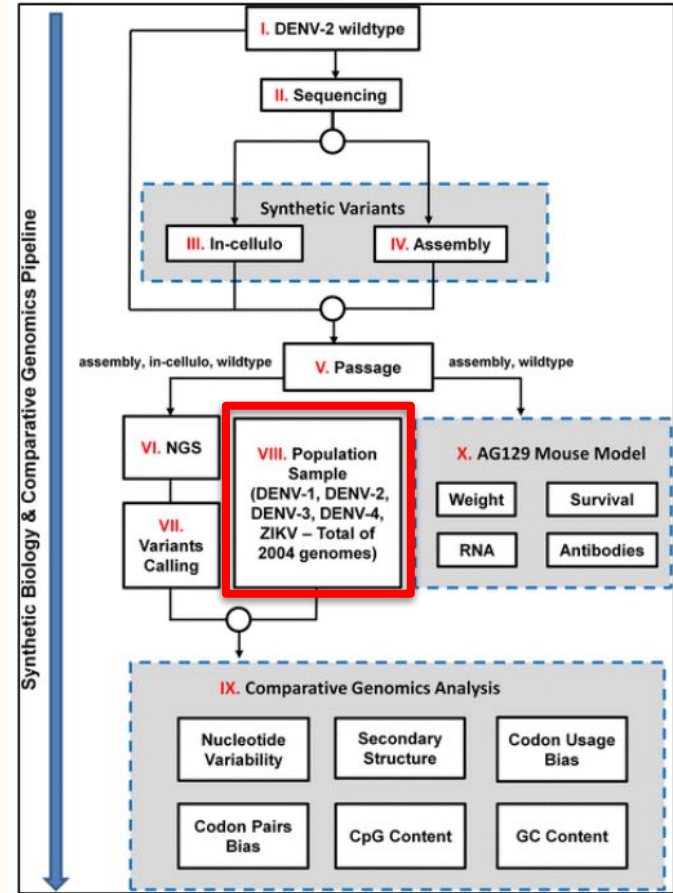


Next Sequencing generator y variantes(II)

- Alineamientos entre sintéticas con algoritmo **BWA-MEM**
 - 10x más rápido que otros algoritmos
- Selección de mejores alineamientos con **Picard MarkDuplicates**
 - Ordena según criterio, marca duplicados...
- Comparaciones entre ellas con **FreeBayes**
 - **SNP** e **Indeles**

Comparaciones(I)

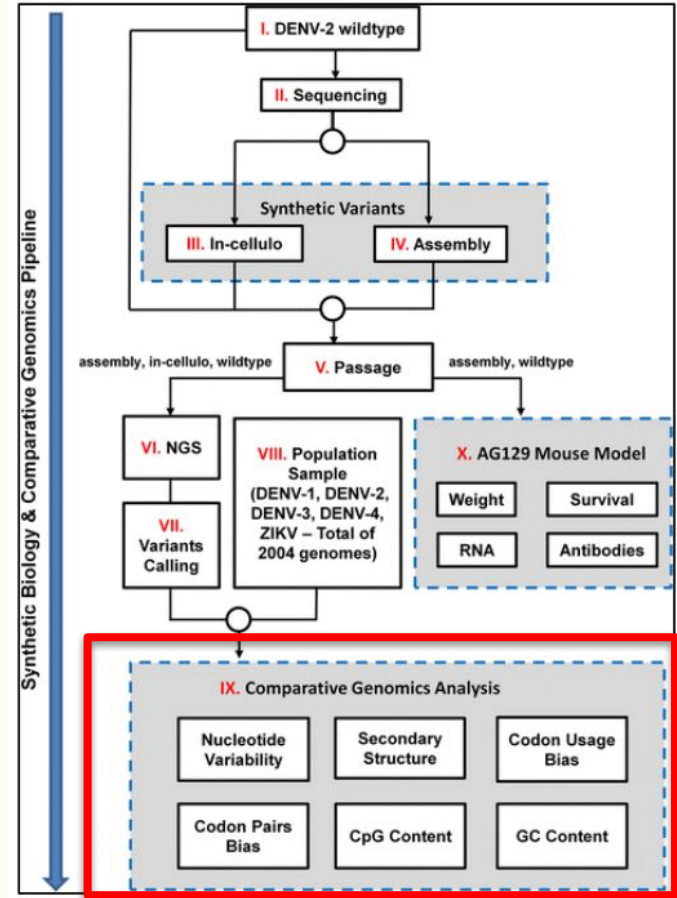
- Comparados con 2004 muestras de virus
- Muestras obtenidas de **GenBank**
- Alineaciones con **Clustal Omega**



Comparaciones(II)

- Comparación variabilidad de cada posición
- Usando entropía normalizada de Shannon
- Más variabilidad en sintético

$$V_i = \frac{\sum_{j=1}^n p_j \log_2 (p_j)}{\log_2 n}$$



Positions of SNVs in genome

positions	variability	SNVs	VAF	
50005		GT(C) → GT(V)	0.06	s
9567		GTT(V) → GTA(V)	0.50	*
9468		CAG(Q) → CAA(Q)	0.08	s
8925		TAT(Y) → TAC(Y)	0.09	s
8913		GAA(E) → GAG(E)	0.09	s
8828		AAG(K) → AGG(R)	0.07	s
8214		ACA(T) → ACG(T)	0.06	s
7435		TG(L) → CTG(L)	0.06	*
7152		GCA(A) → GCG(A)	0.48	W
6981		TTA(L) → TTG(L)	0.14	W
6665		TAC(Y) → TGC(C)	0.10	W
6588		TTC(F) → TTT(F)	0.10	W
6519		CTG(L) → CTA(L)	0.13	W
5846		ACC(T) → ATC(I)	0.08	Z
5556		AGT(S) → AGC(s)	0.13	W
5139		GCC(A) → GCA(A)	0.09	*
5061		GAC(D) → GAT(D)	1.00	*
4806		GTC(V) → GTT(V)	0.06	*
3708		CTC(L) → CTA(L)	1.00	W
3465		TC(C)S) → TCT(S)	0.14	*
3453		AAT(N) → AAC(N)	0.12	*
3290		GAC(D) → GGC(G)	0.22	Z
3037		CTC(L) → TTC(F)	0.07	Z
2989		GAC(D) → AAC(N)	0.14	Z
2982		GCC(A) → GCT(A)	0.06	s
2639		CCA(P) → CTA(L)	0.44	*
1935		TGC(C) → TGT(C)	0.85	W
1932		CCA(P) → CCG(P)	0.16	W
1683		GAT(D) → GAC(D)	0.06	s
1565		AGG(R) → AAG(K)	0.10	Z
1211		CTA(L) → CTT(L)	0.05	W
953		ATA(I) → ACA(T)	0.14	Z

b

Conclusiones

- ADN Sintético en desarrollo
- Retos nuevos como sintetizar el genoma humano
- Aspectos diferentes:
 - Bueno: Creación de vacunas
 - Malo: Virus a la carta

Referencias

- Goz, E., Tsalenchuck, Y., Benaroya, R.O. et al. Generation and comparative genomics of synthetic dengue viruses. BMC Bioinformatics 19, 140 (2018)
- Chao, Ran et al. “Recent advances in DNA assembly technologies.” FEMS yeast research vol. 15,1 (2015)
- Ma, Siying et al. “DNA synthesis, assembly and applications in synthetic biology.” Current opinion in chemical biology vol. 16,3-4 (2012)

Muchas gracias

—