BioNumerics® FEATURED APPLICATIONS

WHOLE GENOME SINGLE NUCLEOTIDE POLYMORPHISM (wgSNP)

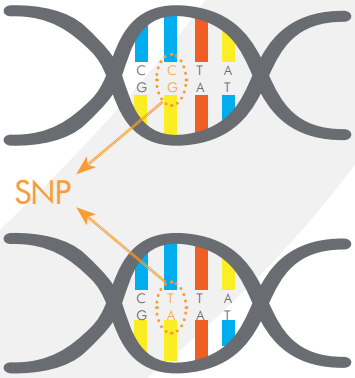
The ultimate resolution in strain typing

What is wgSNP?

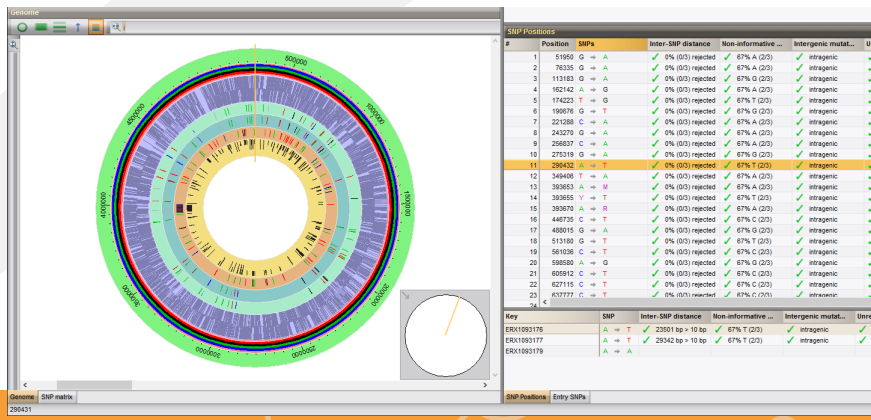
A Single Nucleotide Polymorphism (SNP) is a variation in a single nucleotide, which occurs at a specific position of the genome. SNPs are always defined with respect to a reference sequence. When performed on whole genome sequences (WGS), this analysis is referred as whole genome SNP (wgSNP) analysis. In any SNP search, it is crucial that the reference and sample sequence are aligned, since this is the only way in which base calls per position can be compared meaningfully. wgSNP analysis in BioNumerics relies on all sequences being collinear, i.e. in the same frame and having the same length, which avoids a time-consuming sequence alignment. A SNP search or SNP analysis can be regarded as a post-analysis on (aligned) sequences, in which SNPs are determined on one or more sample sequences, in relation to a reference sequence.

wgSNP analysis in BioNumerics

1. **Automated import from sequence read sets** from various sequencers (Illumina, PacBio, IonTorrent) and sources. Data can be retrieved from NCBI, EMBL-EBI, Illumina Basespace, Amazon S3 or local file servers.
2. **Choice of reference sequence** imported from FASTA or GenBank file, or downloaded directly from online repositories such as NCBI.
3. **Mapping against the reference sequence** using either your local computer or an external calculation engine (e.g. Amazon cloud services).
4. **wgSNP assessment and filtering** determine base differences against the reference sequence and retain only relevant, high-quality SNPs.
5. **Calculate population modelling networks** in the finest and most comprehensive cluster analysis application available today, using standard or custom cluster analysis templates.



NOTE:
The minimal configuration for the wgSNP functionality consists of the "Genome Analysis Tools", the "Sequence Data" and the "Tree and Network Interference" modules. A calculation engine project is only required when reference mappings are performed on the calculation engine.



YOUR ADVANTAGES

WHY USE BioNumerics FOR YOUR wgSNP ANALYSIS?

- ✓ Multiple reference sequences possible
- ✓ Flexible mapping analysis
- ✓ Various SNP filtering templates
- ✓ Quality assessment of retained SNPs
- ✓ Easy to perform wgSNP clustering

wgSNP in BioNumerics

Multiple reference sequences possible

The selection of the reference sequence is key in a wgSNP analysis, as only the genomic information in common between the reference sequence and the sample sequence will be included in the analysis. In other words: any gene, integron, plasmid, etc. that is present in the reference but not in the sample (or vice versa) will be left out. In BioNumerics, different reference sequences - either fully annotated closed genome sequences or draft genomes consisting of multiple contigs - can be defined for multiple wgSNP analyses.

Flexible mapping analysis

The most trivial way to ensure that genomic sequences are collinear for all the isolates under investigation is to map the trimmed sequence reads against the same reference sequence. This can be done locally on your desktop computer or using an external calculation engine such as the pay-per-use Amazon cloud calculation engine which is seamlessly integrated in BioNumerics and provides extremely fast turnaround times.

Various SNP filtering templates

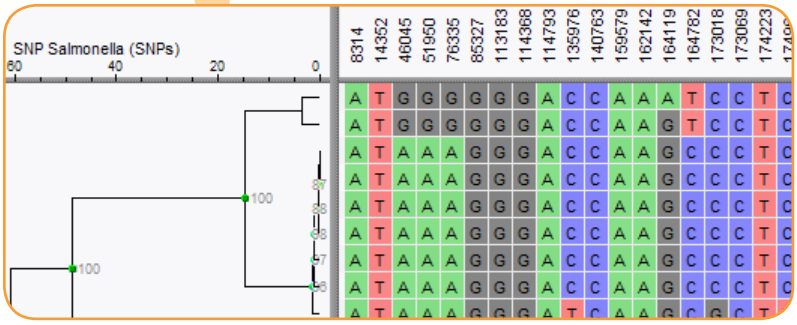
For phylogenetic analyses and strain typing it is very important to retain only the relevant, high-quality SNPs. Therefore, after determination of all base differences, true point mutations need to be distinguished from artefacts due to e.g. sequencing errors or larger indels and rearrangements.

BioNumerics offers this functionality through a number of SNP filters combined into SNP templates, which cover most common use cases. You can also create your own SNP templates, optimized for specific research and/or organisms, and share these between co-workers.

Quality assessment of retained SNPs

An overview of the total and retained number of SNPs for all entries is indicated, as well as the circular genome view up to base level to visually assess the effect of active SNP filters. Detailed quality information on the SNP positions can be accessed and viewed for individual samples. Moreover, underlying sequences and assemblies can be easily accessed.

Easy to perform wgSNP clustering



The SNP matrix, containing those SNPs that are retained by the applied filters, can be further analyzed in the Comparison window.

The rows in the SNP matrix correspond to the samples and the columns represent the positions where at least one SNP is retained after filtering. A wgSNP clustering can be initiated on the SNP matrix using a wide range of clustering options.

CONVINCED? INTRIGUED? TRY IT FOR YOURSELF!

www.bionumerics.com

Scan the QR codes to access each step

1. Make sure you have a **BioNumerics** license (also see note on first page).

2. Request a calculation engine project to perform whole genome analyses.

3. Watch tutorial movies or download sample data for use in **BioNumerics**.

Headquarters
Keistraat 120 • 9830 Sint-Martens-Latem • Belgium
+32 922 22 100 info@applied-maths.com

USA and Canada
11940 Jollyville Rd., Suite 115N • Austin, TX 78750 USA
+1 512 482 9700 info-us@applied-maths.com

APPLIED MATHS
A BIOMÉRIEUX COMPANY