

An automated workflow for high throughput MLVA using the BioNumerics® software

L. Vauterin, P. Vauterin

Applied Maths NV, Keistraat 120, B-9830 Sint-Martens-Latem, Belgium

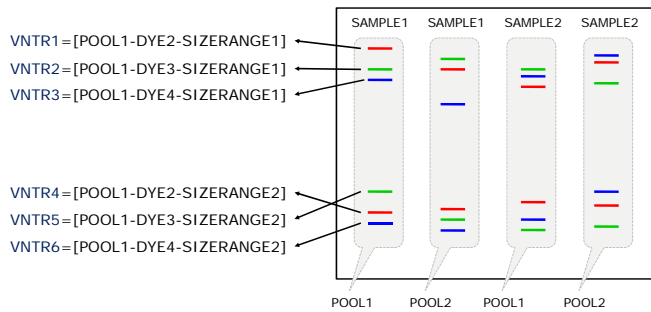
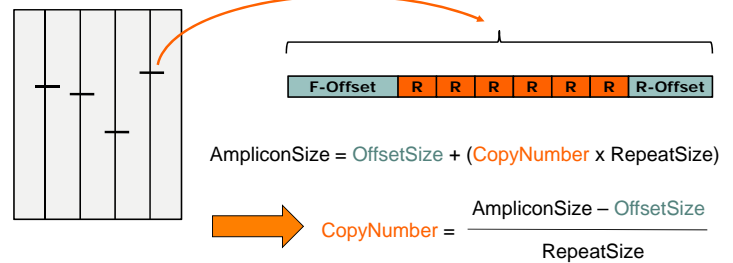
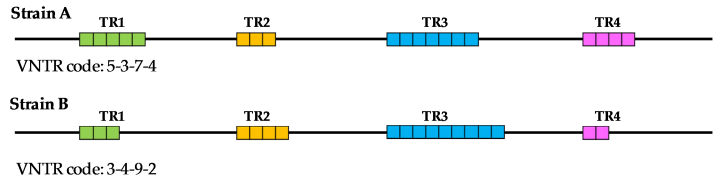
INTRODUCTION: Multi Locus VNTR Analysis (MLVA) is a method to sub-type microbial isolates based upon the Variable copy Numbers of Tandem Repeats (VNTR). The use of capillary sequencers for the determination of the VNTR copy numbers allows for rapid and inexpensive high-resolution typing of microbial isolates. The cost- and time-effectiveness of the technique can even be increased by pooling multiple VNTRs with non-overlapping size ranges. However, the resulting complexity of the pooling schemes and the amount of data generated require an automated workflow for the acquisition and processing of the trace files. We report here on an automated solution for MLVA using the BioNumerics® software.

PRINCIPLE: A VNTR typically exhibits a large range of copy numbers, even among highly related bacterial strains. For a selected set of tandem repeats, comparison of the copy numbers between bacterial strains can be used to obtain insight about the relationships at a micro-evolutionary level.

In practice, VNTR loci are selected that are sufficiently and complementary discriminatory for the organisms studied, and conserved primers are designed outside the tandem repeat for each VNTR. Thus, the size in bp of each PCR-amplicon is the sum of the size of the tandem repeat plus the offsets at both ends. Knowing the repeat size, the copy number can easily be calculated as

$$\text{CopyNumber} = \frac{\text{AmpliconSize} - \text{OffsetSize}}{\text{RepeatSize}}$$

For economy reasons, several VNTRs are sometimes pooled, i.e. they are marked with the same dye and loaded as a mixture in the same column of a capillary sequencer. A condition is that the mixed VNTR PCR products have size ranges that do not overlap. E.g., using 4 dyes and 2 non-overlapping VNTRs, 6 VNTRs can be determined per capillary run (one dye contains a reference sample for size calculation).

DATA PROCESSING: The MLVA setup has to be entered initially in the software. This involves entering the pooling strategy: a *pool* is a mix of VNTR amplification products loaded together in the same capillary. This includes the different dyes used and optionally, the compatible VNTRs with non-overlapping size ranges. Thus, each VNTR is defined by a pool, a dye and (optionally) a size range. The size range is defined by the **repeat length**, the **offset** and the **copy range**. As such, the software knows exactly within which size range it should look for a specific VNTR. Note that the copy range is only important in case different VNTRs are pooled with the same dye.

Due to experimental error, real VNTR amplicon sizes often differ slightly from expected sizes based upon copy numbers. Therefore, a **tolerance** can be entered in bp. Obviously, the tolerance should always be less than RepeatSize/2.

The software can process tab-delimited peak table files from AB (Genemapper) and Beckman. Sample names, pools and dyes are parsed automatically based upon a parsing string.

RN	dye	est frag size (nt)	pk height (rfu)	...
IS001_P2.B05_021201Z	D2	60.17	2015	...
IS001_P2.B05_021201Z	D2	390.11	2525	...
IS001_P2.B05_021201Z	D3	70.05	2410	...
IS001_P2.B05_021201Z	D3	400.51	2458	...
IS001_P2.B05_021201Z	D4	80.48	2153	...
IS001_P2.B05_021201Z	D4	280.48	2153	...

Sample Pool ID Dye Peak size Peak height

Once the settings for VNTRs and parsing have been entered, the software can automatically process thousands of MLVA runs, thereby creating reports listing unresolved VNTRs and any other problems. The resulting VNTR information is stored in integer-type character sets where each VNTR represents one character.

ANALYSIS: VNTR data can be analyzed as **categorical** characters (each different copy number is a different allele) or as **quantitative** characters. In the latter case, the larger the difference between copy numbers, the less related the organisms are considered. The Minimum Spanning Tree algorithm applied on VNTR data in BioNumerics has proven to be invaluable for epidemiological study and population genetics of bacterial populations.

