

Color Balance Calculator

Write a Java program to calculate the „Color-Balance“ of DNA-sequences.

Background:

In Next Generation Sequencing, multiple DNA sequences are read simultaneously, base-by-base by a sequencer. In the first cycle, the first base of multiple DNA sequences is read.

A chemical solution that contains light sensitive molecules is added. Depending on which the first base of the DNA sequence is, different molecules attach to it. The sequencer uses two types of lasers to capture this event. First, the red laser is fired, and the first photo is taken. Then, the green laser is fired and the second photo is taken. The following table shows how each laser ‘sees’ DNA bases.

DNA-Base	Red-Laser-Photo	Green-Laser-Photo
G	Not visible	Not visible
T	Not visible	Lights up
C	Lights up	Not visible
A	Lights up	Lights up

A new chemical solution that removes the first DNA base from the sequence is added by the sequencer and the second cycle begins. (You can search for „sequencing by synthesis“ or „NGS sequencing“ on the internet, for a full explanation.)

The Software:

We need to evaluate combinations of artificial DNA-sequences („indexes“): in photos of every cycle (in both, the red-laser-photo and the green-laser-photo) at least one DNA-sequence / base should be visible.

Please write a simple program, that will receive the following 3 DNA-sequences as input (its up to you to decide, in which format the input needs to be supplied (e.g. without numbers and spaces) or if you prefer a web application or the command line reading from stdin or file).

1. G T C A G T C A
2. A G T A G T A C
3. C T C T G A C A

For each cycle from 1 to 8, the software should output, how many of these 3 sequences will light up on the red-laser-photo, and how many will be visible on the green-laser-photo.

A correct output would be:

- red / green
- 1. 2 / 1
- 2. 0 / 2
- 3. 2 / 1
- 4. 2 / 3
- 5. 0 / 0
- 6. 1 / 3
- 7. 3 / 1
- 8. 3 / 2

You are free to present the output / information in any other way, this is only an example-output.

The software should be able to read and process a variable amount of DNA-sequences (more or less than 3) of variable length (more or less than 8 cycles).

Please send the source code as plain text, or as a link to a git / github repository to: f.jostock@dkfz-heidelberg.de
You may also contact this email address with any questions.