



POLISH NATIONAL AGENCY
FOR ACADEMIC EXCHANGE



STER
PROGRAMME

<p>name of the unit:</p> <p>INSTITUTE OF APPLIED COMPUTER SCIENCE, Lodz University of Technology</p>		<p>symbol:</p> <p>I-24 http://www.kis.p.lodz.pl</p>
<p>head of the unit:</p> <p>Szymon Grabowski, PhD, DSc</p>	<p>potential promoters:</p> <p>dr hab. Szymon Grabowski</p>	<p>contact person:</p> <p>tel. (+48) 42 631-27-50 Szymon.Grabowski@p.lodz.pl</p>
<p>scope of activities:</p> <p>Design and implementation of bioinformatics tools, i.a., compressors for popular data types and formats (FASTQ, FASTA, genome collections).</p> <p>Efficient MEM (maximal exact matches) searching in bioinformatics data (pairs of genomes).</p> <p>Efficient algorithms for detecting similar sequences in collections.</p> <p>Efficient <i>k</i>-mer counting for sequencing data.</p> <p>Compact/succinct data structures (e.g., succinct representation for canonical Huffman code, with $O(1)$-time symbol encoding/decoding; an article under review, preprint: https://arxiv.org/abs/2108.05495).</p>		<p>graphic material</p> <p>. M B G C . . G C K V .</p> <p>Multiple Bacteria Genome Compressor</p>
<p>present activities:</p> <p>Development of a multiple bacteria genome compressor, MBGC (T.M.Kowalski, Sz.Grabowski).</p> <p>Development of efficient parallel MEM-finding algorithms (Sz.Grabowski, W.Bieniecki).</p> <p>Design and implementation of an efficient <i>k</i>-mer counting algorithm (Sz.Grabowski, W.Bieniecki, T.M.Kowalski, in collaboration with researchers from Ruđer Bošković Institute, Zagreb, Croatia).</p> <p>Design and implementation of an efficient FASTA compressor (Sz.Grabowski, R.Susik, T.M.Kowalski).</p>		
<p>Future activities:</p> <p>Following the topics being currently developed (see „present activities“).</p> <p>Development of the PgRC compressor of sequencing data (FASTA) (T.M.Kowalski, Szymon Grabowski, 2020), in collaboration with Piotr Duch.</p> <p>Design and implementation of indexing algorithms for repetitive data (i.a., with applications in bioinformatics).</p> <p>Finding similar sequences in collections (of, e.g., genomes) based on sketching.</p>		
<p>Keywords: text algorithms, bioinformatics, data compression</p>		
<p>List of intenship proposal in this research team:</p> <p>1) Compression of repetitive data (in particular, in bioinformatics), with random access.</p>		



POLISH NATIONAL AGENCY
FOR ACADEMIC EXCHANGE



STER
PROGRAMME

- 2) Indexing collections of repetitive data.
- 3) Efficient algorithms for sequence similarity detection.

[List of attachments:](#)

Major publications of the team in 2019-2022:

Szymon Grabowski, Tomasz Marek Kowalski: MBGC: Multiple Bacteria Genome Compressor. GigaScience, DOI:10.1093/gigascience/giab099, 2022 (accepted)

Szymon Grabowski, Tomasz Marek Kowalski: Algorithms for all-pairs Hamming distance based similarity. Softw. Pract. Exp. 51(7): 1580-1590 (2021)

Tomasz Marek Kowalski, Szymon Grabowski: PgRC: pseudogenome-based read compressor. Bioinform. 36(7): 2082-2089 (2020)

Szymon Grabowski, Wojciech Bieniecki: copMEM: finding maximal exact matches via sampling both genomes. Bioinform. 35(4): 677-678 (2019)

Tomasz Marek Kowalski, Szymon Grabowski, Kimmo Fredriksson: Suffix Arrays with a Twist. Comput. Informatics 38(3): 555-574 (2019)

Robert Susik, Szymon Grabowski, Kimmo Fredriksson: Revisiting Multiple Pattern Matching. Comput. Informatics 38(4): 937-962 (2019)