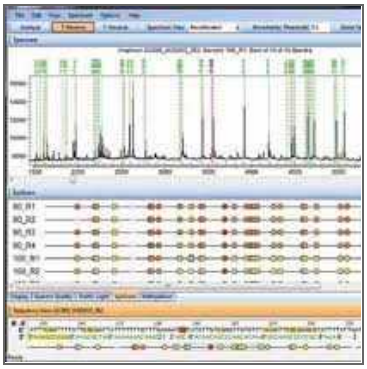


OCT 24, 2023

DOI:
dx.doi.org/10.17504/protocols.io.j8nlk4yk1g5r/v1

Protocol Citation: Louis-Stéphane Le Clercq, Desire Dalton, Antoinette Kotze, Paul Grobler 2023. Designing an EpiTYPER bisulfite sequencing assay for age estimation in Acinonyx jubatus based on human orthologues. **protocols.io** https://dx.doi.org/10.17504/protocols.io.j8nlk4yk1g5r/v1

MANUSCRIPT CITATION: Le Clercq, L.S., 2023. *Biological clock measures: Assessing the association between the circadian and epigenetic clock as predictors of migration phenology and biological aging in wildlife* (Doctoral thesis, University of the Free State).

Designing an EpiTYPER bisulfite sequencing assay for age estimation in Acinonyx jubatus based on human orthologues

Louis-Stéphane Le Clercq[1,2], Desire Dalton[3], Antoinette Kotze[1,2], Paul Grobler[1]

[1]University of the Free State; [2]South African National Biodiversity Institute; [3]Teesside University

Biological clock measures the association between the circadian and epigenetic clock as predictors of migration and age

Tech. support email: **leclercq.l.s@gmail.com**

Louis-Stéphane Le Clercq
University of the Free State, South African National Biodive...

ABSTRACT

Age is key factor in animal ecology as it can be used to assign animals to important age classes, ranging from immature young to reproductive adults and eventual old age and fragility. Different groups contribute to different aspects that need to be considered when modeling current and future population dynamics as part of continued conservation efforts. Due to the need of an accurate molecular method for assigning age, several studies have explore various aspects of epigenetic clocks. Epigenetics is a collective term for mechanisms that modify DNA and DNA packaging, independent of genetic sequence. One widely studied epigenetic feature is DNA methylation; a process that adds a methyl group to the 5’ cytosine of Cytosine-Guanine pairs (CpG’s). Studies have revealed that within genes, nearly a third of all CpG sites are influenced by age. Given its consistency, the epigenetic clock is a promising avenue of chronological age prediction which has been illustrated in many human studies. This protocol illustrates how CpG's with known age-correlations from human studies can be used to (1) identify orthologous regions in other species and (2) design primers to assay differential methylation using EpiTYPER mass array technology.

GUIDELINES

None

Protocol status: Working
We use this protocol and it's working

Created: Jun 15, 2021

Last Modified: Oct 24, 2023

PROTOCOL integer ID: 50801

Keywords: Methylation, Epigenetics, MALDI-TOF, Age determination, EpiTYPER, Human, Cheetah

Funders Acknowledgement:
National Research Foundation (RSA)
Grant ID: 112062

MATERIALS

- UCSC Genome Browser
- NCBI BLAST
- EpiDesigner website
- R
- RSeqMeth

SAFETY WARNINGS

None

ETHICS STATEMENT

Protocol approval for the present study was obtained from the protocol committee of the Department of Genetics, University of the Free State (approval number: Res18/2020). Ethics approvals were obtained from the University of the Free State (approval number: UFS-AED2020/0015/1709) as well as the South African National Biodiversity Institute (approval number: SANBI/RES/P2020/30). Appropriate research permits were also obtained from South African regulatory authorities including the Department of Agriculture, Land Reform, and Rural Development (Section 20 permit: 12/11/1/1/18(1824JD)) and the Department of Environmental Affairs (Threatened Or Protected Species (TOPS) permit: O-52903).

BEFORE START INSTRUCTIONS

You need to know the CG values for the CpG's you would like to design the assay.

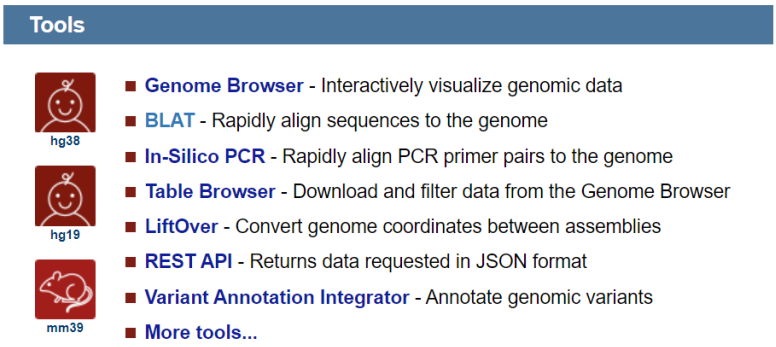
	Gene	CG value	Reference
	ASPA	cg02228185	Vidal-Bralo et al. (2017)
	EDARADD	cg09809672	Bocklandt et al. (2011)
	ELOVL2	cg21572722	Bekaert et al. (2015)
	FHL2	cg22454769	Giuliani et al. (2015)
	FUT3	cg17471102	Vidal-Bralo et al. (2017)
	ITGA2B	cg25809905	Weidner et al. (2014)
	GRIA2	cg25148589	Polanovski et al. (2014)
	PDE4C	cg17861230	Weidner et al. (2014)
	PENK	cg16219603	Giuliani et al. (2015)
	TET2	cg08924430	Polanovski et al. (2014)

Table of genes and cg names used to design assays.

Retrieving CpG from human sequencing

1 Most research papers on humans give a "CG" value which corresponds to a CpG site in the human genome based on Illumina sequencing. The following steps were used to retrieve the human sequence for reported CG values using the University of California Santa Cruz (UCSC) Genome Browser (https://genome.ucsc.edu/).

1.1 On the landing page for the website, select "Genome Browser".

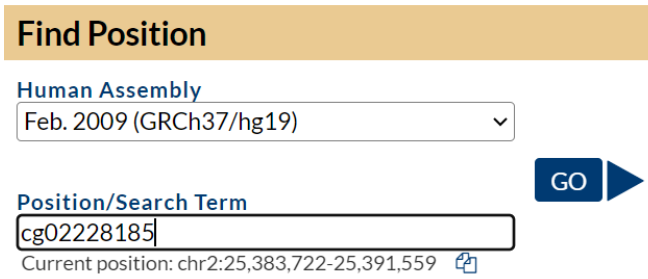


Genome Browser is listed under "Tools" on the landing page.

1.2 From the drop down menu, select the "Feb. 2009 (GRCh37/hg19)" genome build, as this is the

version with the mapped CG values.

1.3 In the "position/search term" box, enter the desired CG value to look up e.g., cg00123456 and click "Go" to perform search.

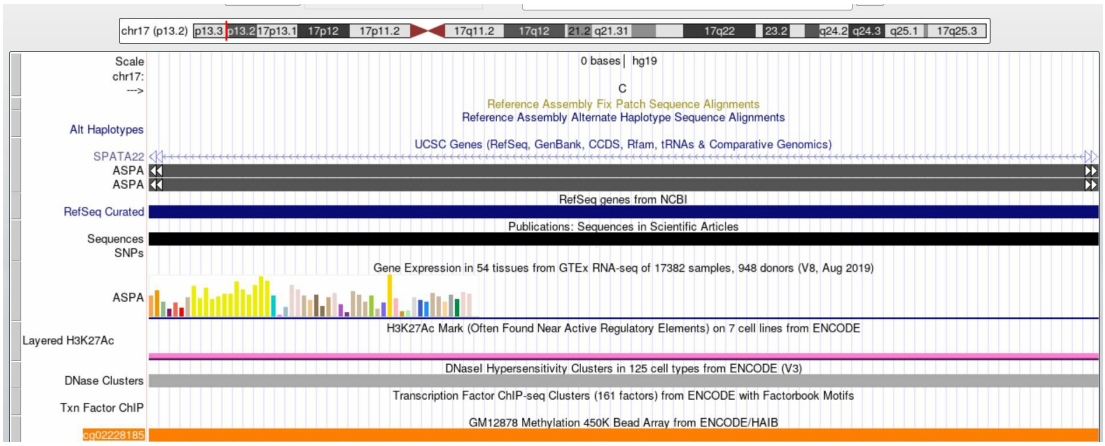


Example of search setup with the correct assembly and desired cg search term indicated.

1.4 The top of the results page will show the specific position e.g., Chr 19; 18.343,902.

1.5 The left panel of the scaffold should indicate a track for the CG value that was searched. Click on the panel (orange) for more details.

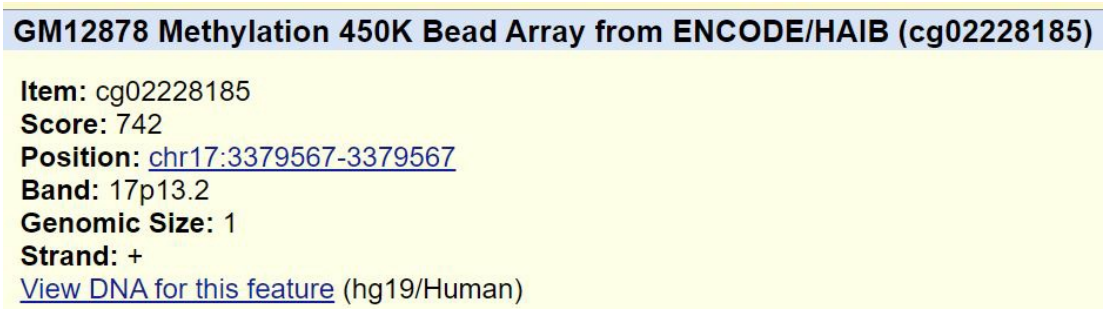
Expected result



Genome view of the CpG mapped on the human genome. The bottom of the genome view shows the specific cg (orange).

1.6 Click on "View DNA for this feature" and then select "Get DNA".

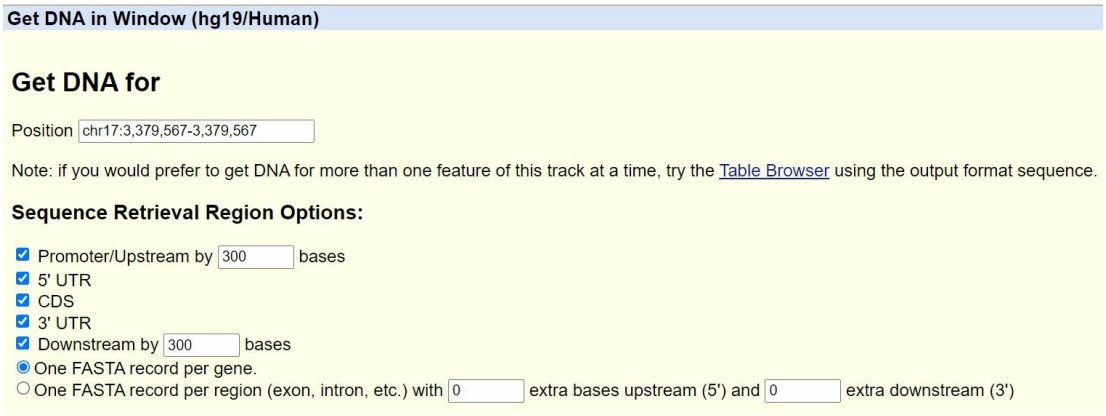
Expected result



New view that loads after clicking on the cg panel.

1.7 Specify to add 300-400 base pairs upstream and downstream of the CpG to ensure that the target CpG is in the middle and you have enough sequence to design primers.

Expected result



Menu that appears to select DNA sequence for export.

1.8 Save the sequence in the FASTA format.

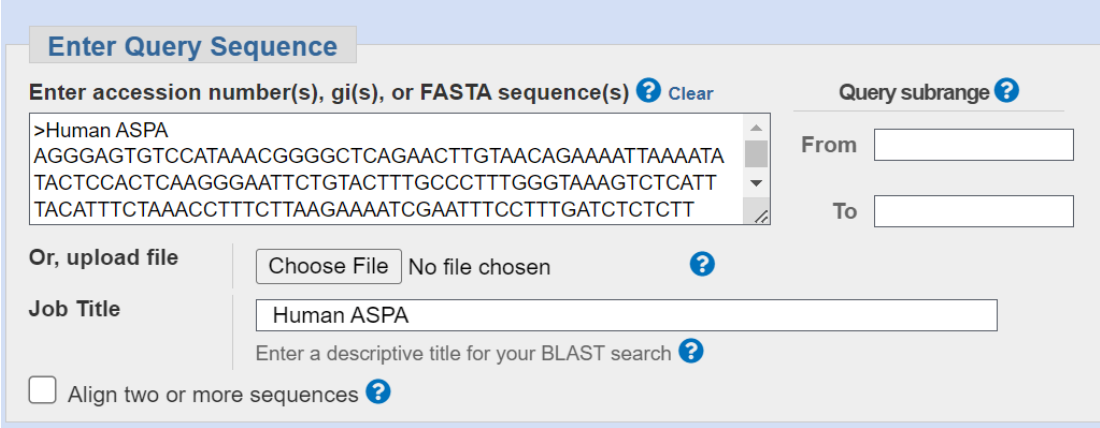
Finding the animal orthologues

2 The next steps are to find the orthologous gene sequence for the target species (e.g., *Acinonyx*

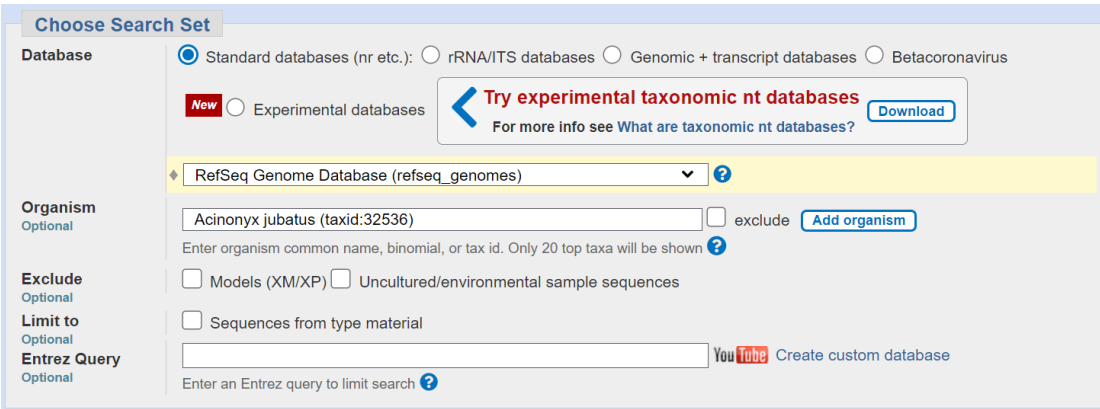
jubatus) using NCBI Blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

2.1 Select "Nucleotide Blast".

2.2 Copy and Paste the FASTA sequence into the sequence box or select the FASTA file to upload.

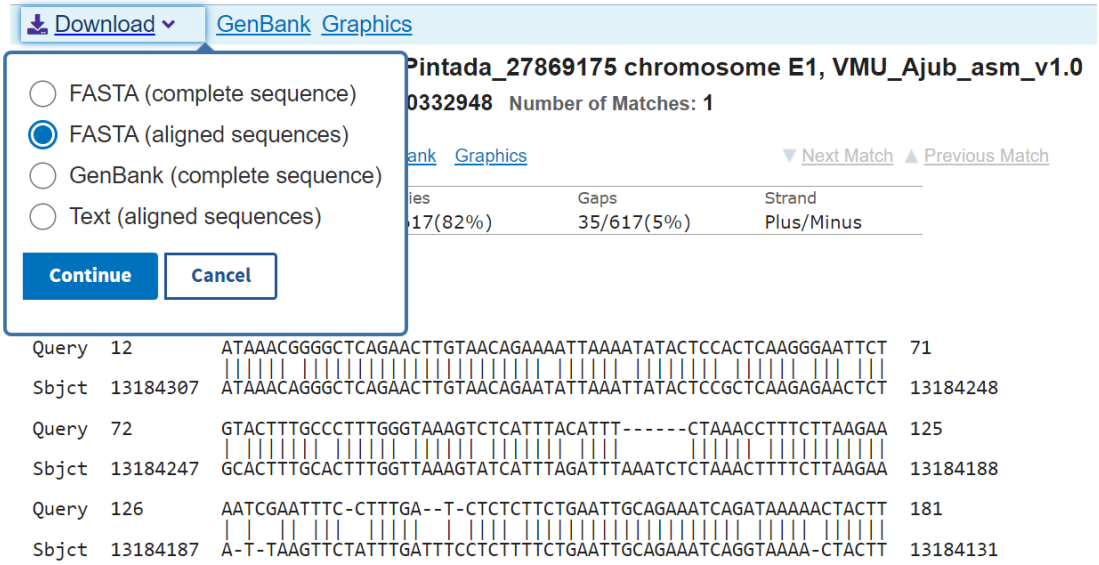


2.3 For "Database" select the "RefSeq Genome Database" and for "Organism" specify the target species e.g., Acinonyx jubatus (taxid:32536).



2.4 Download the FASTA for the complete aligned sequence.

Expected result



Note

Because the target CpG is located at the 300th base pair, be sure the resulting match covers that region.

Note

Tip: If the BLAST result does not cover the full length of the sequence used in the search (600 bp), you can view the result in the assembly viewer and use the sliders to select a region of around 600 bp that includes the BLAST match.

Related Information

Genome Data Viewer - aligned genomic context

Once you selected a specific BLAST match to view, the result page will have the option to see the match in an aligned genomic context on the right-hand side of the page.

Designing Primers for the EpiTYPER assay

3 The next steps are used to design the EpiTYPER primers on their EpiDesigner website (https://www.epidesigner.com/).

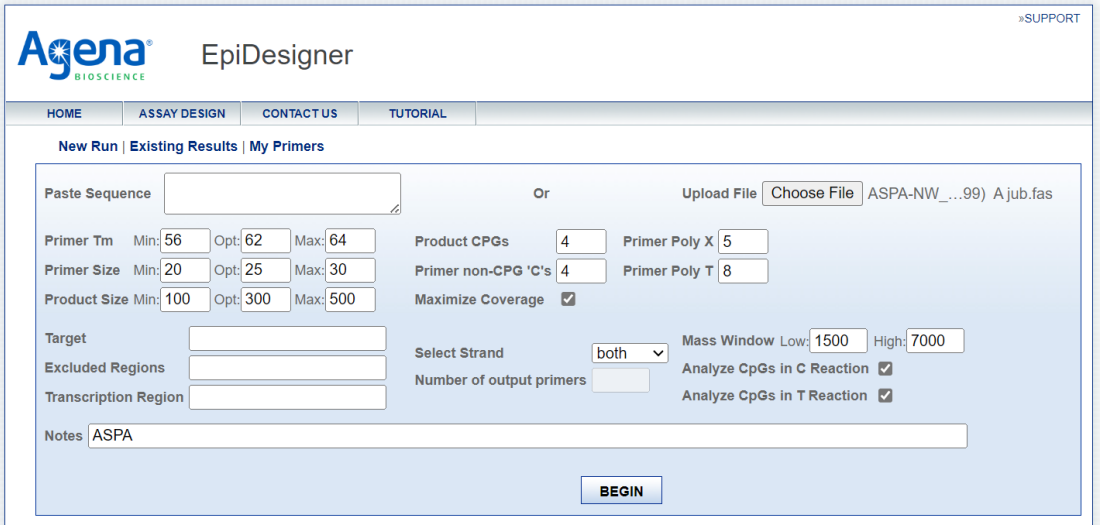
3.1 On the landing page, click "Start" to begin a new experiment design.

3.2 Once the input page has loaded, input the target sequence in FASTA format by either copying and pasting the sequence into the box or selecting the file.

> Note
>
> **Tip:** EpiDesigner seems to work better if you attach a file rather than pasting the sequence into the box.

3.3 Input the desired primer parameters or alternatively use the recommended base settings as is.

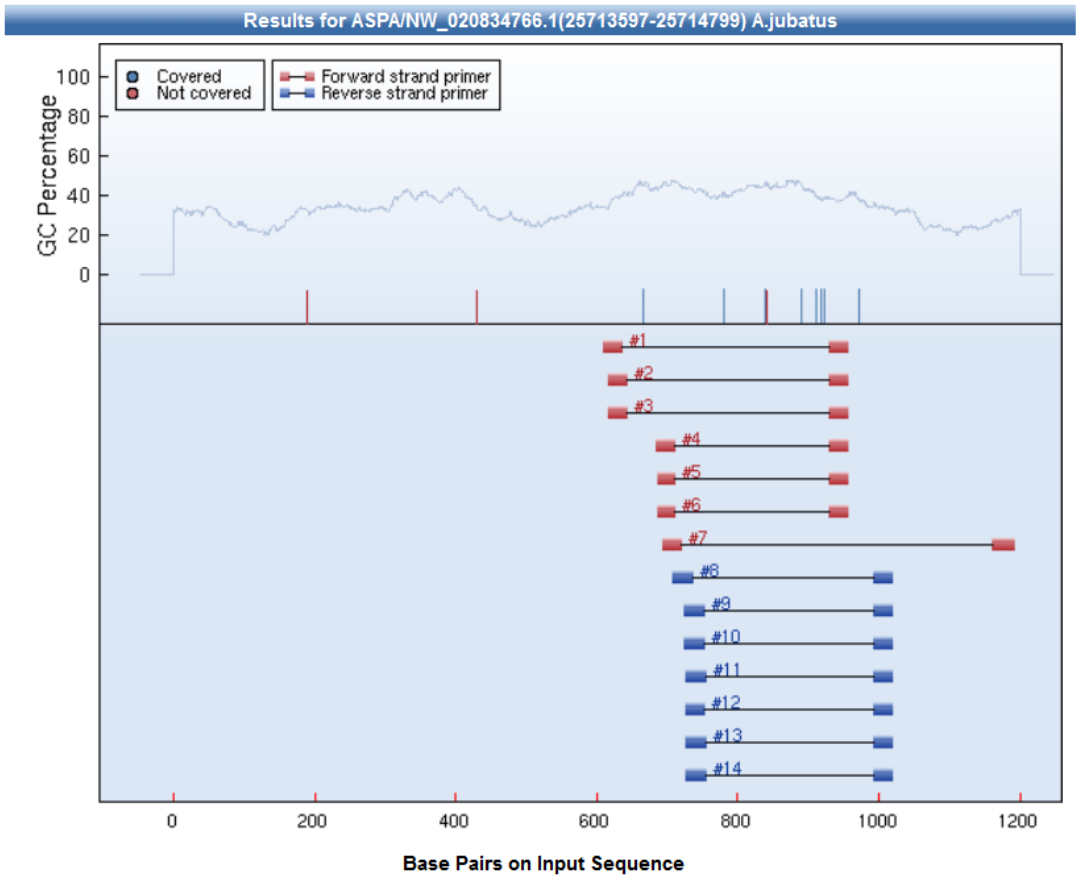
3.4 Leave the Target, Excluded, and Transcription Region (Advanced) setting empty to design primers across the full sequence. Click on "Begin" to design the primers.



Example of input screen for EpiDesigner. The sequence file is chosen and primer design was done by selecting both the forward and reverse strand. The gene name was used as a "Note" to keep track of results for different genes.

3.5 The results will appear starting with an interactive diagram of the sequence, the detected CpG's, and the mapped product for each primer pair.

Expected result



Example of diagram generated for the target sequence indicating the 14 possible amplicons for the region based on different primer pairs.

Below the image, the primers and their details are listed for each product shown on the diagram.

Expected result

		Primer	Start	Size	Tm	GC%	C's	Sequence
☐	1	LEFT PRIMER	612	25	59.56	28	4	TTGTTATTTTTGGAGGAATTTATGG
		RIGHT PRIMER	956	25	58.81	24	4	AAAAATAATCCATTTCACAACAACA
		PRODUCT Size: 345, No of CpG's : 8, Coverage : 5						
☐	2	LEFT PRIMER	618	25	59.97	28	4	TTTTTGGAGGAATTTATGGTAATGA
		RIGHT PRIMER	956	25	58.81	24	4	AAAAATAATCCATTTCACAACAACA
		PRODUCT Size: 339, No of CpG's : 8, Coverage : 5						
☐	3	LEFT PRIMER	618	25	59.97	28	4	TTTTTGGAGGAATTTATGGTAATGA
		RIGHT PRIMER	957	26	58.85	23	4	TAAAAATAATCCATTTCACAACAACA
		PRODUCT Size: 340, No of CpG's : 8, Coverage : 5						
☐	4	LEFT PRIMER	686	26	59.79	35	4	ATTGAGATTTAGAGAATAGGGTTGGA
		RIGHT PRIMER	956	26	59.94	23	4	AAAAATAATCCATTTCACAACAACAA
		PRODUCT Size: 271, No of CpG's : 7, Coverage : 4						

List of primers for each amplicon with positions, size, Tm, and sequences. Below each primer pair the product size and number of CpG's covered are indicated.

3.6 Potential primers that appear to be optimal can be selected by the left-hand tick boxes and exported in several format.

Testing CpG coverage for selected Primers

4 The final steps are performed in R (4.0.6) using RSeqMeth (https://github.com/cran/RSeqMeth) to determine the fragmentation patterns and elucidate which CpG's can be assayed.

Note

The desired amplicon from each gene, indicated in a column from the exported primer design results, needs to be saved as a plain text file e.g., "ASPA.txt"

4.1 Once you have downloaded RSeqMeth, open R and execute the analyses with the following code:

```
>source("Path to RSeqMeth\\R\\ampliconReport.R")
```

This is used to load the function is used to analyze the desired amplicon.
Then execute:

```
>ampliconReport("Path to amplicon text file.txt")
```

4.2 Six files are written to the same directory as the text file once completed, three for the analysis of "T spectra" and three for the analysis of "C spectra".

Expected result

```
> source("C:\\Users\\User\\Desktop\\RSeqMeth\\R\\ampliconReport.R")
> ampliconReport("C:\\Users\\User\\Desktop\\RSeqMeth\\ASPA.txt")
Report written to: C:\Users\User\Desktop\RSeqMeth\ASPA T Report.csv
Predicted spectra written to: C:\Users\User\Desktop\RSeqMeth\ASPA T Spectra.pdf
Predicted fragmentation written to: C:\Users\User\Desktop\RSeqMeth\ASPA T Fragmentation.pdf
Report written to: C:\Users\User\Desktop\RSeqMeth\ASPA C Report.csv
Predicted spectra written to: C:\Users\User\Desktop\RSeqMeth\ASPA C Spectra.pdf
Predicted fragmentation written to: C:\Users\User\Desktop\RSeqMeth\ASPA C Fragmentation.pdf
>
```

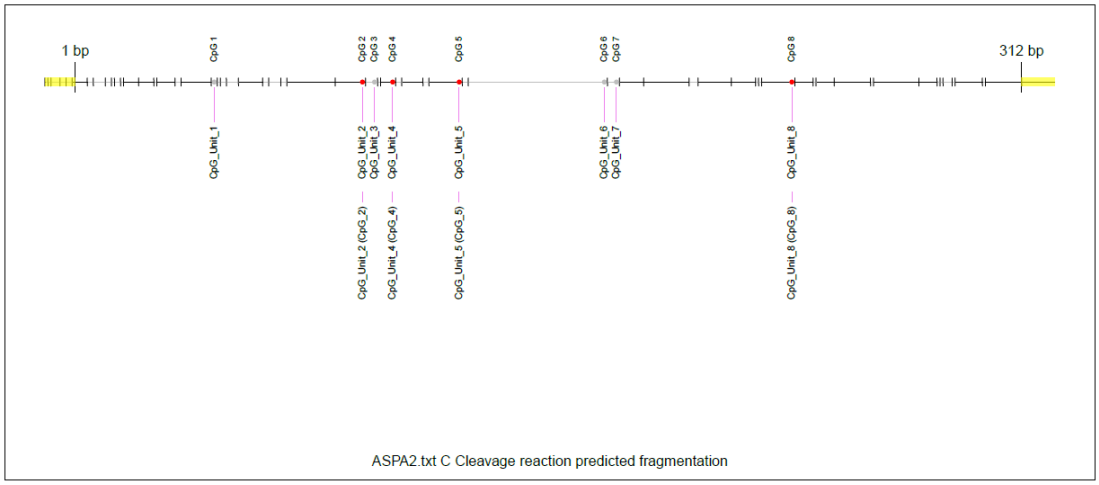
Example of code run and files written for the analysis of fragments.

Note

Depending on which sequencing kit you will be using you only need to look at either the results for the "T spectra" OR the results for the "C spectra".

4.3 The following results can then be viewed and assessed.

Expected result



Cleavage reaction predicted fragmentation showing clustering of CpG's per fragment. Clusters indicated in red are covered and can be tested while those in grey cannot be assayed.

These CpG's and their mass are listed in the table (CSV) output and indicates reasons why selected sites cannot be assayed such as overlapping fragment sizes, size duplication, or low mass.