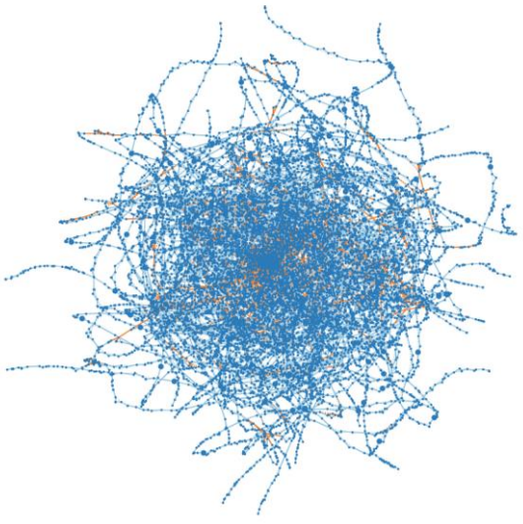
Resource Friendly *de novo* Genome Assembler

Scientists from GIS, Prof Mile Sikic and Postdoctoral Fellow Dr Robert Vaser, published their work on a novel *de novo* genome assembler called Raven in *Nature Computational Science*.

De novo assembly is similar to puzzle solving, but in one dimension. The process starts by finding overlapping DNA fragments, continues with constructing a graph of these overlapping and simplifying it until we have a chain of pieces. The biggest obstacles represent repetitive regions which make the graph tangled.

A sample graph of the human genome is displayed below. Analysing knots in the graph, the authors realised that if they cut the elongated connections (in orange), they can accurately reconstruct a much longer portion of the genome. Therefore, they employed a method based on graph drawings to identify these connections.



The Raven assembler lays the graph in a two-dimensional space to find elongated edges (orange) that connect distant parts of the genome, which are usually present due to sequencing artefacts or repetitive genomic regions.

Compared to the state-of-the-art, Raven maintains similar reconstruction results across different species and genome sizes while being one of the fastest assemblers that uses the least amount of memory on the majority of benchmarked datasets. The small computational requirements make it an excellent option for laboratories with limited funding.

The source code is freely available under the MIT license at https://github.com/lbcb-sci/raven, while precompiled executable can be found in Bioconda and Homebrew. The published paper is accessible under the DOI: 10.1038/s43588-021-00073-4.