

# UPDATE, Spring 2017

Version 01, June 2017

#### **COMPARE Quick Facts**



Coordinator, Frank Aarestrup

Technical University of Denmark



Co-Coordinator, Marion Koopmans Erasmus Universitair Medisch Centrum Rotterdam

Project period

01 December 2014 – 30 November 2019

Funding

€20 million (approximately)

 $\square$ 

Contact

<u>compare@food.dtu.dk</u> 🛜 <u>www.compare-europe.eu 🔰</u>@CompareEurope



This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 643476.



#### Contents

Note from Coordinators	2
WP1 – Risk assessment and risk-based strategies for sample and data collection	4
WP2 – Harmonised standards for sample processing and sequencing	5
WP3/6 – Frontline diagnostics	6
WP4/7 – Foodborne pathogen surveillance, outbreak detection and epidemiological analysis	7
WP5/8 – Detection and response to (re-) emerging diseases	9
WP9 – COMPARE platform	10
WP10 – Risk communication tools	12
WP11 – User consultations	13
WP12 – Barriers to open data sharing	15
WP13 – Dissemination and Training	17
WP14 – Cost-effectiveness framework	18
WP15 – Management	19



#### Note from Coordinators

As COMPARE hits the half-way mark (2.5 years of a 5-year project), it is a time to start sharing results but also keep an eye on the horizon, because the landscape has changed since COMPARE started.

Infectious disease in the current era still holds some basic truths:

- Dynamics of common infectious diseases are changing
  - Demographic change, population density, anti-vaccine, AMR, etc.
- New diseases emerge frequently
  - Deforestation, population growth, health system inequalities, travel, trade, climate change
- Effects are difficult to predict due to complexity of problems
  - Rapid flexible response
- Public health and clinical response depend on global capacity for disease surveillance
  - Rapid sharing, comparison and analysis of data from multiple sources and using multiple methodologies

Therefore, COMPARE must be a system with tools that enables collaborative preparedness and outbreak research. The system must be sector-, domain- and pathogen-independent (One Health); the tools are flexible, scalable and open-source based; the platform is easy-to-use for sharing data and information; and it must be built from a sustainable infrastructure.

Looking at our achievements thus far, COMPARE has developed uniform protocols and workflows for next generation sequencing detection for a range of viruses, bacteria and parasites in a multitude of sampling types, and we have matured customized analytical workflows for foodborne diseases, emerging diseases and clinical samples.

In addition, a number of common pilot studies have been initiated:

- tracking of avian influenza outbreak in wildlife
- support of outbreaks (Ebola, H5N8 AIV, norovirus, AMR-colistine, Zika)
- metagenomic analysis of a global urban sewage snapshot as experimental surveillance tool

Datahubs that facilitate the upload, sharing and download of sequences and metadata have been launched in support of the pilot projects. These datahubs can also be linked with (automatic) analysis tools that, at a minimum, can provide a quality control of sequence data (in comparison with others). The development and linking of automatic visualization tools is another area that can only increase the impact of the data being shared.

All pilot projects must also take into consideration barriers to data sharing. And in relation to this, the initial analysis of the impact of the Nagoya Protocol has been made. The work COMPARE partners have made in this area is being shared with other European projects, such as EVAg, ZikaAlliance and Virogenesis.



We have noted that since the start of COMPARE, the landscape has changed a bit; still more institutes and research organizations are more familiar with WGS and their needs for data to compare and analyze have also increased. There are also advances in solutions to fast data upload or continuous data upload even if internet or power interruptions occur. We are also flexible and take into consideration the changing needs of supra-international organizations that are more and more embracing the opportunities of sharing and analyzing sequencing data along with a minimal set of metadata.

The second half of COMPARE looks to be just as interesting with the expectation of new research developments as well as improved IT infrastructures and systems.





### WP1 – Risk assessment and risk-based strategies for sample and data collection

There has been good progress within WP1 in the time period up to the end of the second project year and start of year 3, and we are beginning to see the fruits of our labour. We have seen the completion of our initial conceptual frameworks and literature reviews/inventories completed.

As evidence of this, we have successfully submitted an inventory of transmission modelling frameworks, and updated our sampling protocol database for collection, devices, media, shipment and storage conditions for different scenarios, hosts and pathogens. We hope to make these epi datasets available more widely through a user-interface. Our first case studies across the WP have been selected in association with other work packages, data requested and collated. We are now parameterizing the model frameworks and tools for specific pathogen examples.

We have initially simulated results for the estimated spatial risk of incursion of Lumpy skin disease with maps produced at different levels of geographical detail. For our developing microbial risk assessments we are using Listeria monocytogenes whole-genome sequence data to update our framework using machine learning techniques, and there has been further development of the phylogenetic modelling with swine flu.

All this research is dependent on the collaboration between specialists in diagnostics, pathogens, bioinformaticians and mathematical modelling techniques – a real strength of COMPARE bringing such expertise and relevant institutions together across the One Health domains.



Figure 1. WP 1 members at the face-to-face meeting in the UK.



## WP2 – Harmonised standards for sample processing and sequencing

Careful sample handling is a crucial step in gaining high-quality information from next generation sequencing to ensure maximum benefit for clinical and public health.

Within WP2, several metagenomic studies were conducted to detect new viruses or variants of well-known viruses, and to investigate their evolution, spatiotemporal patterns and zoonotic potential for instance.

A screening of swab samples of squirrels for VSBV-1 (Variegated Squirrel Bornavirus 1) demonstrated a prevalence of 3.3% and 8.8%, respectively, for two different squirrel species, highlighting the existing transmission risk to humans (Schlottau et al. 2017). In addition, avian influenza was in the focus regarding new outbreaks and analysis of worldwide distribution.

Concerning WP2 tasks, sequencing of main experiments for sample handling and storage is currently under analysis within task 1. In addition, pre-tests for urine samples with Zika virus were completed. Related to task 2, different matrices were intensively tested for optimization of sample disintegration and processing. Then, samples with different pathogen types were investigated also using the optimized workflows. Furthermore, sequencing technologies are continuously tested within task 3. Regarding task 4, software tools were examined for the detection of insertions and deletions (indels) and the investigation of parasites. Within task 7, the GMI Virus Proficiency Test started in autumn 2016 with part 1 (dry lab), which was finished and summarized at the General Meeting (March 2017). Part 2 (wet lab) is still ongoing. The next ring trial on food metagenomics is already in its planning stage.



Figure 2. Deliverables 2.2 ("Standard protocol sample processing") and 2.4 ("Data analysis pipeline") have been completed and submitted in time (30 November 2016). Both are currently under review and will be deposited on the COMPARE Share Site



### WP3/6 - Frontline diagnostics

As planned during the WP 3/6 meeting in The Netherlands in 2016, WP3/6 has been organized in four different work bundles. As part of Work Bundle 1 ("Pathogen typing"), the LMM-UAntwerpen WGS pipeline for clinical diagnostics (and for non-bioinformaticians) has been developed and Deliverable 3.1 ("Analytical workflow for clinical diagnostic application") is ready to be submitted. Next steps are the addition of new modules that are important in a clinical setting and the validation in hospital-based microbiology labs (Deliverable 6.1).



Figure 3. Screenshot from the LMM pipeline for clinical diagnostics.

Regarding Work Bundle 2 ("AMR prediction from WGS data"), Deliverable 3.2 ("Prediction algorithm for antimicrobial resistance markers in sequence data") is ready to be submitted and an online database to submit WGS data and AMR metadata is available for COMPARE partners. The next step is to gather data and metadata



for 1000 *E. coli* genomes with matched MIC data for at least, but not limited to ciprofloxacin, and then go on with the development of different machine learning approaches by the designated COMPARE teams.

For Work Bundle 3 ("In-hospital transmission"), whole genome MLST(wgMLST) and core genome(cgMLST) analysis using both Bionumerics v7 and BigsDB (https://pubmlst.org/) schemes for *S. aureus* has been done to index allele differences in shared genomic loci, and compare both schemes in order to address their discriminatory power, ability of phylogenetic inference and tree congruency. The next step is to study the population dynamics within the different households in order to address the detection of putative transmission events among the populations.

Two metagenomic studies are currently ongoing for Work Bundle 4 ("Diagnostic metagenomics for pathogen detection"). One is a shot-gun metagenome study of feces from swine with diarrhea; the causative agent, Porcine epidemic diarrhea virus (PEDV), was identified and metadiagnostic analysis via FLI RIEMS pipeline revealed co-infection with bacteria. A pilot study for the prediction of antibiotic resistance markers will be done to test pipelines and algorithms with clinical metagenomics data. The second study is a pilot study regarding the analysis of UTI samples with metagenomic sequencing and comparison with culture results. Relative abundance was not enough to discriminate between positive/negative samples since some samples with negative cultures showed high relative abundance of pathogens. However, the combination of DNA quantity per sample with relative abundance correlated well with culture results. The next step is the application of this technique to a larger set of samples.

## WP4/7 – Foodborne pathogen surveillance, outbreak detection and epidemiological analysis

During this period, we have finalized the creation of the datahub Vivaldi (bacteria) and Puccini (parasites) in collaboration with WP9. The working group 4/7 has focused on sequencing and adding data to the datahub Vivaldi which has >1000 bacterial genomes uploaded at present (March 2017). Analysis of these data has been initiated for several studies: Source attribution studies on WGS data, initially focusing on Salmonella Typhimurium and its monophasic variant; long term evolutionary studies of different bacterial clones to describe persistence and spread; and analysis of several historical datasets with known outbreaks from different bacterial organisms in order to evaluate the utility and suitability of different cluster detection methods for whole genome outputs.

The outcome of these analyses will be recommendations on which cluster detection methods are optimal for which organisms. Also, the working group will make suggestions for the final platform in terms of which functionalities we would like for epidemiological analysis and source attribution prediction.





Figure 4. A tanglegram illustrating the coherence between different analysis approaches for WGS data. Depicted is a SNP tree compared with a cgMLST tree. Further analyses of other approaches are ongoing for a final recommendation of which analysis tools to include in COMPARE platform.

The research specifically related to virological methods has mainly been done as cross-WP activities. Of relevance for the foodborne working group, is the development of methods for detection of norovirus in food samples such as oysters. Because viral contamination is very low, a protocol was designed to concentrate viral particles based on elution with Ph modification, filtration to eliminate bacteria and then a final concentration step. This protocol allowed for us to obtain full length genome of norovirus in some bio-accumulated oysters.

A random (meaning non-targeted) viral metagenomics protocol for clinical serum and feces has been developed and assessed. This protocol was used to identify a gastro-enteritis causing virus that our diagnostic qPCR assay failed to identify; this false-negative was caused by mutations in the primer binding site.



#### WP5/8 – Detection and response to (re-) emerging diseases

A paper by the Global Consortium for H5N8 and Related Viruses (coordinated by COMPARE partners in the context of a Pilot Project) was published by *Science* (DOI: 10.1126/science.aaf8852). In this paper, co-authored by 39 individuals from 32 institutes around the world, and based on a combination of analysis of H5N8 viral sequences, epidemiological investigations, waterfowl migration, and poultry trade, showed that long-distance migratory birds played a major role in the global spread of H5N8 avian influenza virus in 2014 and 2015. Further, it showed that the hemagglutinin of clade 2.3.4.4 virus was remarkably promiscuous, creating reassortants with multiple neuraminidase subtypes.

This improvement in our understanding of the circumpolar circulation of avian influenza viruses in migratory waterfowl will help to provide early warning of threats from avian influenza to poultry, and potentially human, health.

As a follow-up to the study, the Global Consortium analyzed H5N8 and related viral sequences from the 2016-2017 global outbreaks. The pilot studies on the added value of NGS approaches over analyses of consensus data are still ongoing.

A good amount of work has been done on the sewage projects, including protocol development for nucleic acid purification steps, NGS procedures and bioinformatics pipelines. A zoonotic infection with swine influenza A/H1N1 virus was detected in the Netherlands, and characterized using NGS.







#### WP9 – COMPARE platform

During the reporting period, we have made substantial progress in a number of areas. While continuing to operate the COMPARE datahubs, COMPARE's infrastructure for pre-publication structured data sharing, we have introduced a number of improvements.

These include a more systematic and agile approach to the configuration of data/metadata and file manifests for new datahubs and the introduction of the new 'AMR Antibiogram' data type and associated standard to support the sharing of measured minimal inhibitory concentrations of antimicrobial agents to provide training data for the COMPARE AMR Working Group's machine learning initiative (see WP 3/6).

On the computational side, we have integrated within the COMPARE cloud compute environment two analytical workflows that are now available for systematic and autonomous operation (using the SELECTA system) upon incoming raw sequence data. The first of these is the 'Center for Genomic Epidemi9ology (CGE)' analysis workflow (now in its second iteration) for bacterial identification, MLST typing and AMR gene identification, etc., from isolates; and the second is the 'SLIM' system for the identification (using a K-mer approach against a viral family-specific reference protein database) of virus from mixtures. While these workflows have their origins in existing pipelines and draw on significant analytical method experience, the integration work allows these analyses to be operated at scale across large numbers of incoming data from COMPARE. We continue to work on the RIEMS viral analysis, parasite and evergreen tree workflows for future integration into the compute environment.



Figure 6. A schematic of the notebook for a pilot project.

Work has continued on different aspects of the COMPARE Data Portal. Due for release later in 2017, this web and programmatic interface will provide an entry point for all COMPARE data (be they in datahubs or publicly available) in an intuitive environment. The back-end service, a data search and retrieval API that operates across



pre-publication and public data, was released in beta in early 2017, and the user authentication components have matured during the reporting period. A part of the Data Portal is the COMPARE Notebook system, which allows users with some scripting ability to rapidly access, explore and visualise data. This has progressed with Notebooks for various pilots and demonstration projects and a number of technical advances including the 'Dockerisation' of the Notebook infrastructure for portability and integration into a single sign-on system.



#### WP10 – Risk communication tools

In the reporting period, WP10 has further developed the first version on the message map and some preliminary communication tools. In particular, activities focused on the creation of several templates to be used in preparation of the message map, including templates for:

- Situation Analysis
- Audience and Audience Segmentation
- Barriers
- Influencers
- Efforts
- Health Communication Support
- Phasing
- Message Elaboration
- Strategic Approach
- Message Design
- Channels and Tools

A message map template, including guidelines for completing and guidance for writing messages according to different situations and contexts, has been created. The message map template has been developed both as an MSWord text and an MSExcel spreadsheet. The Excel spreadsheet includes a dynamic tool for the assessment of message contents. Messages have been linked to the stakeholder dynamic list (spreadsheet) in order to connect different message sets to specific stakeholder groups. A short Message Debriefing Manual (v.1.0) has been also finalized. Moreover, the first version of two short manuals, devoted respectively to "Effective Communication" and to "Propaganda Techniques", has been completed.



#### Message Mapping Tool

Note: Sort by selecting all cells then click "Data"in the navigation menu and "Sort". You can sort by Score (Largest to Smallest)

Messages	Score	Effectiveness	Credibility	Resonance
Risks	0.0	0.0	0.0	0.0
Prevention	0.0	0.0	0.0	0.0
Take Action	0.0	0.0	0.0	0.0
Assurance	0.0	0.0	0.0	0.0
What we don't know	0.0	0.0	0.0	0.0
Treatment	0.0	0.0	0.0	0.0
Test	0.0	0.0	0.0	0.0
Sex	0.0	0.0	0.0	0.0
Non Pregnant	0.0	0.0	0.0	0.0
Travel	0.0	0.0	0.0	0.0

Figure 7. A screenshot from the draft Message Mapping tool being developed for COMPARE.



In parallel with this main group of activities, we finalized the preparatory materials for the workshop on risk communication, to be held during the COMPARE 2017 General Meeting. The materials included 1) a dossier on "Risk Communication and Zika"; 2) the academic version of LIWC2015, which is the gold standard software for computerized text analysis; 3) a computerized analysis of ZIKA communication on social media carried out by using the Receptiviti software; 4) an online psychometric test on risk taking attitude (modified COMPARE-Dospert); and 5) slide presentations about effective communication techniques and message map template.

From September to December 2016, WP10 has also regularly published the weekly bulletin on Risk Communication and EIDs (12 issues). During January and February, the weekly bulletin was suspended to redesign its layout. Bulletin publication started again in March 2017 with a new layout.

Liaisons with other COMPARE WPs included WP14, as per the preparation of WP14 survey on perceived benefits of outbreak prevention, and with the GABRIEL network (led by the Centre International de Recherche en Infectiologie - CIRI), concerning the validation of the message map.

#### WP11 – User consultations

During the period September 2016 – February 2017, several changes in the Expert Advisory Panel (EAP) composition took place. Actions were undertaken, such as replacement of a few EAP members who changed jobs and addition of a new EAP member, to ensure continuous and active participation of EAP members as well as a balanced composition of EAPs. These EAP members are representing key user stakeholder organizations and individual experts that are representative of these user stakeholders.





Figure 8. EAP member Changwen Ke, Center for Disease Control Guangdong, China, speaks with Coordinator Marion Koopmans and PhD student Reina Sikkema during the COMPARE General Meeting 2017.

The third round of EAP meetings has been planned and organized. The members of the EAP were invited to attend the COMPARE General Meeting (1, 2, and 3 March 2017). In particular, EAP members were encouraged to participate in the WP meetings and cross-WP workshops on Barriers and Risk Communication providing their expert advice and perspectives on the topical discussions. At the end of the second day of the General Meeting, 2 March 2017, a COMPARE Executive Board and EAP member meeting was held. During this meeting, the EAP members were invited to provide their feedback to the Executive Board on the progress of the project and strategic choices made by COMPARE thus far.

In addition, EAP members who did not participate in a survey from WP12 offered during the summer of 2016, participated in autumn 2016. This survey aimed to identify the barriers for data sharing that are related to the ownership of data and regulations assigning ownership of data. For this, WP11 and WP12 closely collaborated to engage the EAP members in the survey, and initial observation were presented as a poster during the COMPARE General Meeting. Results from this survey will be presented under WP12.



#### WP12 – Barriers to open data sharing

During the months of September and October 2016, a literature review was performed in order to obtain a comprehensive overview of the non-technical barriers for the sharing of Microbial Genetic Data (MGD) as described in "white, black and grey" literature. This review provided inputs for the WP12 Deliverable 12.1 as a report for the EC.

During the period (September 2016 – February 2017), extra interviews and an improvement in the qualitative analysis was performed to fine-tune and finalize the in-depth study about ownership barriers and respective regulations for the timely sharing of MGD in open-access platforms. This study consisted of interviewing 52 key opinion leaders inside and outside the EU from different domains (human, animal and food) and sectors (research institutes, commercial sector, national surveillance centers and supranational organizations). Besides giving inputs for the Deliverable 12.1, this qualitative research also resulted in an oral and poster presentations at the Yong COMPARE meeting 2017, titled: Four main reasons why people will not share Microbial Genetic Data in COMPARE (M.Y. van Roode, C.S. Ribeiro e.a., COMPARE/RIVM/Erasmus MC/VU University, February 2017). Additionally the results of this study are being assembled to be published as a concept article in an international peer-reviewed journal.

Starting from November 2016, the WP12 working group has been preparing Deliverable 12.1 that consists of a report to the EC and the COMPARE Consortium, listing the identified non-technical barriers for the sharing of MGD in COMPARE. Besides providing an inventory of barriers, the report also explains, analyzes and describes them within the current context. The deliverable report is being currently finalized and will be soon available for the COMPARE community.

Also in the months of January and February 2017, the WP12 team worked on preparing a workshop that took place during the COMPARE General Meeting (2 March 2017) and the GMI 10<sup>th</sup> Meeting (16 May 2017). This preparation consisted of developing a document for the participants of the workshop, which provided background information about the barriers to be discussed in the workshop. In addition, we provided an overview of the activities and insights obtained in WP12 so far, and elaborated on the development of case studies (real-life situations) that were presented for discussion and voting during the workshops of both COMPARE 2017 and GMI-10.





Figure 9. WP12 working group performing the workshop on barriers at the GMI-10 meeting.

Frequent networking, teleconferences and mail discussions took place with different experts and project leaders from partner projects (GMI, EVAg, ZikAlliance). Besides the sharing of information and experiences, expert advice and support was provided in relation to compliance with regulations and requirement for sharing MGD and materials.

The collaboration between COMPARE and other EU-funded projects continues to be a pertinent issue. On 13/14 October 2016, WP12 participated in a meeting with the EVAg group and members of the European Commission to discuss issues about the Nagoya Protocol and the future of Biological Resource Centers. Additionally, WP12 was invited on October 26 in a meeting with the EC and Industry to discuss a document produced by the Commission (DG Environment) to clarify the impact of the Nagoya Protocol for the vaccine and diagnostic industries in Europe and standardize compliance requirements. In this meeting, WP12 advocated for developing a similar document/approach for the Public Health sector, which relies on timely sharing of data and materials.



### WP13 – Dissemination and Training

The COMPARE public website is continuously updated with information regarding upcoming events, news of notes and the most recent publications.

The COMPARE Twitter account (@CompareEurope) shares related news about COMAPRE and from the project partners.



Figure 10. COMPARE Twitter account(@CompareEurope)

COMPARE Workpackage Snapshots are shared throughout the Consortium to keep all involved in the project what achievements are being made.

Regarding Training, COMPARE will utilize e-learning that can be accessed online via the COMPARE portal. A draft overview of the various training materials is under development.

COMPARE will develop modular workshops for the organizations in the Expert Advisory Panels (EAPs) and others. In these workshops, principal investigators will present the analytical tools and software tailored to the specific needs of the workshop audience.



#### WP14 – Cost-effectiveness framework

Erasmus University Rotterdam (EUR) has conducted a review of the existent literature on 'value of safety'; synthesizing the methodologies used in empirical research papers that value safety. The reviewed papers come from different scientific fields, including environmental economics, transport economics, food safety, crime, and health economics; proving that any future policy or project that requires safety valuation can benefit from the results of this paper. As the outcomes from these various fields are highly incomparable, the main focus of this study is on the methodology and covariate results, as these are the most comparable aspects of all the papers. The results of this review feed into the design of the questionnaire, which will be tested, refined and subsequently used in several countries to establish the value of safety associated with interventions like COMPARE. This work is relevant for Task 4 (Develop and apply a methodology to value safety in several countries).

Furthermore, EUR has been working on the first case study meant to illustrate the potential costs and benefits of COMPARE. The case study revolves around the Ebola epidemic in Western Africa and aims to estimate the costs and benefits of early and targeted response in an epidemic scenario in low-income countries. Results indicate that large health benefits as well as wider societal cost savings could be achieved as a result of early and targeted methods and tools using case studies). Finally, we have started a collaboration with the WHO in order to set up a case study with the aim to estimate costs and benefits of upgrading current disease surveillance networks with NGS. Besides being relevant for Task 5, this work is also relevant for Task 3 (Define the baseline (current system and standard methods)).



Figure 11. Cost of Ebola.



Civic Consulting has been working on the focus of the case studies and the related methodology. We have worked on developing the methodological approach and determining the baseline scenario, based on the previous identification of cost and benefit types, existing methodologies and empirical studies. We have also reviewed options for case studies and conducted research and interviews.

Results so far suggest focusing the case studies on estimating costs and benefits of a scenario in which existing surveillance/reference laboratory functions or networks, which are often targeted at specific pathogens, are upgraded with NGS. The case studies would include looking into potential costs savings and/or synergies due to cross-pathogen methods for pathogen identification (e.g. in a scenario of a "catch all" surveillance/reference lab system).

Research into how benefits can be quantified and how approaches to quantifying costs can be refined is ongoing. In terms of scope, the focus would be on EU countries or regions and the relevance of cases in third countries is being considered. A pilot case study will be conducted on the Friedrich Loeffler Institute (FLI), with a visit to the institute for data collection planned for April 2017. Other options for case studies with organisations that either have already upgraded their existing surveillance / reference laboratory function or network with NGS or are planning or in the process of implementing such a pilot project have been considered, and we are in contact with organisations in Denmark and in the UK in this regard.

#### WP15 – Management

The appropriate organizational structures and processes have been put in place to respond to the EC's as well as partners' needs and to ensure COMPARE's compliance with the EC Grant Agreement and the COMPARE Consortium Agreement.

COMPARE successfully submitted the First reporting Period Report and distributed payments based on the report. Information about the project and the achievements from the workpackages are continuously collected.

In the next period, COMPARE will submit an Amendment to the Grant Agreement to reflect the changes that have been made to the Consortium.



