Bio-informatica week 1

HIV- Human Immunodeficiency Virus

AIDS = Acquired Immune Defiency Syndrome

Virussen bestaan uit:

Genoom; DNA of RNA; enkelstrengs of dubbelstrengs.

Capside; capsomeren (eiwitmantel)

Dierlijke virussen hebben een membraan.

Virussen hebben geen stofwisseling en kunnen zichzelf niet voortplanten!

Genoomanalyse:

Grootte van het genoom

Coding/ noncoding regionen van het genoom

Expressie profielen van het genoom

GC% van regionen in het genoom of het gehele genoom

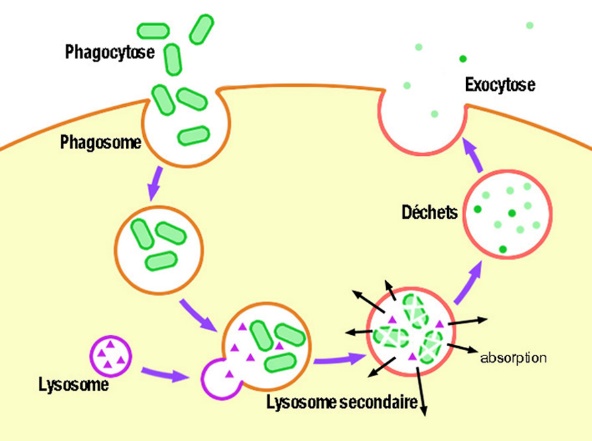
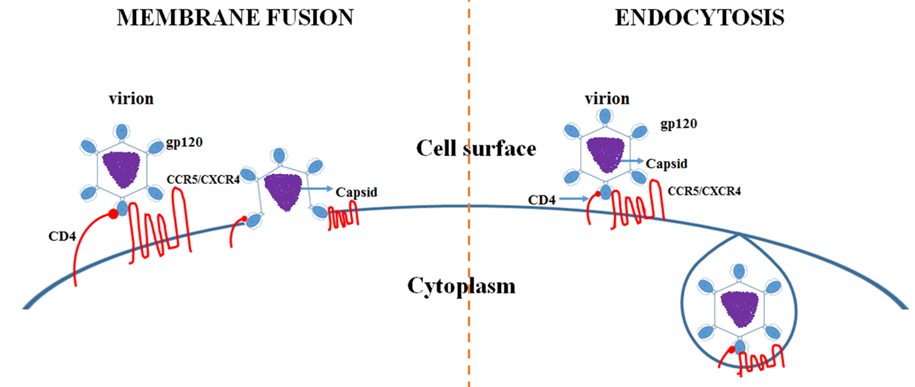
Overeenkomsten met andere organismen; Evolutie

Centrale dogma:

**DNA** -transcriptie- **mRNA** -translatie- **polypeptideketen** -post translationele vouwingen- **functioneel eiwit**

Indeling dierlijkenvirussen:

1. ds DNA
2. ss DNA
3. ds RNA
4. ss RNA
   1. ss RNA 🡪 serves as mRNA
   2. ss RNA 🡪 template for mRNA synthesis
   3. ss RNA 🡪 template for DNA synthesis

[](https://www.google.nl/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwinp_KA7ozaAhVJjqQKHXSDAE8QjRx6BAgAEAU&url=https://brunoanselme.wordpress.com/2011/11/30/endocytose-et-exocytose-laraignee-le-camembert-et-le-ruminant/&psig=AOvVaw3_4MbpIp-XbO2cFieaY_kD&ust=1522252334995548)[](https://www.google.nl/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwj4rY6o6YzaAhVMy6QKHcWDB9UQjRx6BAgAEAU&url=https://www.intechopen.com/books/advances-in-molecular-retrovirology/role-of-host-proteins-in-hiv-1-early-replication&psig=AOvVaw2mJ2UOaQq5xcmzRHt1NsW9&ust=1522250787150551)

Het HIV genoom bevat:

2 identieke RNA strands

9 genen 🡪 9 primaire eiwitten 🡪 18 uiteindelijke eiwitproducten

Membraan

Klein virus, 9181 nt voor HIV1 en 10359 nt voor HIV2

Het verschilt per reading frame welke eiwitten er geproduceerd worden.

Infecteert T-helpercellen

HIV eiwitten:

Structural eiwitten: Regulatory eiwitten: Accessory eiwitten:

POL (NEF) (NEF)

ENV TAT VIF

GAG REV VPR

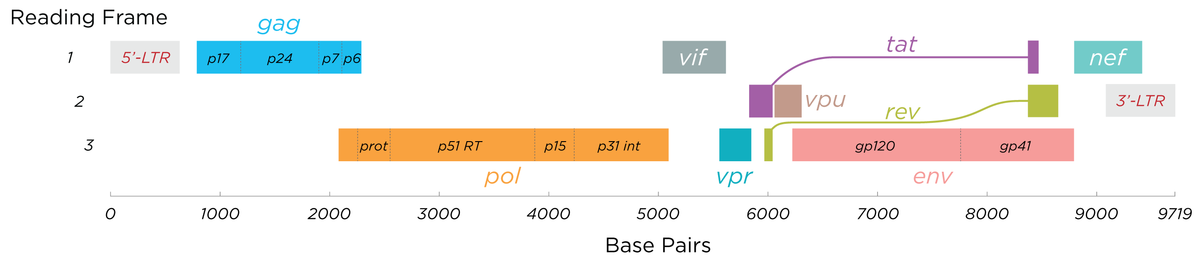
VPU

HIV enzyme:

Reverse Transcriptase: vertaalt (dubbelstrengs of enkelstrengs) mRNA naar DNA

Integrase: integreert virus genoom in het DNA van de hostcel

Protease: knipt eiwitten in kleinere stukjes

[](https://www.google.nl/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwi0jKPo-YzaAhWQy6QKHaxLCa4QjRx6BAgAEAU&url=http://www.wikiwand.com/en/Structure_and_genome_of_HIV&psig=AOvVaw1F75RQjiRwACLQdpwXrC6Q&ust=1522255494763721)

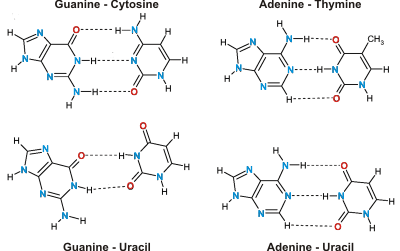
Bio-informatica week 2

Metagenomics: studie van collectieve genomen in monsters uit de omgeving 🡪 verschillende monsters uit verschillende ecosystemen vergelijken.

Karyotype:

GC rijke regionen = donker en stabiel

AT rijke regionen = uiteindes van chromosomen

[](https://www.google.nl/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwjA7qLA_IzaAhXI_aQKHeIGCMgQjRx6BAgAEAU&url=http://www1.lsbu.ac.uk/water/nucleic_acid_hydration.html&psig=AOvVaw2VbguuGxPcCs3Ymr592EkC&ust=1522256165543185)GC% = \*100%

GC ratio =

G&A 🡪 purines

C&T 🡪 pyrimines

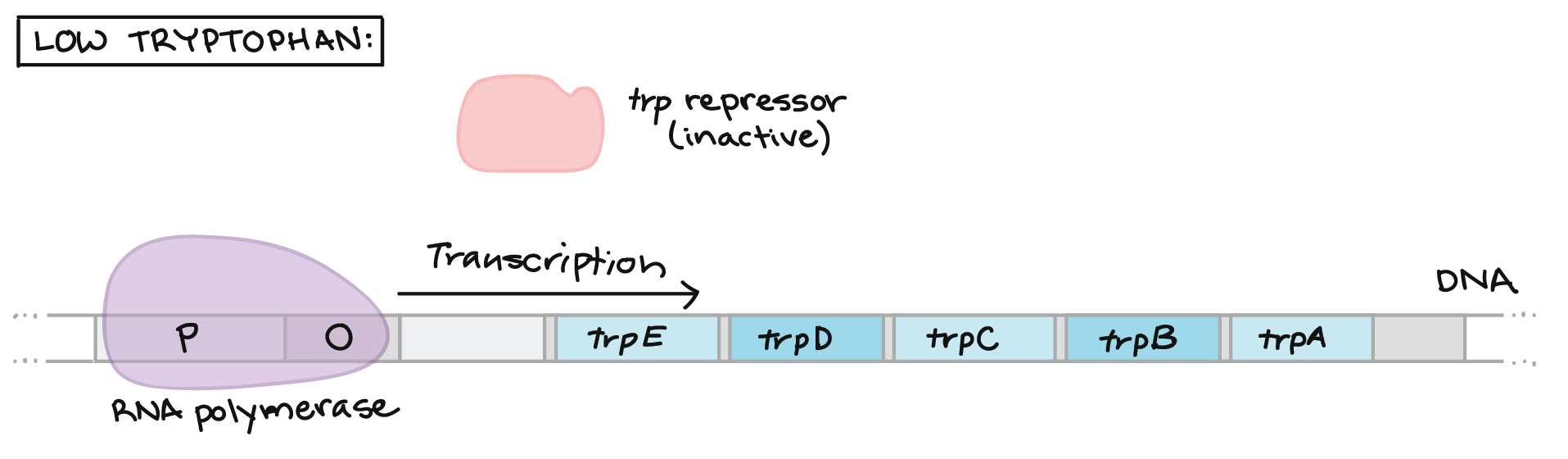
C&G 🡪 3 waterstofbruggen

T&A 🡪 2 waterstofbruggen

Hetrochromatine 🡪 compact DNA, komt het meeste voor in inactieve genen

Euchromatine 🡪 losser DNA, komt het meeste voor in actieve genen

|  |  |
| --- | --- |
| Prokaryoten | Eukaryoten |
| Minder genen | Meer genen |
| Hoge gen dichtheid | Lage gen dichtheid |
| Aantal genen = grootte genoom | Aantal genen != grootte genoom |
| GC% = 25-75% | GC% = 40-45% |
|  |  |
| Kleiner genoom: | **Groter genoom:** |
| Geen repeads | Alternative splicing |
| Gebruik van operons (plaatje) | Post translationele modifications |
| Geen intronen | Regulaties van RNA moleculen complexe promotors |
| Kleine regulerende regio’s |  |

[](https://www.google.nl/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwj39PWOgI3aAhWK-qQKHeY3B08QjRx6BAgAEAU&url=https://www.khanacademy.org/science/biology/gene-regulation/gene-regulation-in-bacteria/v/operons-and-gene-regulation-in-bacteria&psig=AOvVaw0fh05oahaKpBzwVB4qSUnP&ust=1522257180252288)

-Een gen is een nucleotide sequentie dat codeert voor (meerdere) functionele producten, RNA of eiwitten (kan ook beide).

-Non coding-genes kunnen coderen voor tRNA of rRNA

**-**Isochores zijn stukjes DNA (300 kb regionen) met een redelijk stabiel GC%

L1&L2 🡪 ‘empty space’

GC% <40%

Gen arm

85% van weefsel specifieke genen

Genen die alleen tot uiting komen in bepaalde weefsels

H1,H2&H3 🡪 (47%, 52%, >52%)

Gen rijk

Kortere intronen en genen

Weinig repeads

80% van huishoudgenen

Genen die elke cel nodig heeft

Bio-informatica week 3

Hoe kunnen we genen vinden in prokaryoten:

~ 100-500 aminozuren lang

Correcte codon bias

**ORF**, OpenReadingFrame 🡪 van start tot stopcodon

**RBS**, RibosomeBindingSite gevolgd door ATG

Stop gevolgd door terminatie signaal loop+TTTT

Hoe kunnen we genen vinden in eukaryoten:

Correcte codon bias

ORF, OpenReadinFrame van minimaal 90 bp

CpG islands 🡪 hoog aantal GC eilanden bij transcriptie start site

RBS, RibosomeBindingSite gevolgd door ATG

5’ CAP

Poly-A-tail signal

Splice sites

Codon Bias 🡪 voorkeur van een bepaald organisme voor bepaalde codons. Dit is afhankelijk van het aantal beschikbare tRNA’s

Codon Usage 🡪 gebruik van codons in een gen, dit kan afwijken van de codon bias.

Verklaring voor bias&usage:

Codon Bias is gerelateerd aan gewenste evolutionaire verandering (GC%) of keuze van tRNA’s

Codon Usage is gerelateerd aan translatie efficiëntie of gewenste evolutionaire verandering

Codons met CpG worden vaak vermeden

Low-usage codon

Muteert gemakkelijk in TG

C in CpG is vaak gemethyleerd

Transcriptie inhibitie

CpG is vaak regulator voor huishoudgenen

Er is een sterke correlatie tussen:

GC content op de 3e positie van codons in een gen

De GC content van een regio

Evolutie 🡪 genetische variatie als gevolg van mutaties

Mutaties:

Somatisch 🡪 niet overerfbare mutaties

Gametisch 🡪 overerfbare mutaties

Oorzaken:

Mutagenen 🡪 fouten in DNA herstel

Replicatie 🡪 fouten in DNA replicatie

Noncoding regionen muteren 2x zo snel

Missence mutatie 🡪 ander aminozuur

Nonsense mutatie 🡪 vroeg stopcodon

Silent mutatie 🡪 zelfde aminozuur.

Natuurlijke selectie:

omgevingsdruk is van invloed op welke genen efficiënt zijn en welke genen niet efficiënt zijn voor een bepaald organisme.

Proces waarbij een individu met gunstigere eigenschappen meer overlevingskans heeft en een reproductieoordeel heeft.

Wat heb je voor evolutie nodig?

Template DNA

Kopieer mechanisme

Replicatie; meiose; bevruchting

Variatie

Crossing over; gene conversie; mutaties

Selectie

Evolutie in HIV:

RNA 🡪 reverse transcriptase 🡪 DNA

Geen proofreading

1 op de 1000 bp mutatie

Veel nakomelingen

Immuunsysteem en medicijnen

Codon Bias die overeenkomt met de Codon Usage van de hostcel.

Bio-informatica week 4

4 structuurniveau’s:

Primarie structuur

Aminozuur volgorde

Peptidebindingen 🡪 condensatiereactie

N terminus 🡪 **aminogroep**

C terminus 🡪 **carboxylgroep**

Secundaire structuur

Alfa helixen, beta sheets en beta strands

Waterstofbruggen tussen N&C terminus

Beta sheets 🡪 2 of meer strands langs elkaar gelegen

Beta strands 🡪 maximaal gestrekt stukje polypeptide

Betaturn 🡪 bochtje in een strand, vaak glycine en proline.

1 turn = 4 aminozuren

Tertiare structuur

Vouwingen door de restgroepen van de aminozuren

Interacties tussen restgroepen

Waterstofbruggen

Zwavelbruggen

Ion-bindingen

Hydrofobe interactie

Polair naar kern gericht

Vanderwaalskrachten

Quaternaire structuur

Het samenvoegen van meerdere polypeptiden

Zelfde interacties als de tertiare structuur.

Aminozuren

Koolstof C

Stikstof N

Aminogroep NH2

Zuurgroep HO--C=O

Restgroep

Polycondensatiereactie🡪 afsplitsing van H2O

Hydrofoob:

Geen ring: ala, leu, ile, val **alifatisch**

Met ring: phe, tyr, trp, (his) **anomatisch**

Hydrofiel:

Polair: asn, gln

Alcoholic: ser, thr, (tyr)

Charged: arg, lys, asp, glu, (his)

Allebei niet:

Sulfur: met, cys **zwavelbruggen**

Special: gly **geen restgroep**

Pro **restgroep zit aan stikstof vast**

Non coding regio’s

Pseudogenes 🡪 genen die hun functie zijn verloren als gevolg van mutaties

Repetitief DNA 🡪 meerdere kopieën van een stuk sequentie

75% bestaat uit transposable elements

Replicatie en recombinatie

Transposable elements

Stukken DNA die zich kunnen verplaatsen binnen het genoom.

Alu elements 🡪 30 bp

Familie met vergelijkbare sequenties

Line-1 elements 🡪 6500 bp

Zeer lage transpositie snelheid

Simple sequence DNA zijn korte kopieën van stukjes DNA

Eukaryoten kennen 2 soorten transposable elements:

Transposons 🡪 copied; insertion

Retrotransposons 🡪 reverse transcriptase; insertion

Transposons en evolutie:

1. kopieën van transposons kunnen leiden tot crossing over
2. inserties van TE’s in eiwitten

* geen eiwit productie

1. inserties in regulatiore sequenties

* productie wordt verhoogt of verlaagt

1. TE’s kunnen genen verplaatsen
2. TE’s kunnen nieuwe tRNA transcriptie varianten maken.

Multigene families zijn collecties van 2 of meerdere identieke of vergelijkbare genen (globine)

Genoom evolutie:

Duplicatie van totale chromosomensets

Polyploïde 🡪 haploïde, triploïde

Verandering in chromosoomstructuur

Duplicatie en inversie

Duplicatie van genen

Het ontstaan van nieuwe genen

Retransposons en ongelijke crossing over

Exon duplication en exon shuffeling.

Homologe sequenties:

2 sequenties hebben een gemeenschappelijke voorouder sequentie

2 sequenties zijn waarschijnlijk homoloog als:

2 eiwit sequenties (>100 aminozuren) > 25% identity hebben

2 DNA sequenties (>100 nucleotiden) > 70% identity hebben

Bij een lagere identity kan er nog steeds sprake zijn van homologie

Ortologe sequenties:

Als hetzelfde gen in meerdere organismen voorkomt

Paraloge sequenties:

Verschillende versies van een gen komen voor in één organisme.

Bio-informatica week 5

Alignments:

Sequenties zo indelen zodat vergelijkbare gedeeltes van de sequenties met vergelijkbare eigenschappen worden uitgelijnd.

2 sequenties langs elkaar leggen en naar verschillen en of overeenkomsten kijken

Eiwitten, DNA, genoom, gen

Zegt iets over verwantschappen tussen de sequenties

Dezelfde geconserveerde stukjes in meerdere organismen

Homologe, ortologe en paraloge sequenties achterhalen

Evolutionaire events identificeren

Identificatie van geconserveerde sequenties

Structuur en functie van een eiwit ontrafelen

Pairwise alignment:

Positives 🡪 overeenkomende functies van eiwitten

GAPS simuleren Indel events; inserties, deleties

Snel en simpel

Niet altijd betrouwbaar

Score berekenen aan de hand van identities en similarities van de aminozuren

Blast

Multiple alignment:

Zwaar voor de computer

Iedere sequentie wordt met iedere sequentie vergeleken

Scores kunnen worden gebruikt voor een fylogenetische boom

Betrouwbaar

cluselOmega