

Bio Project

Stijn Rosaer¹ and Olivier Van Houtte²

1- Universiteit Antwerpen - Computer Science
Prinsstraat 13 2000 Antwerpen - Belgium

Abstract. Zoals we weten, is er interesse in het storen van data als DNA. Er bestaan al veel methoden om dit DNA op te bouwen en terug uit te lezen, maar de encodings wijze staat nog niet vast. In dit verslag achterhalen we ofdat er misschien tijd en/of plaats efficiëntere encodings systemen zijn dan gewoon bit voor bit op te slaan en of deze in de praktijk toepasbaar zijn.

1 Inleiding

Als we aan DNA denken, worden we al snel herinnerd aan de 4 basen. Adenine (A), guanine (G), cytosine (C) en thymine (T). Met deze basen kunnen we de simpelste vorm van encoding opstellen, namelijk 2 bits per base. We kiezen dan voor elke base een unieke combinatie aan bits.

00 = A, 01 = G, 10 = C, 11 = T

We kunnen dan elke DNA string (met een even aantal bits) omzetten naar DNA.

0111001110 -> GTATC

Het kan echter efficiënter.

2 Encoding

In plaats van simpelweg 2 bit aan elke base toe te wijzen, stellen we nu eerst alle mogelijke sets op bestaande uit unieke combinaties van de 4 basen. Er zijn er zo 15.

A			
G			
C			
T			
A	G		
A	C		
A	T		
G	C		
G	T		
C	T		
A	G	C	
A	G	T	
A	C	T	
G	C	T	
A	G	C	T

We kunne aan hier de getallen 0 t.e.m. 14 aan toewijzen. we kunnen deze zelfde getallen ook representeren met 4 bits. Hierbij komen op een probleem. Met 4 bits kan je de 16 getallen representeren, namelijk 0 t.e.m. 15. 1 Getal meer dan de 15 getallen die we met de base-sets bekomen. We zullen dus 1 bit moeten laten vallen. Dat brengt ons op 3 bits die de getallen 0 t.e.m. 7 voorstellen. We kiezen dan 8 sets uit onze 15, en wijzen aan elk van deze sets 1 getal tussen 0 en 7 toe.

0		A	
1		G	
2		C	
3		T	
4		A	G
5		A	C
6		A	T
7		G	C

3 Correctheid

4 In de praktijk

5 Conclusie

And that's the way the cookie crumbles.

References

- [1] Choi, Y., Ryu, T., Lee, A.C. et al. High information capacity DNA-based data storage with augmented encoding characters using degenerate bases. Sci Rep 9, 6582 (2019). <https://doi.org/10.1038/s41598-019-43105-w>