Microbiome bioinformatics: large scale bioinformatics of intercellular signaling in complex bacterial communities

Quorum sensing systems are one of the most well known mechanisms of cell-cell communication and cooperation that play a crucial role in bacterial community formation and complex bacterial behavioral patterns. It allows cells of some bacteria to sense the proximity of each other via the environmental concentration of a QS signal molecule that members of the same species release into the environment. When this concentration is above a certain level, certain cellular functions will be activated in each cell so that the population will behave in a synchronized fashion. This synchronized behavior will then allow a bacterial community to carry out tasks that a single cell cannot, such as virulence, bioluminescence, biofilm formation and many more.

The aim of the research is to conduct a large-scale survey of such quorum sensing systems in the continuously growing body of genomic data, which includes not only the assembled bacterial genomes, but WGS and 16S microbiome sequencing information as well. We developed a bioinformatic tool, qs-finder to automatically detect and cluster the operon structures. There are various reasons for studying these systems: for example, as pathogens use quorum sensing for regulating toxin production, development of possible compounds blocking these communication pathways are promising complementary treatments to antibiotics.

QS systems

Our research group has been developing a bioinformatic pipeline suitable for identifying quorum sensing systems in bacterial genomes. One of the key parts of the pipeline is database search with manually curated Hidden Markov Models (HMM). From many similar sequences of a protein coded by quorum sensing genes of different bacterial species an HMM profile can be built with which proteins resembling the original sequences can be efficiently detected [1]. After identifying the proteins, the corresponding genes can be identified and their subsequent location genomic organization can be analyzed. By examining the neighbourhood of these gene topologies a class label can be assigned. One of the main advantages of this method is that unusual topologies are also identifiable, even those that are not described in literature. Furthermore, the approach is a very flexible process with many adjustable parameters.

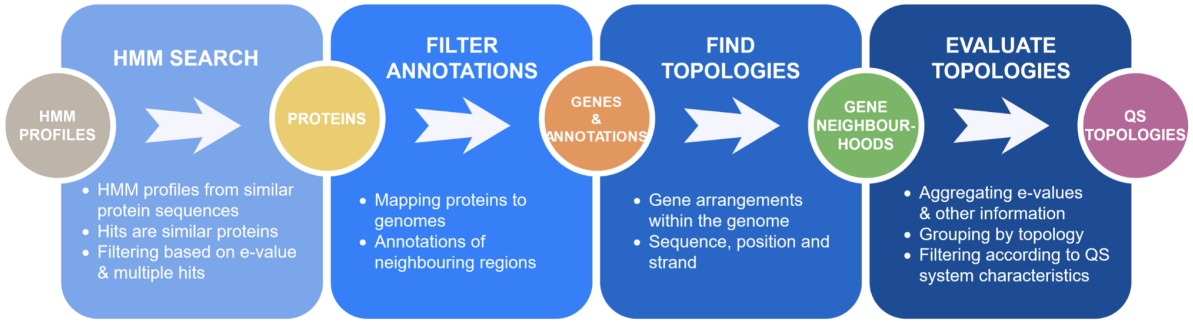

Figure 1. *Simplified schematics of the workflow of the QS-Finder pipeline. The pipeline uses pre-built, manually curated and cleaned databases. Steps are: I. HMMER search, preliminary filtering using various conditions such as hit structure, quality of the similarity, etc, II. mapping the candidate hits to genomic structures. III. Identification and quality assessment of the identified topologies using the spatial and arrangement information. IV. Post-processing and evaluation of candidate topologies using complex aggregation processes, i.e. phylogenetic and other taxonomic information (i.e. gram status, pathogenicity, etc.*

First we surveyed the known systems and marker databases of bacterial quorum sensing systems. The literature was systematically monitored in the grant period. Next, we developed the first prototype of the pipeline and databases and we focused on the validation of well known and previously described systems. We reproduced the analysis of two known bacterial signaling systems (quorum sensing system) [2,3], the AHL (acylhomoserine lactone) system characteristic of Gram-negative bacteria and the ComQPA system mainly found in Gram-positive bacteria, especially Bacillus [4]. The components of the systems are described by various HMM recognizers, HMM profiles. The prototype of the pipeline (shown in Figure 1) was implemented in python, as it is capable of automatically determining the structure of the operon and describing the possible variation and the genomic locations. Based on the literature studies and manual curation we compiled new profile systems for both the AHL and COMX systems as well as a system description strategy that can be extended to other QS systems as well. Preliminary results showed that the number of operons found is 5-10 times higher than the number known, which confirms the relevance of the study (Figure 3). To handle the increased amount of data (genomic and WGS metagenomic), a high-capacity data storage facility was put into operation (financial support was covered by other grants of the research group).

The preliminary analysis also revealed that the identification of such systems contains many false positive hits, therefore we introduced a complex filtering system to avoid such scenarios. In order to and obtained structure to be called a true case of QS signaling, it is supposed that bacteria need three general characteristics as shown in Figure 2: (i) an ability to secrete a signaling molecule, an autoinducer (see below), (ii) an ability to detect the change in concentration of signaling molecules, and (iii) an ability to regulate gene transcription as a response. Minimum systems should have such characteristics which are described by rules. (For more complex classification of QS systems, please see [4]).

topology	pattern freq
R+ || R-	40178
I+ || I-	20461
L+ || L-	11651
I-|R- || R+|I+	5088
I-|M-|R+ || R-|M+|I+	2987
I-|L+|R- || R+|L-|I+	2124
I+|R- || R+|I-	1867
R-|R+ || R-|R+	901
I-|X1|M-|R+ || R-|M+|X1|I+	764
I-|M-|X2|R- || R+|X2|M+|I+	645

Figure 2. *Frequently obtained AHL operon organization structure. Only the sequences classified as bacteria were analyzed.*

Another well studied QS mechanism in Gram positive bacteria is the comQXPA locus in Bacillus subtilis and related bacteria which encodes a QS system consisting of 4 genes. Here the signal is a peptide that is transported across the membrane via active transport. The peptide in the extracellular space is sensed by the extracellular part of a transmembrane receptor ComP. The intracellular part of the receptor is a histidine kinase which will phosphorylate a DNA-binding protein ComA. Once phosphorylated, ComA will bind to the chromosome and upregulate the production of a ComX protein that includes the peptide signal.

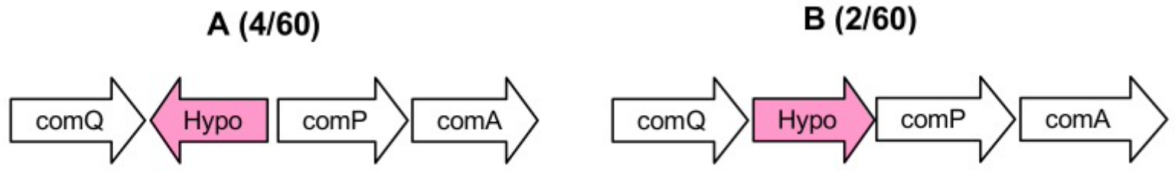

Figure 3*. Typically obtained comQXPA system arrangement. Often, more commonly the shorter genes (here the precursor of the signal molecule) is annotated as hypothetical protein, which likely to be an annotation error.*

In a previous work by Dogsa et. al [4] 60 comQXPA systems were identified with similar methods from which we found all that were possible to. Some contigs were removed from the database and in case of some others, the annotations changed, for example some what was once considered a gene, is annotated now as a pseudogene. In addition to the known ones we found new operon candidates not mentioned in literature before. Due to the exponentially growing sequence databases providing more and more data each year and our improved pipeline becoming more complex, 531 system candidates were identified in total. By examining these putative operons we discovered that 97% of the topologies can be separated into two groups: one of which has an overlap between the comX and comQ genes and those that do not have any overlap between their genes.

The next phase we systematically extended our research to, are less studied QS systems and their general general genomic structures covered in the literature, such as iQS, GBL, etc. The final list of systems are shown in Table 1.

We classified the QS according to the functional steps required for QS signalling which allows one to identify broad mechanism types I-IV. For instance, type I designates QS systems wherein the signal is a small molecule and both inbound and outbound traffic is carried out by diffusion. Type II includes QS systems wherein the small molecular signal

leaves the cell diffusion, but it does not reenter the cell, rather it binds to a transmembrane receptor which triggers a cascade that will ultimately activate the both signal production and the downstream target genes. In the type III. system the signal molecules are peptide and inward signal transport carried out by active transport, meanwhile in systems of type IV the peptide signals bind to membrane receptors. For more detailed description of the classification please see [4].

Table 1: Examples for different quorum sensing systems. “Gene1” → “gene2” stands for “receptor gene” → “response regulator gene” in the fifth column. More details are found in the text.text.

Type	System	Signal molecule	Synthesis gene	Receptor and regulatory gene	Typical species
I.	AHL (N-AHL, AI-1)	acyl-homoserine lactones	luxI; lasI	luxR; lasR	Widespread system, e.g.: *Vibrio fischeri; Pseudomonas aeruginosa*,
I.	Quinolone signalling	quinolones (AQS, HHQ, PQS, HMAQs)	phnA,B. pqsA,B,C,D,E, (pqsH,L); trpE,G, hmqA,B,C,D,E,F ,G	pqsR (mvfR)	*Pseudomonas aeruginosa; Burkholderia spp.*
I.	IQS	2-(2-hydroxyphenyl)-thiazole-4-carbaldehyde	ambB,C,D,E	iqsR (unknown)	*Pseudomonas aeruginosa*
I.	phc QS	3-OH-PAME, 3-OH-MAME	phcB	phcS → phcR	*Ralstonia solanacearum*
I.	Photopyrone QS (PPYs)	alpha-pyrones	ppyS	pluR	*Photorhabdus luminescens*
I.	Dialkylresorcinols (DARs), cyclohexanediones (CHDs)	Dialkylresorcinols (DARs), cyclohexanediones (CHDs)	darA,B,C	pauR	*Photorhabdus asymbiotica*
I.	GBLs (gamma-butyrolactones, A-factor)	CHB	scgA,X	scgR	*Streptomyces chattanoogensis* Gram-positive!

Type	System	Signal molecule	Synthesis gene	Receptor and regulatory gene	Typical species
II.	alpha-hydroxyketones (AHK) QS	CAI-1 ((Z)-3-aminoundec-2-en-4-one) LAI-2	cqsA; lqsA	cqsS → … → luxR; lqsS, lqsT → lqsR	*Vibrionaceae* (*Vibrio harveyi*); *Legionella pneumophila*
II.	HAI-1 QS	N-(3-hydroxybutyryl)-HSL	luxM	luxN → … → luxR	*Vibrio harveyi*
II.	Diffusible signal factor: (B/C/I)DSF	cis-2-unsaturated fatty acids	rpfF	rpfC(,H) → rpfG → clp	*Burkholderia spp.; Xantomonas citri; Pseudomonas aeruginosa*
II.	AI-3	epinephrine, norepinephrin-like molecule	(unknown)	qseC → qseB; qseE → qseF	*Escherichia coli*
II. III.	AI-2 (autoinducer 2, QS-2, DPD)	furanosyl-borate diesters	luxS	luxP → luxQ → … → luxR; lsrA,B,C,D → lsrR	Very frequent in both Gram-positives and Gram-negatives, e.g.: *Vibrio harveyi*; *Escherichia coli*

Type	System	Signal molecule	Synthesis gene	Receptor and regulatory gene	Typical species
III. Peptide signal	Short hydrophobic peptide (SHP) signalling	SHP, CIP, XIP	hsp2,3; comS	rgg2,3; comR (rgg4)	*Streptococcus spp.,Lactobacillus spp., Listeria spp.*
III. Peptide signal	RNPP family	CSF; PapR-AIP; NprX/NprRB-AIP; iCF10, cCF12; PhrA-AIP	phrC; nprX; papR; ccfA, prgQ; phrA	rapB,C; nprR; plcR; prgX; trpA	*Bacillus subtilis, B. cereus, B. thuringiensis, Enterococcus faecalis; Streptococcus pneumoniae D39*
IV. Peptide signal	ComQXPA	ComX	comX	comP → comA	*Bacillus subtilis*
IV. Peptide signal	Arg-type cyclic pheromones	AIP (I-IV); GBAP; Lam-AIP	argD; frsD; lamD	argC → argA; frsC → frsA; lamC → lamA, lamK → lamR	*Streptococcus aureus; Enterococcus faecalis; Lactobacillus plantarum*
IV. Peptide signal	Gly-Gly peptides (class II. bacteriocine AMP related)	e.g.: CSP; Blp signal; Sil signal	comC; blpC; silCR	comD → comE; blpH → blpR; silD → silA	e.g.: *Streptococcus pneumonia; Streptococcus pyogenes*
IV. Peptide signal	lantibiotics (class I. small AMPs)	e.g.: nisin	nisA	nisK → nisR	e.g.: *Lactobacillus lactis,*

In the review process we compiled and created manually curated HMM-recognizer. The reviewing process has been continued during the grant period. The final database covered 17 QS systems as shown in Table 1. Another characteristic feature is the modular architecture of QS proteins which has an impact on how such proteins and their genes can be identified in bacterial genomes. As an example, the sensor-regulator protein R of AHL systems used in Gram-negative bacteria consists of two domains, a DNA-binding domain DB and a signal-binding (autoinducer binding) domain AIB.

The DNA-binding domain, which contains a helix-turn-helix (HTH) DNA binding motif, is a very common structural motif. The autoinducer-binding domain is specific for AHL systems. When searching for R homologs we find a large variety of hits to various helix-turn-helix proteins, with only a few of which also regions similar to AIB sequences. Generally the

common domains (i.e. response regulator, histidine-kinase, ABC transporters) are associated with specific QS 'tasks' (i.e. sensing signals, import-export of the signal, etc) and these proteins tend to cluster together as shown in Figure 4.

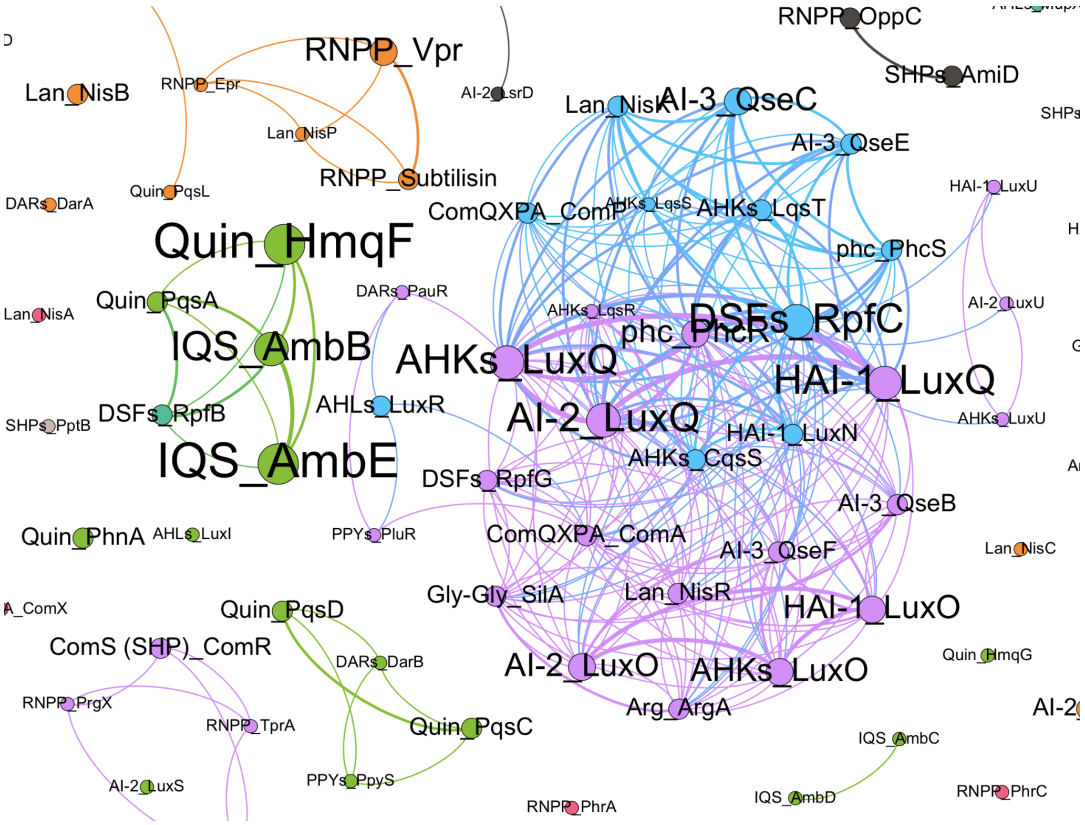

Figure 4. *Domain structure in various QS systems. The protein is connected if they share domains.*

The advent of next generation sequencing, the number of known bacterial genomes is dramatically increasing (Figure 5). The exponential growth of genomic data requires both computational and algorithmic solutions that can keep up with the amount of data the pipelines must handle.

There are several relatively easy-to-use methods for demandings tasks that require High Performance Computing, ie. identifying communication genes from metagenomic NGS result. I. using hardware based algorithmic acceleration i.e. FPGA, multi-core CPU, GPUs, II. using horizontal scaling techniques on computer clusters (Apache Spark, Dask, etc).

Dedicated multi-core architectures such as FPGA, GPU, Xeon-Phi showed limited performance enhancement in our applications, despite the initial positive results. The hashing performance in k-mer applications (FPGA, GPU) were significantly improved compared to CPUs, however in real-world sequencing data (especially in aggregation of k-mer matches subtask), such advantages were diminished.

Recently many generic frameworks and technologies have been proposed for large-scale computational tasks. Such frameworks should solve various problems such as load balancing, cluster management of software, algorithms, and requires state-of-the-art hardware and high bandwidth with low latency network devices, which might not be widely accessible for research groups.

Wa wanted to design the bioinformatics pipelines in a way that is easy-to-deploy and can efficiently work in heterogeneous infrastructure as well. I.e. both on a local server or cloud based environments. We investigated various solutions to make the pipeline scalable and easy-to-deploy, namely, Apache Spark, python-Dask, Kubernetes along with Docke, CWL (Common Workflow Language), serverless solutions such (i.e. Azure function).

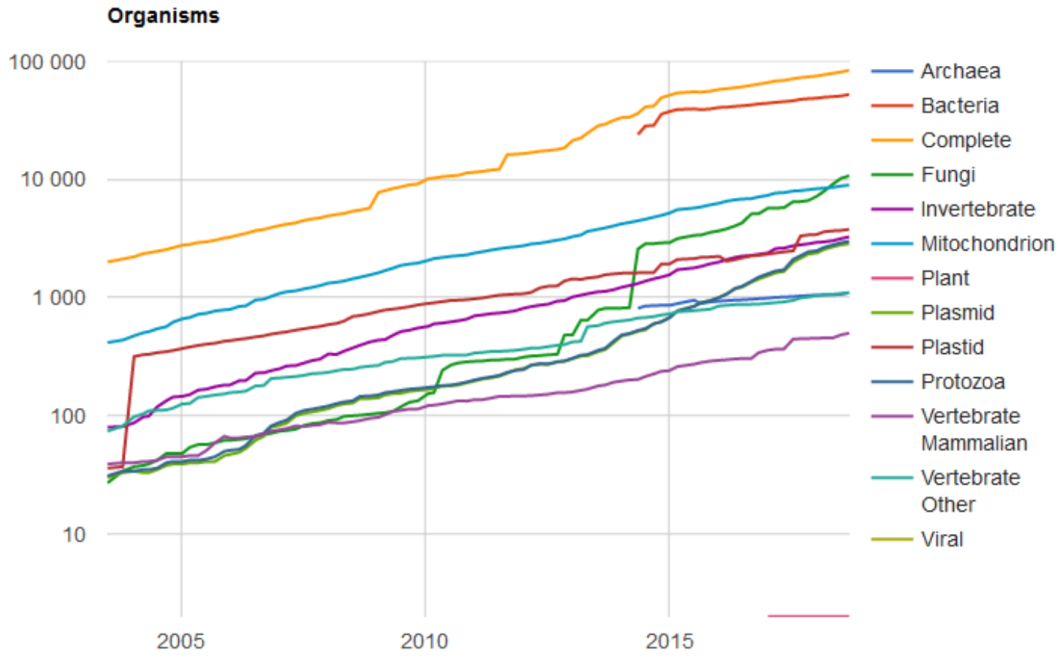

Figure 5. *Exponential increase in the availability of genomic data in Refseq database.*

Common Workflow Language project created for data pipeline execution and look for an already working example using this project. CWL could be an interesting and useful way to create a bioinformatics pipeline, so it would be useful to study how it works and see if it is compatible with other potentially useful technologies like Docker containers.

CWL combining with docker proved to be a portable solution, however the QS-Finder pipeline and marker-search pipelines heavily use large databases, which rendered the deployment process way too slow (i.e. 40 minutes to 3 hours) to be a practical application (loading time is significantly larger, then computational time), however some cloud optimization were performed (i.e. using blob storage, etc).

Next, we continued the benchmarking with other horizontal scaling techniques such as Apache Spark. Apache is a very fast and general engine for large-scale data processing. Spark can run in several places, e.g. Both on a Hadoop cluster and on a local PC itself.

Next, we set up a benchmarking task based on metagenomic samples. The simplified pipeline used in the process is shown on Figure 6. Here we used the optimized python scripts with 23 CPU cores most of the time as reference time. The SAM file which was the result of the alignment of the reads with overlap was about 2.2 TB. After parsing and reducing this file with the sequence-db module we got a TSV file with the size of 681 GB. The computational process was reduced from 300sec to 50 using 4 VM (16 core and 64GB RAM), we concluded that horizontal scaling can be applied efficiently in our applications, as well as cpu based hardware acceleration.

Later, we benchmarked the python DASK framework, which can be easily integrated both into the QS-Finder and marker-search pipelines. The Dask showed a similar performance enhancement pattern in our application, especially in aggregation and filtering steps, compared to Apache Spark.

Furthermore, we also tested the Azure serverless solutions in the benchmarking task (Figure 6). Serverless computing enables developers to build applications faster by eliminating the need for them to manage infrastructure. With serverless applications, the cloud service provider automatically provisions, scales, and manages the infrastructure required to run the code. The name *serverless* comes from the fact that developers do not have to take attention to the underlying infrastructure. With Azure Functions, developers can run so-called functions that are small pieces of codes without worrying about application infrastructure.

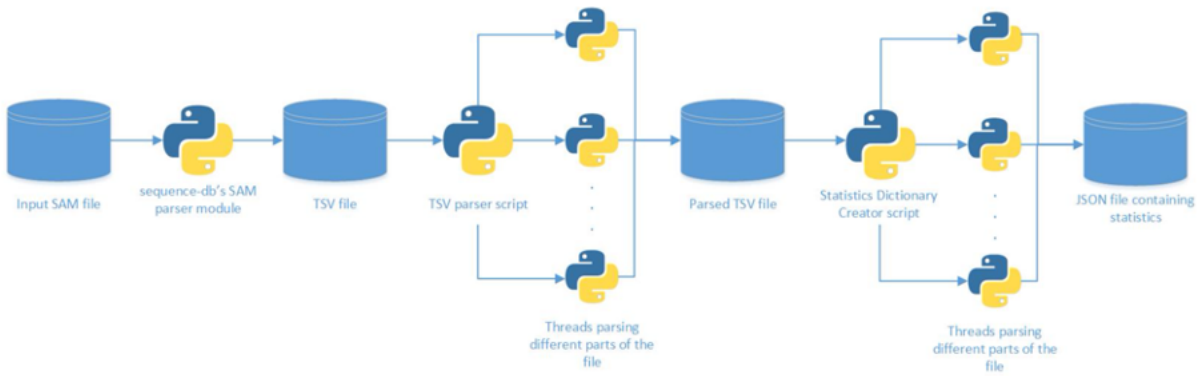

Figure 6. *Benchmarking process, implemented in the microsoft Azure cloud system.*

Followed by the investigation of possible ways of improving the scalability of the pipelines we continued the development of an automated/semiautomated system for mining quorum-sensing (QS) related operons from genomic data. We developed a general framework using a successive search reduction strategy, based on a hierarchical use of presence/absence, composition and structural descriptions for operons as well as a search for gene doublets and gene triplets. With this strategy we were able to analyze various operon types and proceeded towards the automatization of the system.

We also carried out a survey of bacterial archaeal viral and plasmid genomes in order to identify characteristic quorum sensing gene topologies and characteristic chromosomal neighborhoods. We continued and redesigned our warehouse Figure 7, to be able to support various analysis and machine learning tasks (i.e. clustering) of QS Systems.

We continued the optimization of the qs-finder bioinformatics pipeline in order to make it capable of analyzing the comprehensive refseq database and a large number of metagenomic samples. We screened the refseq database (1.14T bp, 1.95m contigs, 73 000 taxa). Also improved and expanded the data warehouse with bacterial annotations, such as gram status to support the subsequent data analysis and visualization tasks.

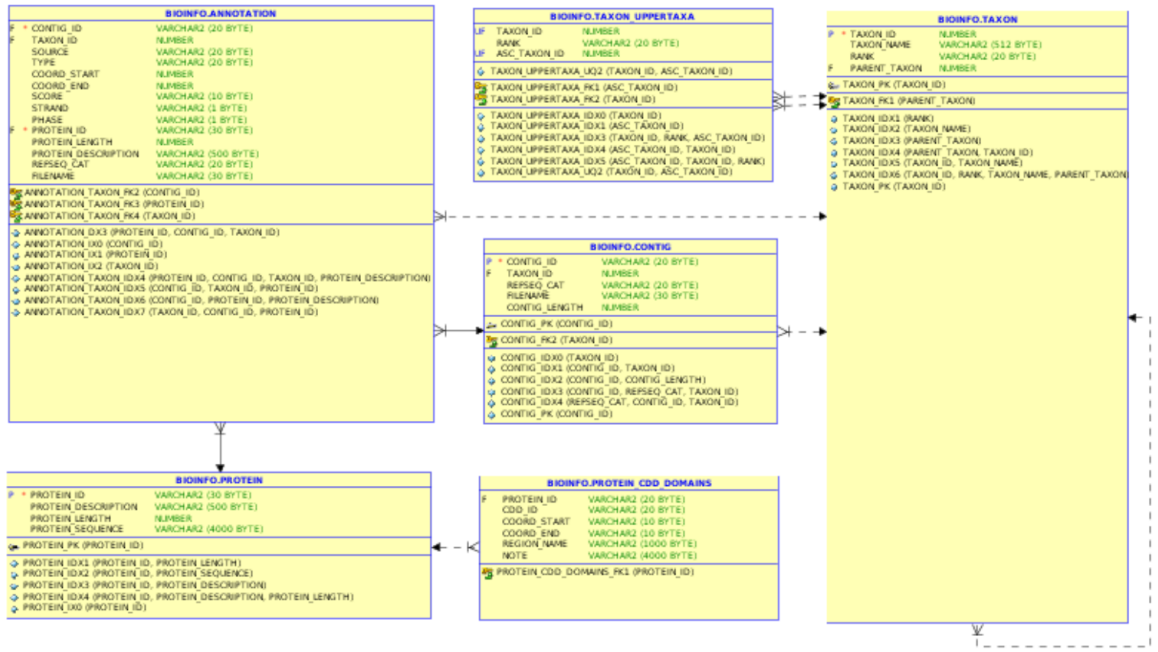

Figure 7. *Sample of the ER-schema of the developed for QS-Survey and QS-Finder (only partial).*

Next we continued the development of the analysis of metagenomics samples and focused on various applications, such as efficient detection of markers, or marker like structures, i.e. bacteria phages. We extended the pipeline and developed an integration process (Figure 8) in a way that is not only capable of characterizing the taxonomical content of a wgs sample, but recognizes certain genomics parts.

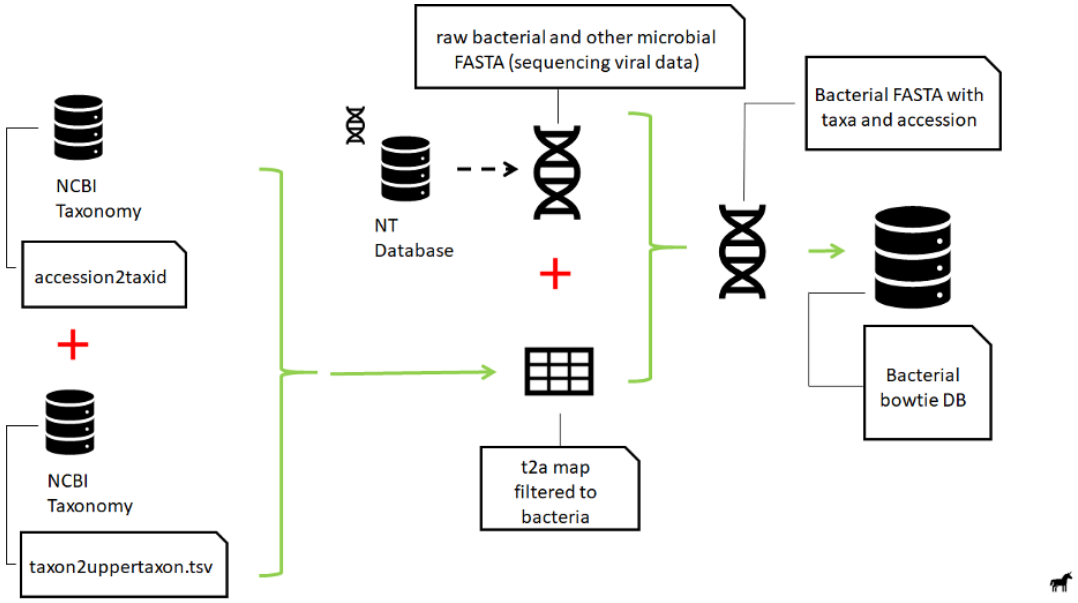

Figure 8. *Bacterial marker database creation process. The input databases are known genomic and taxonomy databases which mighte BLAST or in Refseq format, both sequence representation strategies are supported. The taxonomy can be various format i.e. SILVA for supporting 16S analysis. The output indexed and filtereped K-mer assembled contigs indexed by BWT based methods such bowtie2.*

In this period we also investigated the possible usage of serverless technologies offered by Microsoft Azure by compiling an alternative 16S sequence database to SILVA and systematically benchmarking and defining the possible resulational limitation of using the k-mer based marker database in amplicon sequencing projects. The outcome of the analysis helped us in better designing of 16S microbiome studies, i.e. in the collaboration with Dóra Szabó better understanding the dynamics of the bacterial community in the presence of multidrug resistant Klebsiella pneumoniae and various antimicrobial agents.

Quorum sensing systems are one of the most well known mechanisms of cell-cell communication and cooperation that play a crucial role in bacterial community formation and complex bacterial behavioral patterns. The aim of the research project was to systematically survey and describe these quorum sensing systems and develop novel bioinformatics tools. As part of this, we created a tool (QS-finder) for detecting and describing QS systems using manually created HMM profile systems. We set up a classification system.

QS-finder automatically detects and clusters these systems in metagenomic samples. The tool is available under URL https://pongor2.itk.ppke.hu/qs-finder.

The pipeline is capable of scaling with the continuously increasing size of genomic data by taking advantage of modern multicore architectures, but using the horizontal scaling techniques. We create a dedicated database for 17 QS systems, which are also publicly available https://pongor2.itk.ppke.hu/qs-survey).

References

[1] S. T. Rutherford and B. L. Bassler, “Bacterial Quorum Sensing: Its Role in Virulence and Possibilities for Its Control,” Cold Spring Harb. Perspect. Med., vol. 2, no. 11, pp. a012427–a012427, Nov. 2012.

[2] K. Sonal Choudhary, S. Hudaiberdiev, Zs. Gelencsér, B. Gonçalves Coutinho, V. Venturi and S. Pongor, “The Organization of the Quorum Sensing luxI/R Family Genes in Burkholderia” INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES 14: 7 pp. 13727-13747., 21 p.

[3] Zs. Gelencsér, B. Galbáts, J. F. Gonzalez, K. Sonal Choudhary, S. Hudaiberdiev, V. Venturi and S. Pongor, (2012) “Chromosomal Arrangement of AHL-Driven QuorumSensing Circuits in Pseudomonas” INTERNATIONAL SCHOLARLY RESEARCH NETWORK MICROBIOLOGY 2012 Paper: doi:10.5402/2012/484176 , 6 p.

[4] J. Juhász, S. Pongor and B. Ligeti, “Toward a Systematic Genomic Survey of Bacterial Quorum Sensing Genes: Cross Cutting Regulatory and Genomic Concepts“. In: V. Ravishankar Rai and A. B. Jamuna (eds) Trends in Quorum Sensing and Quorum Quenching: New Perspectives and Applications. (2020)

OTKA 120650 Publications

1. Ádám Nagy, Balázs Ligeti, János Szebeni, Sándor Pongor, Balázs Győrffy. COVIDOUTCOME – Estimating COVID Severity Based on Mutation Signatures in the SARS-CoV-2 Genome Database : the journal of biological databases and curation (**2021**) In press.

2. Barbara Hutka, Bernadette Lázár, András S. Tóth, Bence Ágg1, Szilvia B. László, Nóra Makra, Balázs Ligeti, Bálint Scheich, Kornél Király, Mahmoud Al-Khrasani, Dora Szabo, Péter Ferdinandy, Klara Gyires and Zoltán S. Zádori. The nonsteroidal anti-inflammatory drug ketorolac alters the small intestinal microbiota and bile acids without inducing intestinal damage or delaying peristalsis in the rat. Frontiers in Pharmacology (**2021**) In press (doi: 10.3389/fphar.2021.664177)

3. Szabo Helga, Hernyes Anita, Piroska Marton, Ligeti Balazs, Fussy Peter, Zoldi Luca, Galyasz Szonja, Makra Nora, Szabo Dora, Tarnoki Adam Domonkos, Tarnoki David Laszlo Association between Gut Microbial Diversity and Carotid Intima-Media Thickness MEDICINA 57 : 3 Paper: 195 , 11 p. (**2021**) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31974190

4. Moreno-Indias Isabel, Lahti Leo, Nedyalkova Miroslava, Elbere Ilze, Roshchupkin Gennady, Adilovic Muhamed, Aydemir Onder, Bakir-Gungor Burcu, Santa Pau Enrique Carrillo-de, D'Elia Domenica, Desai Mahesh S., Falquet Laurent, Gundogdu Aycan, Hron Karel, Klammsteiner Thomas, Lopes Marta B., Marcos-Zambrano Laura Judith, Marques Cláudia, Mason Michael, May Patrick, Pašić Lejla, Pio Gianvito, Pongor Sándor, Promponas Vasilis J., Przymus Piotr, Saez-Rodriguez Julio, Sampri Alexia, Shigdel Rajesh, Stres Blaz, Suharoschi Ramona, Truu Jaak, Truică Ciprian-Octavian, Vilne Baiba, Vlachakis Dimitrios, Yilmaz Ercument, Zeller Georg, Zomer Aldert L., Gómez-Cabrero David, Claesson Marcus J. Statistical and Machine Learning Techniques in Human Microbiome Studies: Contemporary Challenges and Solutions FRONTIERS IN MICROBIOLOGY 12 p. 1 (**2021**) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31967038

5. Nagy Ádám, Pongor Sándor, Győrffy Balázs Different mutations in SARS-CoV-2 associate with severe and mild outcome. INTERNATIONAL JOURNAL OF ANTIMICROBIAL AGENTS 57 : 2 Paper: 106272 , 5 p. (**2021**) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31820533

6. Juhász János, Ligeti Balázs, Gajdács Mário, Makra Nóra, Ostorházi Eszter, Farkas Ferenc Balázs, Stercz Balázs, Tóth Ákos, Domokos Judit, Pongor Sándor, Szabó Dóra Colonization Dynamics of Multidrug-Resistant Klebsiella pneumoniae Are Dictated by Microbiota-Cluster Group Behavior over Individual Antibiotic Susceptibility: A Metataxonomic Analysis; ANTIBIOTICS 10 : 3 Paper: 268 , 21 p. (**2021**) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31903338

7. Stercz Balázs, Farkas Ferenc Balázs, Tóth Ákos, Gajdács Mário, Domokos Judit, Horváth Viola, Ostorházi Eszter, Makra Nóra, Kocsis Béla, Juhász János, Ligeti Balázs, Pongor Sándor, Szabó Dóra The influence of antibiotics on transitory resistome during gut colonization with CTX-M-15 and OXA-162 producing Klebsiella pneumoniae ST15; SCIENTIFIC REPORTS 11 Paper: 6335 , 10 p. (**2021**) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31921035

8. Kalcsevszki Regina, Juhász János, Pongor Sándor, Ligeti Balázs Biomarkers for Colorectal Carcinoma Prediction from Compositional Microbiome Data (Poster) (**2020**) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31652847

9. Juhász János, Sándor Pongor, Balázs Ligeti Toward a Systematic Genomic Survey of Bacterial Quorum Sensing Genes: Cross Cutting Regulatory and Genomic Concepts In: V., Ravishankar Rai; Jamuna, A Bai (szerk.) Trends in Quorum Sensing and Quorum Quenching: New Perspectives and Applications

Boca Raton (FL), Amerikai Egyesült Államok : CRC Press - Taylor and Francis Group, p. 31, (**2020**) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31314580

10. Juhász János, Gáspári Zoltán, Pongor Sándor Structure and Oxidative Folding of AAI, the Major Alfa-Amylase Inhibitor From Amaranth Seeds FRONTIERS IN CHEMISTRY 8 p. 180 , 7 p. (**2020**) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31251836

11. Juhász János Mikrobiális közösségek koordinációjának vizsgálata ágens-alapú modellekkel Pongor Sándor Disszertáció benyújtásának éve: 2018, Védés éve: 2019 Megjelenés/Fokozatszerzés éve: **2019** https://m2.mtmt.hu/gui2/?mode=browse&params=publication;30762813

12. János Juhász, Balázs Ligeti, Sándor Pongor Modelling genomic and metabolic responses of microbial communities to antibiotics JEDLIK LABORATORIES REPORTS VII : 1 pp. 77-78. , 2 p. (**2019**) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;30762852

13. Balázs Ligeti, Hőnich Ágnes Kata, János Juhász, Sándor Pongor Surveying bacteriophages of the human gut JEDLIK LABORATORIES REPORTS VII : 1 p. 79 (**2019**) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;30762859

14. Ligeti Balázs, Menyhárt Otilia, Juhász János, Pongor Sándor, Győrffy Balázs; Network-based prediction of drug-combination efficiency based on expression data on off target genes JEDLIK LABORATORIES REPORTS 7 : 1 pp. 75-76. , 2 p. (**2019**) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;30762840

15. János Juhász, Éva Mercédesz Kistóth, Zoltán Nagy, Péter Szolgay, Sándor Pongor, Balázs Ligeti Towards a systematic census of bacterial quorum sensing genes in public databases JEDLIK LABORATORIES REPORTS VII : 1 pp. 72-74. , 3 p. (**2019**) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;30762833

16. Balázs Ligeti, Éva M. Kistóth, Regina Kalcsevszki, Bálint. S. Stellek, János Juhász, Sándor Pongor Fast and sensitive detection of pathogens from high throughput sequencing data In: Porgramme and Books of abstracts - Hungarian Molecular Life Sciences 2019 Budapest, Magyarország, (**2019**) pp. 91-92. , 2 p. https://m2.mtmt.hu/gui2/?mode=browse&params=publication;30762806

17. Juhász János, Bihary Dóra, Jády Attila, Pongor Sándor, Ligeti Balázs Differential signal sensitivities can contribute to the stability of multispecies bacterial communities. BIOLOGY DIRECT 12 : 1 p. 22 (**2017**) https://m2.mtmt.hu/gui2/?mode=browse&params=publication;3271393