

Oxford Nanopore Technology (ONT) Library Preparation and Sequencing

Vanni Benvenga
Bioinformatician
Applied Microbiology Lab
PI: Prof. Dr. Dr. Adrian Egli
Institute of Medical Microbiology
University of Zurich

Learning outcomes

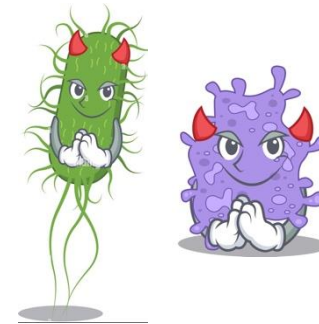
- You know how DNA must be prepared for ONT sequencing
- You will be familiar with technical terms such as **coverage** and **depth**
- You are able to calculate the amount of DNA needed for an ONT library
- You know the principle behind nanopore sequencing

Libraries are collections of DNA fragments ready for sequencing

- Before sequencing, DNA must be prepared to be compatible with the sequencing platform of choice.
- DNA is fragmented, end repair, the addition of barcodes, and pooling of DNA into a library

What are we doing today?

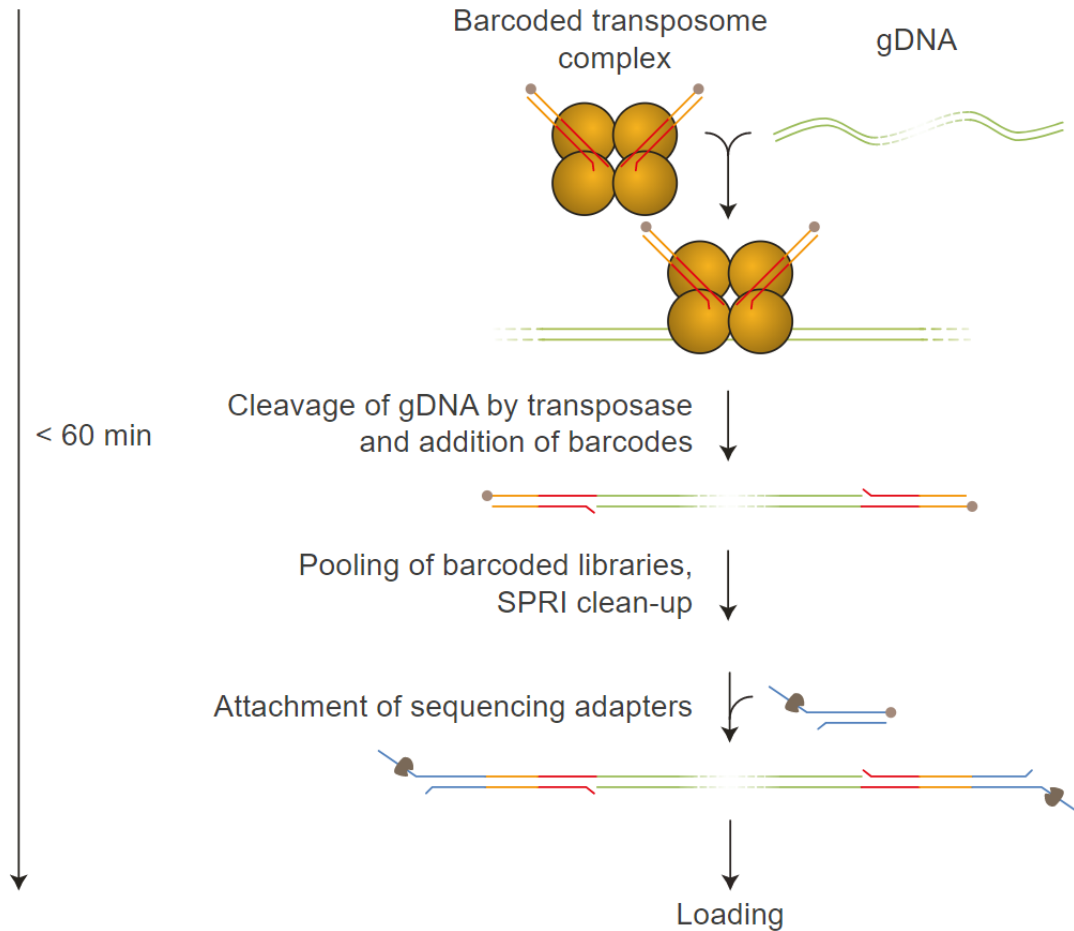
- S. aureus and E.coli - DNA extracted - QIAamp DNA Mini Kit (Gram+ve protocol).
- Library prep - Rapid barcoding kit (RBK114.24).
- Sequencing platform of choice - Illumina and ONT (GridION).
- Single isolate sequencing - Whole Genome Sequencing (WGS).



Why are we doing WGS?

- Who is on the plate? - Species identification.
- To detect resistance and virulence genes.
- Transmission Events and Outbreaks.

How are we doing it?



E. coli

Genome Size: ~5 Mbp
Gram-negative
UTI, virulence



S. aureus

Genome Size: ~2.8 Mbp
Gram-positive
MRSA, transmission

UTI - Urinary tract Infection

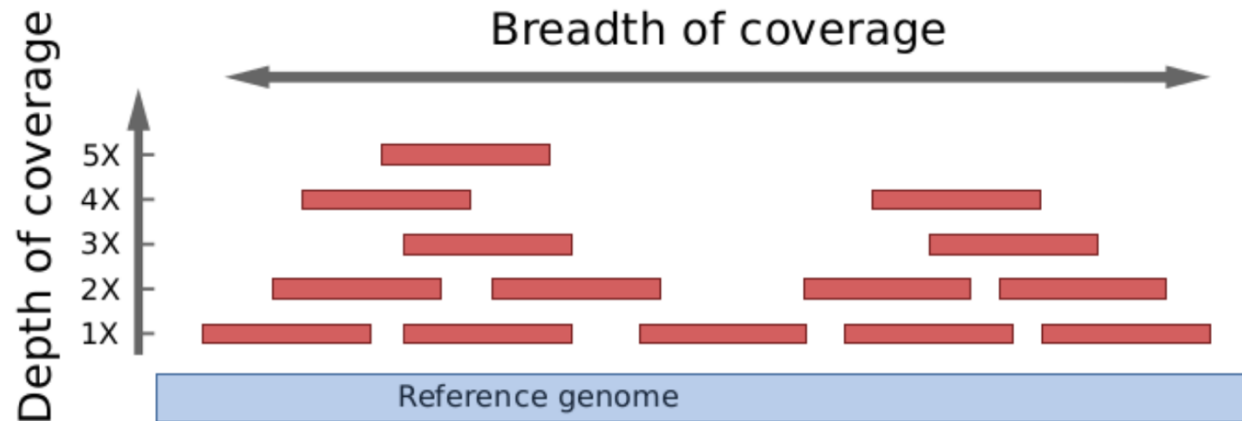
MRSA - Methicillin Resistant *Staphylococcus aureus*

Terms and calculations!

Multiplexing - the pooling of prepared libraries to be sequenced simultaneously on the same flow cell.

How to identify your sample after pooling?

Barcoding - unique sequence identifiers (differs with sequencing platforms and library kits), ligated to the DNA sequence before pooling.



Coverage/Breadth of coverage - how much of the sample (genome) is covered by sequencing.

For Illumina: $C = (L * N) / G$

C - coverage

L - Read length

N - Number of reads

G - Genome Size (Haploid)

Depth of sequencing/Read Depth/Depth of coverage:

How many reads detected a specific nucleotide - reliability of a base call.

You are studying biology to learn math, right?

Example: *S. aureus*

- Multiplexing 9 samples together
- The final volume allowed for each sample is 10 µl
- Sample SaH2P90822 - 10.3 ng/µl
- Sample SaH1P10622 – 118 ng/µl

Equal chance of sequencing - Same starting concentration

$$V_1N_1=V_2N_2$$

$$10 \mu\text{l} * 10.3 \text{ ng}/\mu\text{l} = 103 \text{ ng}$$

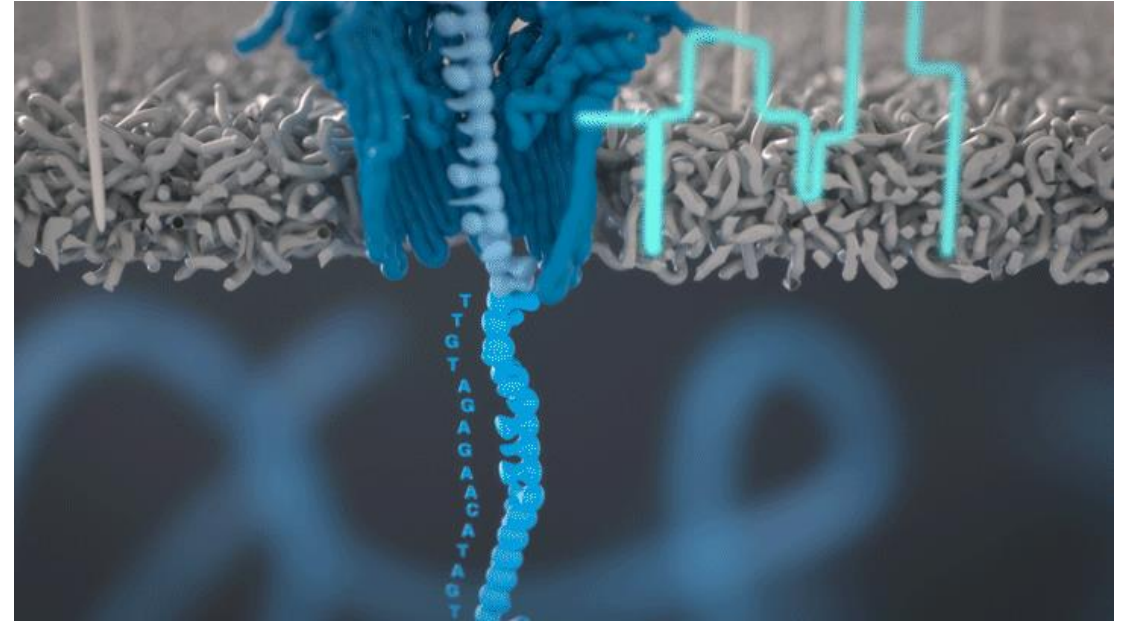
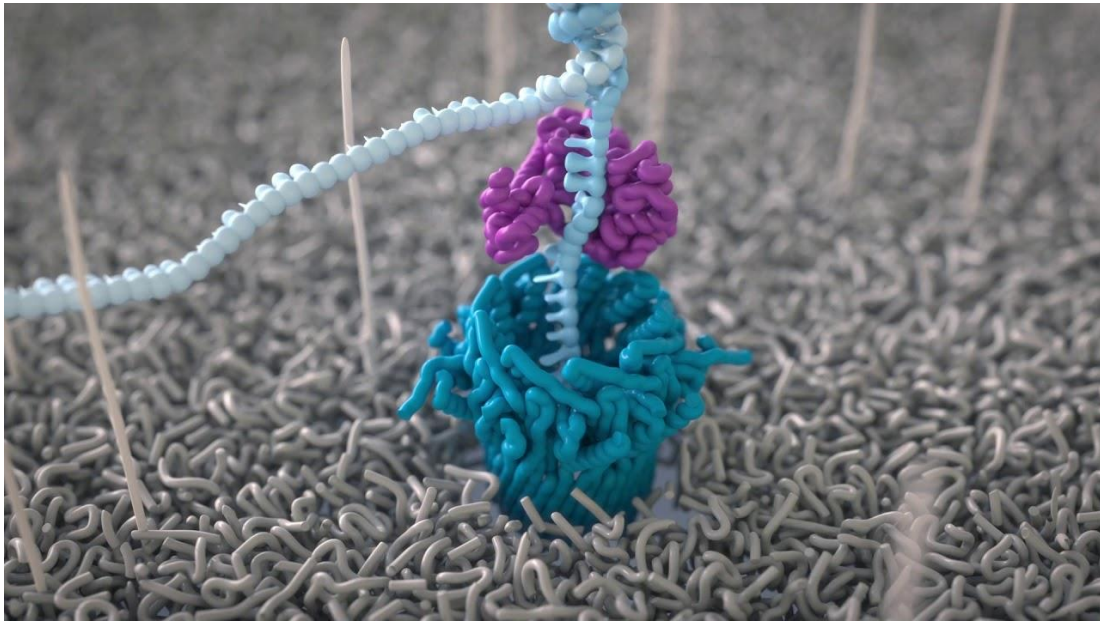
In 10 µl we will have 103 ng

$$X \mu\text{l} * 118 \text{ ng}/\mu\text{l} = 103 \text{ ng}$$

$$X = 0.9 \mu\text{l}$$

To add: 10 µl, Nuclease Free water (NFW) = 9.1 µl

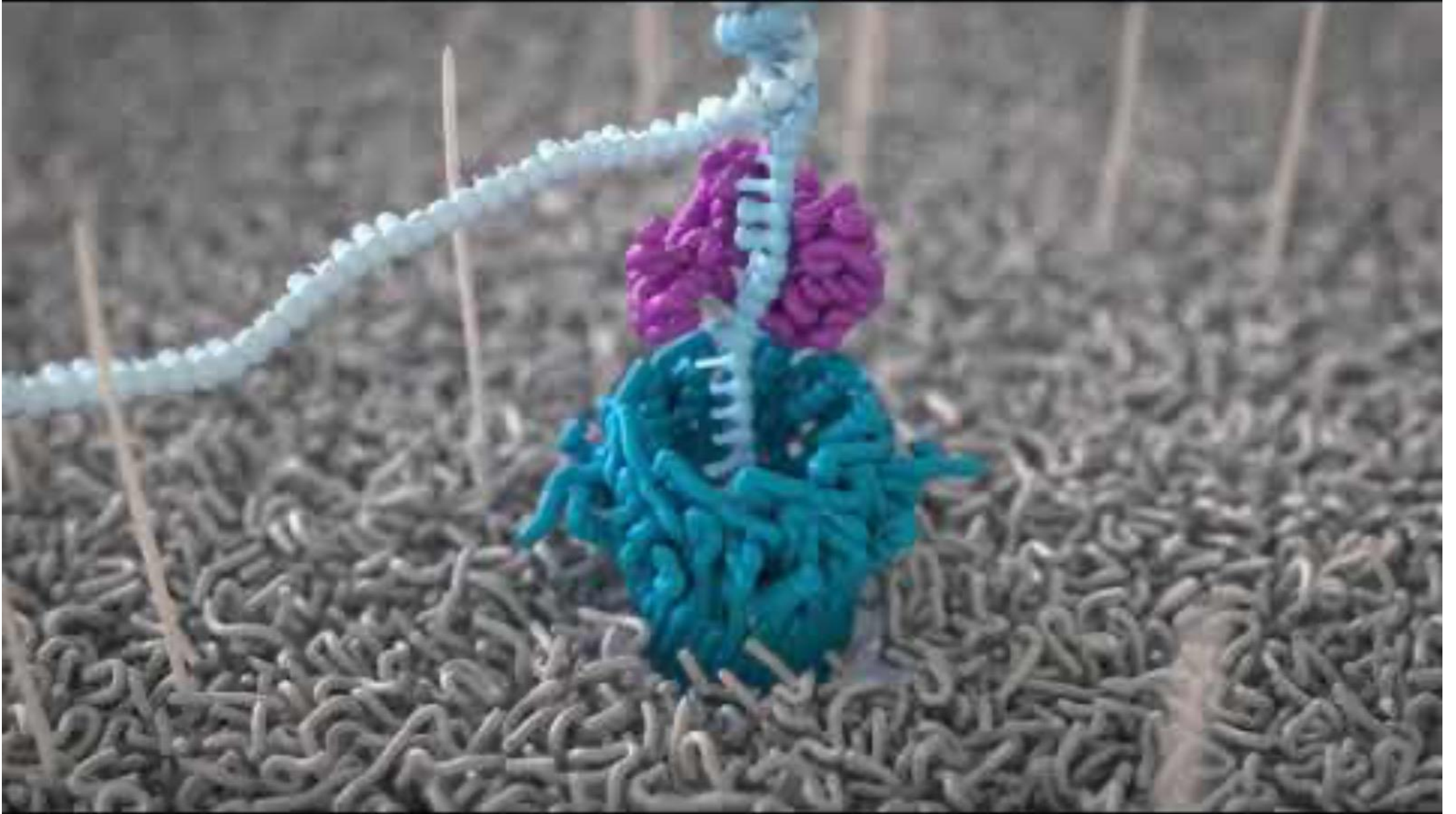
Oxford Nanopore Sequencing



Sequencing run time = 72h

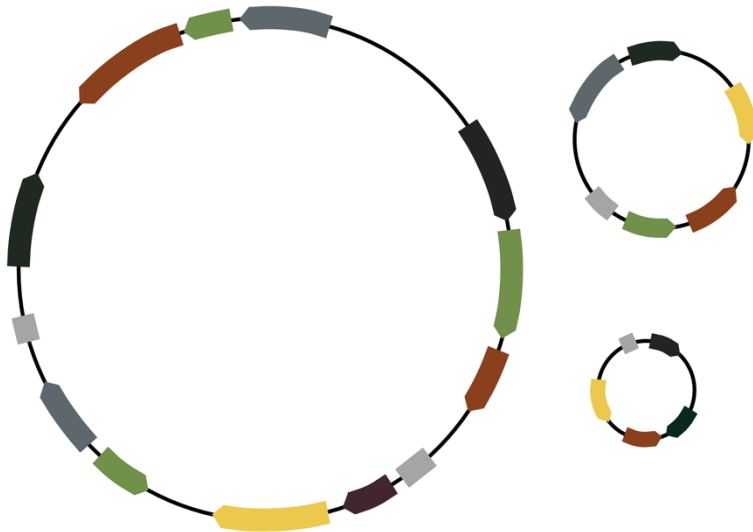
The squiggle form (FAST5/POD5) converted to FASTQ (FASTA file with Quality score) - **base calling** – Dorado – SUP (Super accuracy model - Q20 Simplex).

Video - [here](#)



The Alphabet Soup

What we want



What we get

```
@063352cc-49de-4b04-a309-fff46dd99057 runid=7cfd1f558cd29d6cbbb183ff269af3f987521ca6
read=18 ch=29 start_time=2023-02-03T12:07:26.683366+01:00 flow_cell_id=FAS77291
barcode=barcode01 barcode_alias=barcode01 basecall_model_version_id=2021-05-
17_dna_r9.4.1_minion_384_d37a2ab9
AGCTTTTCATTCTGACTGCAACGGGATTAAAAAAGAGTGTCTGATAGCGCTACCAATGCTGAAAC
TGCTGCCGGCAGACGATACCCGGCGGGCGGTGATAGAGCGCAGTATCCGGCTGGCGCAGGA
TGAGTTTGGCCAATCGATCATGCATGCAATGGTAACCGGAGTGCAATCTACGTACTACCATGGGA
AAATTGAAATAATGCCCTGACGGCCCGGTGGCGTGGCGCCATCGATCTAACTAGCACCACCATAT
TACCACAGGTAACGGTGCGGGCTGACGCGTA
+
"((68:<;977::*("("&%$$"#$$&&&+00<>>=>=>+414<=<<?>7547C?>>350.267;E::<>@BKC@><;20/+
**{,13@@<;::<=FDF?A763320123966><<>?9.'''*+*+,-
;=>><<=:99;=<<??AB@====BB>B>=>=?=<=?@C@B))))49;=@6556BB?::>A@>>>>@@@???A<
<<;<::<>=:799
```