Genetic colourant pathway Detection in *Chlorociboria aeruginascens*

MNEXT MATERIALS & ENERGY X TRANSITION

Nanopore mRNA sequencing to determine Xylindein producing pathway in *Chlorociboria aeruginascens*

Background

The move away from fossil fuels leads us to investigate biological alternatives. Blue colourant usually contains copper and harmful solvents. A blue colorant has been found in in the fungi *Chlorociboria aeruginascens* named Xylindein. This fungal colourant could be an eco-friendly alternative.

To find out more about the production of Xylindein, more needs to be known about the genetic pathway involved. Some research has been performed by the Westerdijk Institute. They found gene g432.t1 to be a suspect. The highest expression of this gene is thought to occur in the light blue phase.

This project aims to utilize Oxford Nanopore's MinION direct RNA sequencing together with bioinformatics to locate this gene, determine & its expression during different stages (figure 1). This data can be combined to find the pathway.

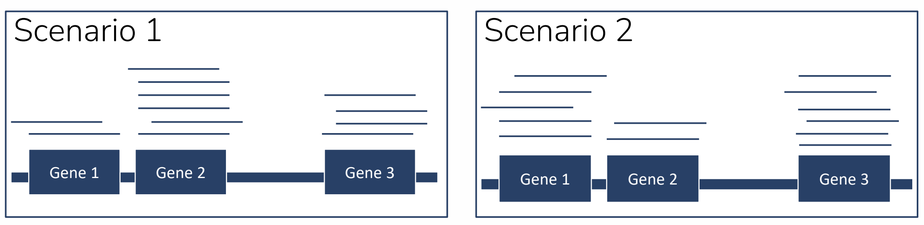


Figure 1 Differential Gene Expression

Method

Wet lab

Total RNA isolation was performed by the Westerdijk Institute. Two biological replicates of *C. aeruginascens* were grown; B & C. Each replicate consists of three samples; white, light blue & dark blue (figure 2). As the blue colour formed, isolation of these took place at 6 (white), 10 (light blue) and 15 (dark blue) days, respectively.

Sequencing took place using the Nanopore direct RNA sequencing kit and a custom barcoding protocol[1]. During the library prep a Spike-in was included in oadevery sample. 5 runs were performed, three of sample B and two of sample C. Before ling the concentration was measured using Qubit. Sequencing took place on a MinION Mk1C. Once the sequencing was completed the experiment was stopped and the data retrieved.

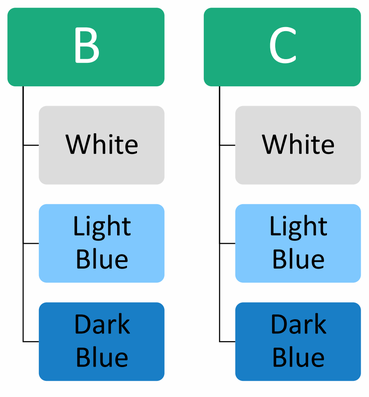
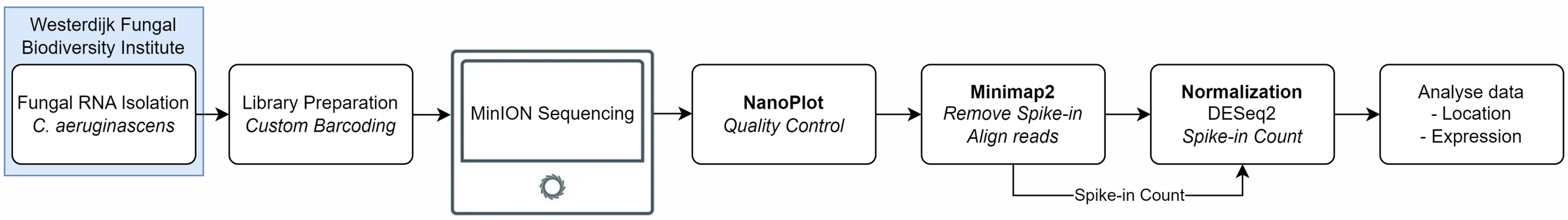


Figure 2 Biological Replicates

Dry lab

The dry lab work took place on a digital working environment run on Ubuntu. First the data was demultiplexed. Quality control was performed using NanoPlot. After sorting by barcode, the Spike-in was removed using Minimap2. After another quality control the data was aligned to a reference genome provided by the Westerdijk Institute using Minimap2. The expression per barcode was then counted by Featurecounts. Normalization of the data took place using the Spike-in data. Normalization was verified using DESeq2. The alignment made with Minimap2 against the reference genome was visualized using IGV[2].

Workflow



Results

The quality control of the reads was performed using Nanoplot. The data of the run was pooled per biological replicate. The important data was selected from the reports and visualized (table 1). The overall quality and read length of the white and lightblue samples of B is lower than the rest of the sequencing data. The read count of the white replicates is also lower overall across the runs. The Spike-in was removed using Minimap2 and counted. This data is used for normalization.

Table 1 Summary of relevant sequencing information of the pooled runs and barcodes

Sample	white-C	lightblue-C	darkblue-C	white-B	lightblue-B	darkblue-B
Number of reads:	158481	328674	298379	39998	226700	221139
Number of bases:	125630024	227717854	240540625	15939213	78018707	144008318
Mean read length:	793	693	806	399	344	651
Mean read quality:	13,3	12,9	13,3	10,8	10,9	12,6
Spike-in read count:	13950	18676	13394	1250	6066	3386

After alignment to the reference genome and the reads were visualized in Integrative Genomics Viewer (IGV)[2] by supplying an annotation file. The suspected involved gene g432.t1 (contig 16, position 1,741,058 - 1,748,197) was located (figure 3).

The expression (number of reads that align with the target gene) was determined using Featurecounts. This data was normalized using the incorporated spike-in data. A separate file was creating storing the data of all individual runs and barcodes.

The expression of the target gene between the different biological replicates can be seen below in table 2.

Table 2 Expression of g432.t1 gene in biological replicates

Expression	B	C
White	22,9	2,5
Light Blue	7,9	6,8
Dark Blue	70,0	24,1

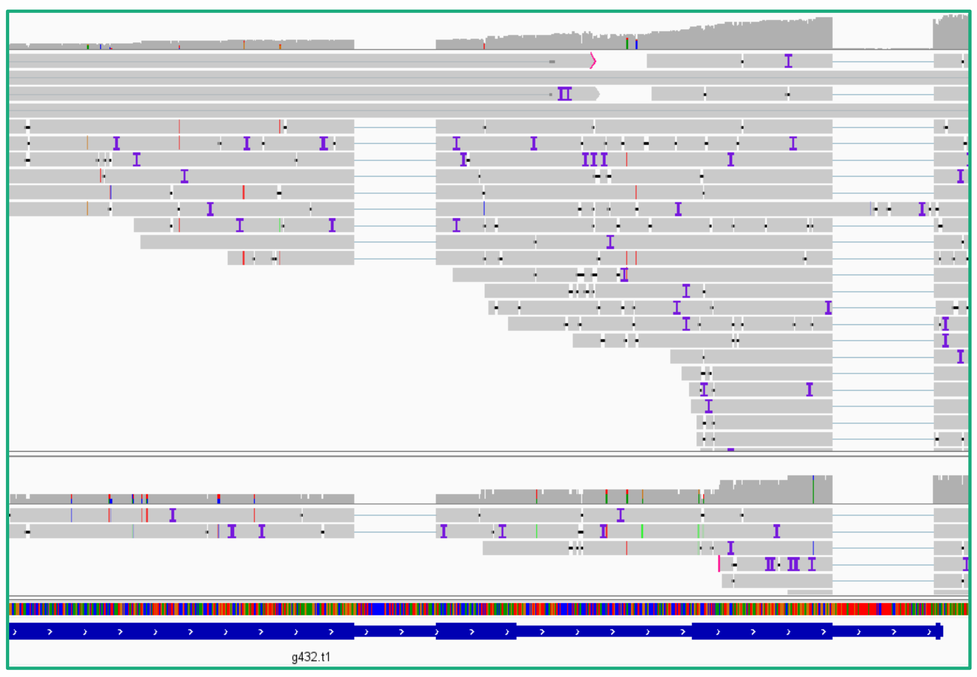


Figure 3 Part of the visualization of alignment to the target gene g432.t1 in IGV.

Discussion

The lower quality of white-B and lightblue-B could have several explanations. The sequencing could be hindered because of proteins, chemicals or solvents from the RNA isolation. Due to the amount of sample provided, replicate B was thawed and frozen more often than C. This could also contribute to diminishing quality.

As of yet the normalization using the spike-in method has not been validated by DESeq2 yet. This makes the expression values (table 2) less reliable. Though this data does convey that expression of the gene does not occur mostly in the light blue phase. This could mean the gene is downregulated though. This is most likely not true because that would mean the fungi would be blue to start with. Makes it so the validity of the normalization is questionable; or the hypothesis is incorrect.

Another point of discussion is that one of the exons present in the target gene g432.t1 shows coverage, where a gap would be expected. This is because exons are excluded in mRNA. This could be because non-spliced RNA was sequenced. Another possible explanation could be that the annotation file is incomplete or incorrect. This is possible cause it is not an official annotation file, but one constructed and provided by the Westerdijk Institute. A solution could be to BLAST (Basic Local Alignment Search Tool) the gene sequence against public databases to find similar genes.

References

1. Smith, M. A. *et al.* Molecular barcoding of native RNAs using nanopore sequencing and deep learning. *Genome Res.* **30**, 1345–1353 (2020).
2. Robinson, J. T. *et al.* Integrative Genomics Viewer. *Nat. Biotechnol.* **29**, 24 (2011).

Bart van den Hout
Research Group: BBB&P TUFUCOL
Contact information: ba.vandenhout1@student.avans.nl
Date: 18 December 2023