Bioinformatica 1

Auteurs:

Mark Sibbald Jeanine de Keyzer Jurre Hageman

Versie:

25 februari 2020

Overzicht lessen:

Les 1: Algemene computervaardigheden Les 2: Refworks/PubMed

Les 3: DTP Les 4: Excel

Les 5: Genbank/BLAST

Les 6: Benchling Les 7: Eindopdracht

Les:

Les 6: Benchling

Leerdoelen

Tijdens deze les staan de volgende leerdoelen centraal:

- DNA-moleculen analyseren met Benchling
- Een restrictie uitvoeren met Benchling
- PCR simuleren met Benchling

Inhoudsopgave

Bioinformatica 1	1
Auteurs:	1
Versie:	1
Overzicht lessen:	1
Les:	1
Leerdoelen	2
Inhoudsopgave	3
Benchling	
Inleiding	
De ziekte van Huntington	
Account voor Benchling maken	
Opdracht 1: Analyse van DNA-moleculen in Benchling	6
GFP Fusie-eiwit	6
Openen/importeren DNA/RNA-sequenties	6
Restrictie enzymen en restrictie analyse in Benchling	8
Opdracht 2: Simulatie van een PCR	12
Tot slot	17

Benchling

Inleiding

Tijdens deze les leer je werken met Benchling. Dit programma is een veel gebruikt programma binnen de moleculaire biologie. Het vereenvoudigt het werken met DNA-sequenties. Je kan kloneringen eerst simuleren op de computer voordat je deze in het lab uitvoert. Dat verkleint de kans op fouten aanzienlijk. Ook is het programma geschikt om primers en probes te ontwikkelen en een PCR te simuleren.

Als model gebruiken we de ziekte van Huntington. Allereerst gaan we een plasmide analyseren waarbij een ziekmakend allel van het Huntington-gen gekloneerd is in een plasmide. Hierbij is het gen gefuseerd aan een GFP-eiwit waardoor het fusie-eiwit groen oplicht onder een fluorescentie microscoop.

Daarna gaan we met een set PCR-primers een simulatie PCR uitvoeren om de ziekte van Huntington te detecteren.

Allereerst wat informatie met betrekking tot de ziekte van Huntington.

De ziekte van Huntington

De ziekte van Huntington is een autosomaal dominant overdraagbare aandoening. Mutaties in het ziekmakende allel van het gen zorgen voor het ontstaan van een mutant eiwit (het Huntingtin eiwit). De ziekte openbaart zich over het algemeen pas op latere leeftijd. De ziekte is een neurodegeneratieve ziekte (tast het zenuwstelsel aan). Mensen krijgen neurologische problemen zoals chorea (spier spasmen) en tal van psychiatrische problemen. Op dit moment is deze ziekte niet te genezen.

Om je een indruk van het ziektebeeld te geven, kun je het volgende filmpje bekijken:

http://www.youtube.com/watch?v=PCVO9c9q4tE&feature=fvw

Account voor Benchling maken

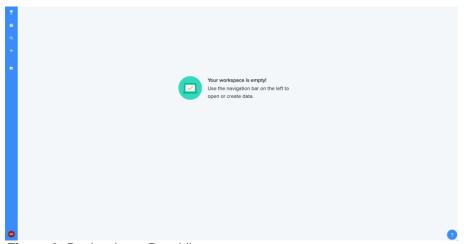
Benchling is een online programma om DNA-moleculen te analyseren en manipuleren. Het wordt vaak gebruikt om een te testen of een manipulatie (klonering, restrictie, CRISPR-Cas, etc.) de juiste resultaten oplevert, voordat het in het lab uitgevoerd wordt. Het programma is gratis te gebruiken voor studenten en het enige wat je hoeft te doen is een account aan te maken.

Ga naar <u>www.benchling.com</u> en klik op 'Sign up' om een nieuw account te maken. Volg de instructies om een nieuw account aan te maken. Gebruik je Hanze emailadres en vul bij 'Affiliation' 'Academic' in (Figuur 1), zodat je de gratis versie kunt gebruiken.



Figuur 1. Sign up voor een Benchling account.

Login in je Benchling account. Het scherm is nog helemaal leeg en ziet aan de linkerkant wat icoontjes staan (Figuur 2).



Figuur 2. Beginscherm Benchling.

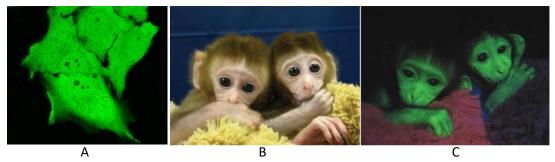
Je kunt je data netjes verzamelen in mapjes, zodat je deze ook weer makkelijk kunt terugvinden. Begin met het maken van een nieuw project door op de '+' te klikken en '*Project*' te kiezen. Noem het nieuwe project '<u>Bioinformatica 1</u>'. Sla alle data die je in het vak '<u>Bioinformatica 1</u>' maakt en gebruikt op in dit project.

TIP: maak gebruik van (sub)folders om het overzichtelijk te houden. Bijna alles wat je in Benchling importeert/creëert wordt bewaard in de map waar je in werkt en dat wordt al snel onoverzichtelijk.

Opdracht 1: Analyse van DNA-moleculen in Benchling

GFP Fusie-eiwit

Een onderzoeker heeft een fusie gemaakt van eGFP met het mutante Huntington gen. Omdat het eiwit groen fluorescerend is kun je het met een microscoop goed zien (Figuur 3A). Er zijn zelfs Huntington-GFP apen gegenereerd (Figuur 3B, 3C).



Figuur 3. Fusie-eiwit Huntingtin-GFP in cellen (**A**). Apen (en andere dieren) met het fusie-eiwit zonder fluorescentie (**B**) en met onder fluorescent licht (**C**).

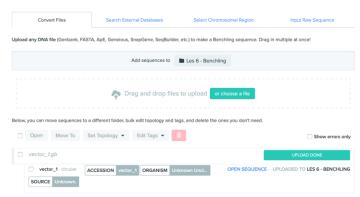
Het Huntington eiwit is gefuseerd met eGFP. Dit betekent dat er één groot eiwit ontstaat. Hiervoor gebruiken we een speciaal plasmide. Het plasmide bevat het GFP-gen zonder stop codon gevolgd door een multiple cloning site. Een multiple cloning site is een stuk DNA met veel restrictie sites bij elkaar. Met behulp van restrictiesites is het gezonde en het ziekmakende allel (een allel is een versie van een gen) gekloond in het plasmide. Met behulp van Benchling kun je controleren of er een insert in het plasmide zit. Daarnaast kun je ook controleren of het ziekmakende allel of het gezonde allel aanwezig is.

De bestanden die je nodig hebt voor deze les vind je op Blackboard in folder bestanden les3 benchlingin een zip-bestand. Pak het bestand uit in de map School\Thema02 die je in les 1 hebt gemaakt. Je gaat nu de bestanden openen in Benchling.

Openen/importeren DNA/RNA-sequenties

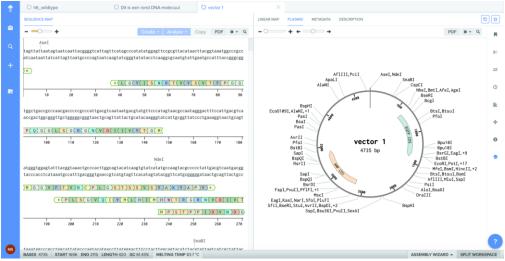
In Benchling kun je met zowel DNA-sequenties als met aminozuursequenties werken. We zullen beginnen met het laden van een DNA-sequentie.

Klik weer op de '+' om een molecuul te laden. Kies voor 'DNA sequences' en 'Import DNA Sequences'. Je kunt direct uit externe (bekende) databases DNA-sequenties (zoals NCBI gene) downloaden, maar je kunt ook een bestand van je eigen harde schijf uploaden (Figuur 4). We beginnen met de tweede optie (later in deze les komt de eerste optie ook aan bod). Van de (uitgepakte) bestanden van Blackboard kies het bestand **vector_1.gb**. Je kunt eventueel een map aanmaken ('Add sequences to' en vervolgens 'create new folder') waar het bestand in gezet wordt.



Figuur 4. Uploaden van bestanden in Benchling.

Dit bestand is een GenBank-bestand welke iets meer informatie bevat dan het FASTA-bestand. Zo staan onder andere Open Reading Frames en andere relevante informatie weergegeven in dit type bestand, welke mee worden geïmporteerd naast de sequentie. Zo staat bijvoorbeeld ook in dit type bestand dat het om een plasmide (circulair DNA) gaat en niet om een lineair DNA-molecuul. Je ziet nu twee vensters: links de DNA-sequentie en rechts de lineaire weergave van het plasmide. Door op het tabblad '*Plasmid*' te klikken, krijg je de circulaire weergave die gebruikelijk is voor een plasmide (Figuur 5).



Figuur 5. Benchling interface met <u>links</u> de sequentie en <u>rechts</u> de plasmide map.

De naam van het plasmide is vector 1 en is 4735 bp lang. Je ziet in beide vensters een aantal restrictie enzymen weergegeven. Met pijlen zijn de open-reading frames (ORFs) EGFP en AMP aangegeven. Een ORF is een deel van een sequentie dat begint met een startcodon en doorloopt gedurende een aantal gedefinieerde codons (100 is de default) voordat een stopcodon in hetzelfde frame gezien wordt. Ook in de DNA-sequentie zijn deze ORFs aangegeven. Je ziet de vertalingen aangegeven van potentiële ORFs (beginnend met ATG en eindigend met een stopcodon). De geannoteerde ORFs AMP en EGFP zijn dus ook aangegeven (in de DNA-sequentie moet je iets verder naar beneden scrollen om deze geannoteerde genen te kunnen zien). EGFP staat voor het "Enhanced Green Fluorescent Protein". AMP staat voor het ampicilline-resistentie gen.

In beide vensters is bovenin een 'tandwiel' te zien. Met deze knop kun je bepaalde kenmerken zichtbaar maken of weglaten. Probeer maar eens om de ORFs weg te laten bij de sequentie en zie wat het verschil is. Zo kun je zelf de voor jou meest overzichtelijke weergave gebruiken. Laat de ORFs voor de rest van de les weer gewoon zichtbaar staan.

Klik in het plasmide op het AMP CDS (Ampicilline Coding Sequence) en je ziet dat het gen geselecteerd wordt in de sequentie.

Vraag 1: Wat zijn de eerste 10 aminozuren (in 1-lettercode) van het **AMP**-resistentie eiwit?

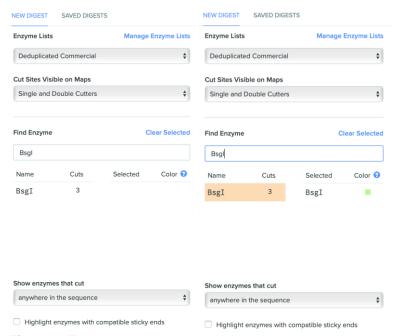
TIP: je kunt de eerste 10 aminozuren selecteren in de sequentie en met de knop 'copy' kopiëren in je antwoordblad.

Rechts in het tabblad van de plasmide map zijn een aantal knoppen te vinden waarmee je de sequentie(s) kunt onderzoeken en/of kunt manipuleren. Klik op de bovenste knop ('Annotations') en er verschijnt een venster met gegevens over de geannoteerde genen (dit nieuwe venster kun je weer wegklikken door op dezelfde knop te klikken). Je ziet een lijst met geannoteerde genen en je kunt de annotatie eventueel aanpassen.

Vraag 2: Wat zijn de start- en stopposities van het gen coderend voor **EGFP**? Voorbeeld antwoord: 123-456.

Restrictie enzymen en restrictie analyse in Benchling

Met een restrictie analyse kun je achterhalen hoe vaak een restrictie enzym knipt in een stuk DNA. Je gaat nu met vector 1 een restrictie analyse doen. Met het schaartje kun je een restrictie analyse nabootsen. Klik op het schaartje en een nieuw venster komt tevoorschijn. Je kunt hier zoeken naar specifieke restrictie enzymen en selecteren voor een digestie (Figuur 6). Je kunt ook zoeken naar alle restrictie enzymen die bijvoorbeeld drie keer knippen in de sequentie (vul in '3' in plaats van een restrictie enzym).



Figuur 6. Zoekvenster voor restrictie enzymen.

Zoek naar het restrictie enzym **MscI** en beantwoord de volgende vragen.

Vraag 3: Hoe vaak knipt het restrictie enzym **MscI** in vector 1? Geef alleen het getal als antwoord.

Als je het enzym selecteert in de lijst, krijg je meer informatie te zien over het enzym, inclusief praktische informatie voor in het lab.

- **Vraag 4:** Op welke positie(s) knipt het restrictie enzym **Mscl**? Als er meerdere posities zijn, scheid de posities met een komma.
- **Vraag 5:** Wat voor type uiteindes worden geproduceerd door Mscl?
 - A) 'sticky ends' of
 - B) 'blunt ends'

Het geselecteerde restrictie enzym blijft in de lijst staan. Om een nieuwe restrictie uit te kunnen voeren, moet het vorige restrictie enzym worden gedeselecteerd. Klik onder het kopje 'Selected' weer op het enzym (MscI).

Ga vervolgens op zoek naar het restrictie enzym **Eagl** en beantwoord dezelfde vragen als hierboven.

- **Vraag 6:** Hoe vaak knipt het restrictie enzym **Eagl** in vector 1? Geef alleen het getal als antwoord.
- **Vraag 7:** Op welke positie(s) knipt dit enzym? Als er meerdere posities zijn, scheid de posities met een komma.
- **Vraag 8:** Wat voor type uiteindes worden geproduceerd door Eagl?
 - A) 'sticky ends' of
 - B) 'blunt ends'

Open nu de andere plasmiden (**vector_2.gb** en **vector_3.gb**). Je kunt tussendoor switchen tussen de verschillende moleculen door op het juiste tabblad (bovenaan) te klikken. Vector 2 en vector 3 bevatten het Huntington gen gefuseerd met het gen dat codeert voor het GFP-eiwit. Het betreft het wild-type (vector 2) of het mutante allel (vector 3). Het mutante allel is langer. De lengte van het plasmide is af te lezen uit de plasmide map.

Vector 2 bevat een foute titel. Klik aan de rechterkant op het icoon met de letter 'i'. Je krijg nu informatie over het molecuul. Je kunt hier de naam van de vector veranderen naar 'vector 2'. Vergeet niet op 'Update Information' te klikken om de verandering van naam toe te passen.

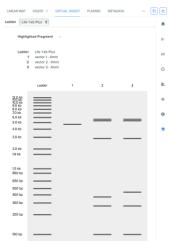
Vraag 9: Wat is de lengte van vector 2 in bp? Geef alleen het getal als antwoord.

Om onderscheid te maken tussen plasmiden kan gebruik gemaakt worden van restrictie-analyse. Benchling kan een simulatie uitvoeren van een restrictie reactie. Je weet dan precies welke fragment groottes je kan verwachten. Knippen gaat weer via het schaartje aan de rechterkant van het scherm. Knip vector 1, vector 2 en vector 3 met Xmnl (deselecteer eerst weer het restrictie enzym Eagl van de vorige opdracht). Zoek naar Xmnl, selecteer het enzym en klik op 'Run Digest'.

Vraag 10: Welke vector (plasmide) wordt niet geknipt door Xmnl?

- A) Vector 1
- B) Vector 2
- C) Vector 3

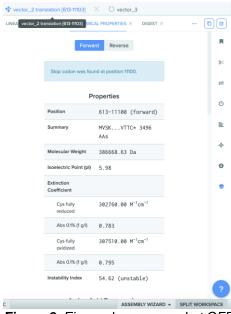
De resultaten van de restrictie zijn nu ook beschikbaar als tabblad en kun je altijd naar terug gaan, zolang je de browser niet afsluit. Na een restrictie verschijnt er ook een tweede tabblad 'Virtual Digest' (Figuur 7). Elke digestie die je in een sessie uitvoert wordt hier aan toegevoegd en geeft een beeld wat je zou kunnen verwachten op een agarose gel.



Figuur 7. Virtuele digestie van de drie verschillende vectoren met het restrictie enzym Xmnl.

Vraag 11: Kijk na de digestie van de drie vectoren bij 'Virtual Digest'. Welk bandje van vector 3 vind je niet terug bij vector 2? Je kunt op de bandjes klikken en zien hoe groot het fragment is (onder 'Highlighted Fragment').

Ga terug naar het venster voor vector 2 en selecteer het GFP-Huntington CDS in de plasmide map. Klik nu in het venster met de sequentie op 'Analyze' en kies 'Analyze As Translation'. Er verschijnt een nieuw tabblad met enkele biochemische eigenschappen van het vertaalde GFP-Huntington fusie-eiwit (Figuur 8).



Figuur 8. Eigenschappen van het GFP-Huntingtin fusie-eiwit.

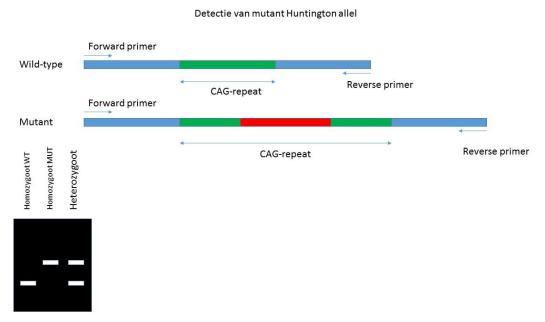
Doe nu hetzelfde voor vector 3 en het GFP-Huntington fusie-eiwit en beantwoord de volgende vragen.

- **Vraag 12:** Wat is het moleculaire gewicht in <u>kilo</u>Dalton van het GFP-Huntington fusie-eiwit van vector 3 (rond af op een geheel getal)? Geef alleen het getal als antwoord.
- **Vraag 13:** Hoe groot (in aminozuren) is het GFP-Huntington fusie-eiwit van vector 3? Geef alleen het getal als antwoord.
- **Vraag 14:** Hoeveel aminozuren verschil zit er tussen de fusie-eiwitten van vector 2 en vector 3? Geef alleen het getal als antwoord.

Opdracht 2: Simulatie van een PCR

Bij deze opdracht ga je een PCR simuleren van een mutant en een wild-type Huntington allel (Figuur 9). Het Huntington-gen bevat een CAG trinucleotide repeat. Je vindt dus een heel stel van deze repeat sequenties achter elkaar. Dit codon codeert voor het aminozuur glutamine. Een normaal aantal glutaminine repeats voor het Huntington-gen is 7-35 repeats. Vanaf 40 repeats is er sprake van de ziekte van Huntington op een bepaalde leeftijd.

De repeat-lengte kan veranderen van ouder op nageslacht. Hierdoor kan een nietziekmakende repeat-lengte van 30 in een volgende generatie plots langer zijn en kan het dan wel de ziekte veroorzaken. Doordat de repeat-lengte verschilt kan een diagnostische PCR duidelijk maken of een persoon drager is (en daardoor ook ziek wordt) van het mutante Huntington allel. In figuur 9 is een overzicht te zien van de test.



Figuur 9. PCR-test om het wildtype allel te kunnen onderscheiden van het mutante allel.

We gaan hiervoor een PCR simuleren met behulp van Benchling op de computer. Voordat je hier mee kan beginnen moet je eerst de DNA-sequenties in handen krijgen. Ga naar NCBI gene en download de mRNA sequentie van het Huntingtongen in FASTA-formaat. Downloaden van gensequenties heb je in een eerdere les geleerd. We gebruiken mRNA omdat de meeste genen van mensen intronen bevatten. Intronen zijn stukjes DNA waarvan de code niet in het uiteindelijke mRNA terecht komt. Ze worden dan ook niet vertaald naar eiwit.

Vraag 15: Zoek in NCBI de (RefSeq) mRNA (isoform 2) sequentie voor het Huntington-gen. Wat is de NCBI mRNA entry voor het Huntington gen? Voorbeeldnotatie: NM_123456.7

Het is nu misschien handig om de tabbladen van de vorige opdracht te sluiten, zodat het overzichtelijk blijft. De moleculen worden bewaard en kun je terugvinden aan de linkerkant van het scherm onder de knop '*Projects*' en kies dan de map die je hebt gebruikt om je bestanden op te slaan.

We gaan nu Benchling gebruiken om de sequentie te downloaden. Importeer dus weer een DNA-sequentie zoals je eerder hebt gedaan (voor vector_1.gb), maar kies nu voor 'Search External Database' (Figuur 10) en vul het antwoord van Vraag 15 in de zoekbalk in. Benchling herkent het nummer en gaat alleen in de NCBI (Gene) database zoeken. Andere databases komen in Bioinformatica 2 aan bod.



Figuur 10. Zoeken naar DNA/RNA-sequenties in externe databases.

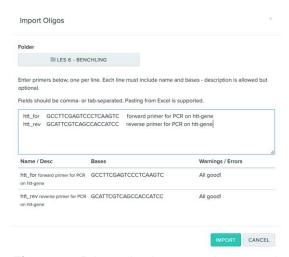
Selecteer de juiste map waar de sequentie in gezet moet worden en klik vervolgens op 'Import' om de sequentie te downloaden. Je hebt nu het de sequentie voor het wildtype Huntington gen. Verander de naam van het molecuul in 'htt wildtype'. Als je de entry codes van een gen weet, kun je dus direct in Benchling de sequentie downloaden zonder eerst naar een andere database te hoeven gaan.

Open nu ook het bestand **htt_mutant.gb** dat in het zip-bestand zit wat je van Blackboard hebt gedownload (Figuur 11). Dit bestand bevat de DNA-sequentie van het mutante Huntington gen. De annotatie van het ORF is ook zichtbaar in Benchling.



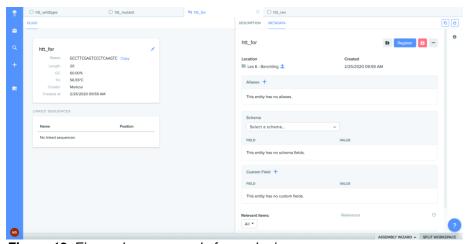
Figuur 11. Mutante Huntington-gen met annotatie.

Op deze twee moleculen kunnen we vervolgens een PCR-simulatie uitvoeren. Voor de PCR hebben we primers nodig die aan beide moleculen binden. Het ontwerpen van nieuwe primers leer je in het tweede jaar bij Bioinformatica 2. De primers die je gaat gebruiken staan nu in een Comma Separated Values (.csv) bestand (primers_htt.csv) dat in het zip-bestand van Blackboard zit. Open het bestand met Excel, kopieer alleen de rijen met de primers, sequenties en beschrijvingen (dus zonder de eerste regel). In Benchling, klik op '+' aan de linkerkant van het scherm en klik op 'Oligo' en 'Import Oligos' (Figuur 12). Vervolgens kun je de gekopieerde rijen uit het csv-bestand plakken in het vak in Benchling. Selecteer de correcte folders waar de primers in geïmporteerd worden en klik op 'Import'.



Figuur 12. Primers kopiëren vanuit een csv-bestand voor het gebruik in een PCR op het Huntington-gen.

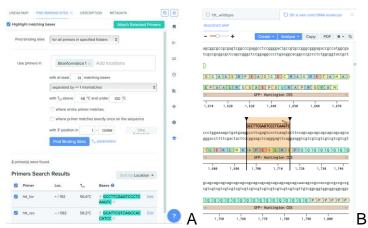
De primers openen niet direct in Benchling, dus moet je ze eerst opzoeken in de folder waar ze zijn geïmporteerd en vervolgens aanklikken (Figuur 13).



Figuur 13. Eigenschappen van de forward primer.

Je ziet nu een aantal eigenschappen van de primer, zoals het %GC en de Tm (annealing temperatuur). Onderaan zie je ook dat de primer nog niet gelinkt is aan een molecuul. In de volgende stappen linken we de twee primers aan de template (het wildtype en mutante Huntington gen).

Selecteer eerst het wildtype Huntington gen. Aan de rechterkant, klik nu op de twee pijltjes ('*Primers*'), klik op '*Attach Existing*' en selecteer het project waar de primers in staan. Klik vervolgens op '*Find Binding Sites*'. Benchling gaat nu zoeken waar de primers kunnen binden. Als het goed is worden de twee geïmporteerde primers gevonden en kun je ze selecteren onder in het scherm en linken aan de sequentie met '*Attach Selected Primers*' bovenin het scherm (Figuur 14A).



Figuur 14. Het zoeken van bindingsplekken van primers in de sequentie (**A**) en het linken van de primers aan de sequentie (**B**).

De primers zijn nu ook zichtbaar in het venster met de sequentie (Figuur 14B). Klik op de Forward primer, hou de SHIFT-knop ingedrukt en klik op de Reverse primer (scroll naar beneden om de Reverse primer te vinden). Het stuk tussen de primers is nu geselecteerd. Alternatief, kun je ook de primers in de lineaire map aanklikken, maar dan moet je inzoomen om de primers te kunnen zien. Zoomen gaat met de '-' en '+' bovenaan het venster. Klik aan de rechterkant van het scherm op de twee pijlen ('*Primers*') en kies het tabblad '*Pairs*' (Figuur 15).



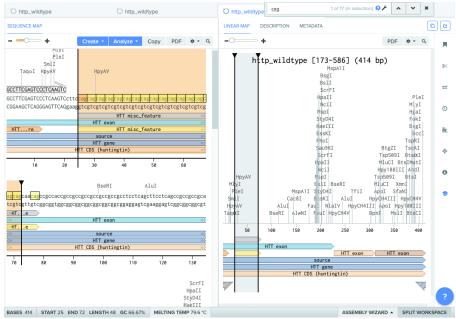
Figuur 15. Geselecteerde primers die binden aan de sequentie. Aangegeven zijn de bindingsposities, annealing temperatuur (T_m) en verwachtte productgrootte.

- Vraag 16: Wat betekenen de '+' en '-' die voor de posities staan?
 - A) '+' betekent dat de primer wel kan binden, '-' betekent dat de primer niet kan binden.
 - B) '+' betekent dat de sequentie van de primer in de coding (+) strand te vinden is en '-' dat de sequentie in de template (-) strand te vinden is.
 - C) '+' betekent dat de primer een positieve lading heeft en '-' betekent dat de primer een negatieve lading heeft.
- **Vraag 17:** Op welke positie bindt de reverse primer op het wildtype Huntington gen? <u>Let op</u>: de positie in bovenstaand screenshot is niet het juiste antwoord. Geef alleen het getal als antwoord.

Klik in het scherm op 'Create PCR Product', vervolgens op 'Copy' en kies een map om het PCR-product in op te slaan.

Vraag 18: Hoe groot wordt het PCR-fragment voor het wild-type allel? <u>Let op</u>: de positie en grootte in bovenstaand en onderstaand screenshot is niet het juiste antwoord. Geef alleen het getal als antwoord.

In de sequentie kun je zoeken naar een unieke volgorde van basen met CTRL + f (Figuur 16). Zoek nu naar de repeats die kenmerkend zijn voor het Huntington gen.



Figuur 16. Zoeken naar specifieke sequenties in Benchling. De gevonden zoekterm wordt aangegeven in de sequentie.

Vraag 19. Hoe lang is de CAG-repeat (hoeveel aaneengesloten CAG-triplets) van het wild-type allel? Geef het aantal repeats. HINT: selecteer na zoeken de laatste repeat en kijk in de zoekbalk welke repeat is geselecteerd.

Link de primers ook aan het mutante Huntington gen en beantwoord de volgende vragen.

- **Vraag 20.** Hoe groot wordt het PCR-fragment voor het mutante allel? Geef alleen het getal als antwoord.
- Vraag 21. Welke fragment groottes verwacht je voor een persoon die heterozygoot is? Geef alleen het getal als antwoord. Indien je meerdere fragmenten verwacht, scheid ze dan met een komma.
- **Vraag 22.** Hoe lang is de CAG repeat (hoeveel aaneengesloten CAG-triplets) van het mutante allel? Geef het aantal repeats.
- Vraag 23. Wat is het verschil in aantal repeats tussen wildtype en mutant? Vergelijk je antwoord met het antwoord van vraag 14.

Tot slot

Lever je antwoorden in op Blackboard voor de les Benchling. Ga bij jezelf na of je de volgende leerdoelen hebt behaald:

- DNA-moleculen analyseren met Benchling
- Een restrictie uitvoeren met Benchling
- Een PCR simuleren met Benchling