STSM Report

COST Action: FA1408
STSM title: Use of Next Generation Sequencing for whole genome analysis of *Cryptosporidium* and *Giardia*
Reference : ECOST-STSM-FA1408-050416-072894
STSM dates: from 05-04-2016 to 12-04-2016
Location: Istituto Superiore di Sanità, Rome, Italy
Host: Dr Simone M. Cacciò

ISS staff: Dr. Anna Rosa Sannella, Dr. Alessia Possenti, Dr. Patrizia Rossi, Dr. Gianluca Marucci

Work carried out on 05.04.2016.

I.
-Welcome at the European Union Reference Laboratory of Parasitology, discussion and adoption of the workplan.
-A water sample concentrate (concentration performed by IMS) originating from Hungary has been checked for the presence of *Giardia* cysts, and the cyst number counted by Thoma counting chamber (1000 cysts/75µl). Beside water sample other samples of human and animal origin (cat, parakeet) were available and have been used in the experiments described below.
-DNA extraction was started using Qiagen Mini Kit using the protocol provided by the company until the step proteinase K treatment (which was applied overnight). At the beginning of the procedure 5 cycles of freezing-thawing were added to rupture the cysts.

II. Next generation sequencing (NGS):
ISS is partner of a large Horizon 2020 European Project (COMPARE) that aimed at implementing the use of NGS for pathogen detection, including parasites. ISS has used Illumina sequencing to generate whole genome sequences of *Cryptosporidium* isolates. NGS generates huge numbers of short reads (100-150 bp), and these large datasets needs to be analyzed by dedicated software. The aim of this training was to introduce the useage of a commercial software and to illustrate the various steps involved in the processing of NGS data.

Discussion on computer work to process NGS data:
Software: CLC Genomics Workbench
Data: Fastq files of Illumina-generated *Cryptosporidium* sequences

Steps in data analysis:
1. Removing adapters (TRIM): removing the oligos at the 5' and 3' end of each DNA fragment in a sequencing library. These adapters are complementary to the lawn of oligos present on the surface on the illumina sequencing flow cells.

2. Removing TAGs (barcodes, indexes): a unique DNA sequence (a TAG) is incorporated into fragments within a sequencing library for downstream in silico sorting and identification. Libraries with unique tags can be pooled together (multiplexing). In the last experiments carried out by ISS, 17 samples have been pooled and sequenced in parallel.
3. Removing wrong (failed) sequence data (e.g. containing a number of Ns above a defined threshold).
4. DE NOVO assembly. This step generates “contigs” by identifying overlapping sequence reads from the pool of NGS sequences.
5. Mapping: BLAST using the NCBI database.
6. Genome finishing module. This part of the software allows refining the mapping of contigs to a reference genome, and identifying poorly represented regions for future experiments (e.g., gap filling).

-Discussion on "coverage" (how many times each base is sequenced to allow reliable and robust identification of genetic variants) and statistical data collected during analysis.

Work carried out on 07.04.2016.

-DNA extraction using Qiagen Mini Kit has been completed using the protocol provided by the company.
-DNA was concentrated using vacuum drier (down to a final volume of 20 µl).
-DNA concentration was measured using a Qubit 2.0 fluorometer (water sample contained 0.742 ng/µl DNA).
-Nested PCR has been performed on the prepared DNA using protocol described by Lalle et al. (2005), which targets the Beta-giardin gene.
-RFLP analysis was started on the nested PCR products described above using Hae III digestion as described by Lalle et al. (2005). Digestion proceeded overnight.
-Whole genome amplification using REPLI-g Midi Kit was also prepared for overnight incubation based on the Qiagen protocol.

Work carried out on 08.04.2016.

I.
-After completion of Hae III digestion, the products were run on a 3% high resolution MetaPhor gel (3.5 h, 60V). After staining with Ethidium Bromide, a gel image was acquired and the restriction patterns observed and interpreted. In the Hungarian water sample *G. duodenalis* assemblage A was identified.
-The whole genome amplification products were checked by running an aliquot (1/10) on a 0.8% agarose gel (1.5 h, 60V) stained by EtBr.

II.
-Discussion on accreditation system at EURLP and accreditation of proficiency test (PT):
Basic standard for PT accreditation is ISO 17043, furthermore other useful standards for microbiology ISO 22117 and for statistical analysis ISO 13528.
1. Planning of PT. Key points: how to avoid communication between labs (random generating code of samples), which kind of information should be provided to labs, timing, sample number/panel, mode of distribution of samples, avoiding damaging of

samples, control of stability and homogeneity, final report/individual report, subcontracting;
2. Document preparation: description of test samples and the criteria for evaluation, PT request form, package control form, detailed procedure description, available instruments, reagents and bias, results sheet;
3. Following the schedule
Final report should contain: design, implementation, summary and comments about negative results; final workshop including presentation of results and showing trends; complaining and correction actions, confidentiality, archiving for 10 years, notes and recommendations from participants.

Work carried out on 11.04.2016.

I.
-WGA products were checked for the presence of bacteria by single round PCR targeting 16S rDNA. The WGA from water sample was highly contaminated by bacteria; this can be reduced by treating the sample after IMS with 0.6% Sodium hypochlorite, followed by 3 washes with PBS (5000 rpm, 5').
-WGA products were purified using Qiagen MagAttract HMW DNA Kit according to the description of the manufacturer. After purification the presence of DNA has been rechecked (by gel electrophoresis as described above and Qubit instrument).
-DNA could then be sent to a company to perform all subsequent steps necessary for NGS. For standard library preparation, 1-2 µg of high molecular weight genomic DNA are needed.

II.
Preparation of *Trichinella* proficiency test

-*Trichinella* species are the smallest nematode parasite of humans, have an unusual lifecycle, and are one of the most widespread and clinically important parasites in the world. The small adult worms mature in the intestines of a definitive host such as a pig. Each adult female produces batches of larvae, which enter the blood (to feed on it), lymphatic system, and are carried to striated muscle. Once in the muscle, they encyst, or become enclosed in a capsule. When a human eats the infected meat (pork, horse), the larvae are released from the nurse cell (due to stomach pH) and migrate to the intestine, where they burrow into the intestinal mucosa, mature, and reproduce.

Delivery of PT test of *Trichinella*

1. Preparation of digestion solution for pork meat (carcass) to release capsules and larvae (250 ml tap water 46 °C, HCl, 1.25 g pepsin, pH=1).
2. Mix meat pieces with the digestion solution in a blender.
3. Incubation (43 °C, 5 min).
4. Filtration through nylon filter tissue (250 µm pore size) into V cup.
5. Collection of larvae from the bottom of the cup into Petri dishes.
6. Examination under stereo microscope and transfer the capsules to watch glasses.
7. Spiking of 3/5 capsules in each low fat pork meat ball. No pipetting, just 150 µl PBS flush.
8. Package in vacuum plastic bags and cooling boxes.

Demonstration of the web communication of PT.

Work carried out on 12.04.2016.

-Summary of the experiments, data generated and protocols used
-Writing of the report

Rome, April 12, 2016

I declare that the content of this report is correct.

Dr. Simone M. Cacciò

EURL for Parasites, Istituto Superiore di Sanità, Rome, Italy