EFSA's Scientific Colloquium on the Use of Whole Genome Sequencing (WGS) of food-borne pathogens for public health protection

Dr. Ernesto Liebana
Head (a.i.) BIOCONTAM Unit

7th Meeting on GMI in the UK,
11-12 September 2014,York

OUTLINE

- Background
- Aims and objectives
- Questions for discussion
- Ongoing work
- Preliminary conclusions

BACKGROUND

- EFSA 10th Anniversary Conference: **The creation of a centralised microbial WGS database is needed** to predict the outcome of pathogen-host interactions.

- SCoFCAH 2012 V**ision paper on the development of databases for molecular testing of foodborne pathogens**: EFSA to establish database for data from isolates in food, feed and animals. Start with PFGE and MLVA data and then gradually include other methods, such as WGS.

- EFSA has **formally agreed to collaborate closely** with **ECDC** and **EURLs** on providing scientific analyses of molecular typing of FB pathogens.

BACKGROUND

- EFSA **mandate from DG SANCO to provide technical support** on building a molecular typing database for FB pathogens (food and animal data), to be operational in pilot phase as from **01/2015** .
  - database, data dictionary, operational tools and periodic analysis (joint) of data.

- The **BIOHAZ** Panel opinions on **"**Evaluation of **molecular typing methods** for major food-borne microbiological hazards and their use for **attribution modelling**, **outbreak investigation** and **scanning surveillance".**
  - Focus on food and feed-borne zoonotic bacteria.

BACKGROUND

- Open Call for tenders on '*Closing data gaps for performing RA on L. monocytogenes in RTE foods*' molecular characterisation employing WGS of strains from different compartments along the food chain, inc. humans (**evaluation of offers currently ongoing**).

- **June 2014:** EFSA organised the **Scientific Colloquium on the use of WGS of foodborne pathogens for public health protection. Overall objective:**

Support and drive EFSA's efforts in the collection of molecular typing data by proactively anticipating the specific requirements and challenges of WGS data.

AIMS AND OBJECTIVES

DG 1: WGS of foodborne pathogens in action

- **Q1. M**ethods available: cost, speed, accuracy, convenience, practicality and feasibility.
- **Q2. M**eans available for data interpretation: accuracy of different approaches, and the ways in which plain language reports can be generated for public health action.
- **Q3. D**ata curation and storage to ensure continuity of existing datasets. How to derive information from WGS data to predict e.g. serotype, phagetype, PFGE type, MLST, resistome, virulome ?

AIMS AND OBJECTIVES

DG2: Curation and analysis of WGS data: bioinformatics solutions

- **Q1. C**hallenges linked to quality evaluation, annotation, interpretation and storage of data, from raw data to genome assembly and analysis results.
- **Q2. H**armonisation of approaches used for data analysis, including development of analysis pipelines and software (e.g. open source *vs.* commercial) and the feasibility of international standards for data analysis.
- **Q3. B**enefits of specialised online databases for sharing WGS data and associated metadata, and algorithms for real-time data analysis and visualisation.

AIMS AND OBJECTIVES

DG3: Cross-sectorial coordination and international cooperation

- **Q1. C**hallenges to integrate WGS into routine monitoring & surveillance, and outbreak preparedness within and across sectors ?
- **Q2. C**oordination of efforts between the food, veterinary and human health sectors.
- **Q3. S**peed and opportunities for collaboration to develop and validate cross-sectorial WGS applications (i.e. human, animal and food). Challenges with capacity building and transition management in introducing new technologies *vs.* traditional typing methods.

WHAT NEXT ?

Ongoing work

- Drafting summary report to be published in the EFSA Colloquium series by the end of 2014:
  - Consultation to all participants (ca. 90): will be given opportunity to comment on complete merged draft.

PRELIMINARY CONCLUSIONS DG 1

Q1. Methods available

- WGS techniques are continuously evolving (Life technologies PGM Ion torrent, Illumina MiSeq, 3rd generation sequencing techniques)
- Setting up sequencing pipeline *vs.* outsourcing sequencing and data analysis to commercial/PH laboratories
- Need to define appropriate quality metrics (including data analysis): EQA programs, role of EURLs
- Need to establish guidelines for use of WGS in detection of FB pathogens (rules on the minimal coverage of genomic data, reproducibility, accuracy)
- Costs: for WGS, for data analysis, suggestions for decreasing costs.

PRELIMINARY CONCLUSIONS DG 1

Q2. Interpretation of data for different applications

- SNP calling *vs.* comparisons of allelic variants (gene-by-gene comparison). Plurality of approaches; depends on precise question
- Not clear yet which method most suitable in food safety
- Need for outcome reported in plain language widely understandable and interpretable
- Need for communication with policy makers
- Harmonisation across sectors

PRELIMINARY CONCLUSIONS DG 1

Q3. Curation and storage of data

- Uncurated approach (e.g. GenBank) not suitable for food safety and public health purposes
- WGS data should be publically available in real-time together with some metadata
- Interoperability of databases and backward and forward compatibility to other sequence based datasets
- Ensure continuity of existing and future databases

PRELIMINARY CONCLUSIONS DG 2

Q1. Quality evaluation, annotation, interpretation and storage of data

- Need for standard protocols and quality metrics (sample preparation, DNA concentration/quality, library preparation, data processing, analysis and storage)
- Quality assessment metrics may depend on technology
- Currently no gold standard for analysis
- Storage and management of large FASTQ files
- Multi-disciplinary forums, case studies, ring trials
- For PH purposes: SOPs for accreditation purposes

PRELIMINARY CONCLUSIONS DG 2

Q2. Harmonisation of approaches for data analysis

- Effective international surveillance depends on common nomenclature
- Harder to define nomenclature based on K-mer/SNP approaches
- Different groups hosting identical pipelines locally *vs.* single location for analysis
- Open source freely available *vs.* closed source commercial software
- Commercial software should not be "black boxes"
- Workflow managers/schedulers, e.g. Galaxy: web-based access to individual tools
- Ever-growing need for computing power

PRELIMINARY CONCLUSIONS DG 2

Q3. Online genomic databases, data sharing and real-time data analysis and visualisation

- Database interoperability to facilitate analysis algorithm development
- Single global database *vs.* federated distributed multi-level databases
- Encryption is needed for secure data transfer of data and metadata
- Sharing of microbial data and analysis driven by One Health perspective *vs.* differing rules for data ownership and release within different sectors
- National differences in data release *vs.* need for agreement and harmonisation of policies
- Any database or resource is only as good as the data that it contains, and how easy it is to access that data

PRELIMINARY CONCLUSIONS DG 3

- **Q1. Integrate WGS into routine monitoring & surveillance, and outbreak preparedness**
  - Concerns related to sharing related epidemiological data
  - Food and veterinary sector handle sensitive data (misuse can have adverse economic impact)
  - First attempt: new molecular typing database of foodborne pathogens (enhance outbreak preparedness at EU level)
  - Many international and national initiatives: difficult to have overview of applicability to routine monitoring and surveillance

PRELIMINARY CONCLUSIONS DG 3

Q2. Coordination of efforts between the food, veterinary and human health sectors

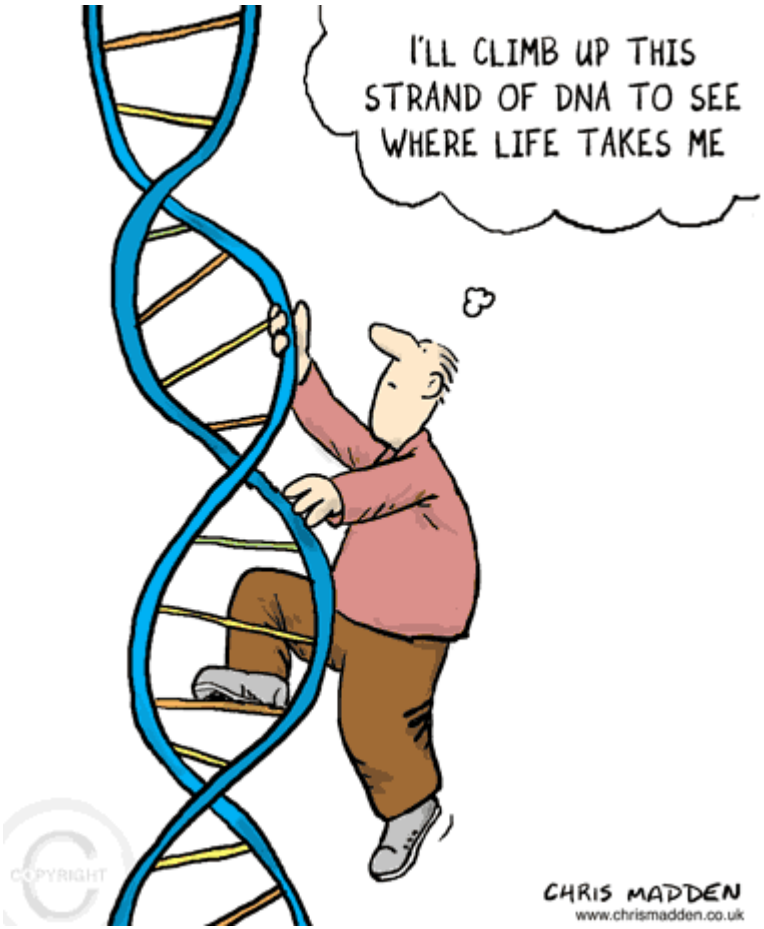
- EC's Vision paper: starting point to develop vision on data sharing across sectors, communities and professional disciplines
- Reporting sequence data alone useless for surveillance and scientific purposes
- Ongoing discussions at ECDC/EFSA level in the context of common joint molecular typing database for PFGE/MLVA
- Minimum dataset to be shared

PRELIMINARY CONCLUSIONS DG 3

Q3. Development and validation of applications across sectors; transition management challenges

- Proof of concept studies needed to demonstrate added PH value, e.g. international FB outbreak/epidemic situations
- Don't underestimate training needs
- EURLs play a crucial role in supporting the transition from old to WGS methods in the food sector
- Close collaboration of EURLs with PH laboratory networks. Cooperative network around EFSA and ECDC
- Start now the investments in national capacities (equipment, application tools and competence building)
- Replacement of old techniques requires comparative analytical and epidemiological validation studies
- Open a discussion with policy makers on the potential impact of WGS development on EU legislation

Questions?


I'LL CLIMB UP THIS STRAND OF DNA TO SEE WHERE LIFE TAKES ME
CHRIS MADDEN
www.chrismadden.co.uk