Workshop on Next Generation Sequencing of Viruses

Auditorium F Jacob, Institut Pasteur, Paris, France

20 & 21 May 2015

Brief Report

The Workshop on *Next Generation Sequencing (NGS) of Viruses* was organised by the isirv-Antiviral Group in conjunction with GISAID and the PREDEMICS Consortium and held at the Institut Pasteur, Paris on 20-21 May 2015. It was fully subscribed with 117 registered participants from 16 countries, 20% of whom were from outside Europe. Fifteen abstracts were submitted, 8 of which were selected for oral presentation and 7 were presented as posters.

The 2-day expert workshop reflected a timely and urgent need to address the analysis and interpretation of NGS data of viruses, in particular as regards genetic variation in (mixed) virus populations, intra-host diversity, and the significance and potential impact of minor variants emerging in response to immune or antiviral pressure.

The programme included 'state of the art' presentations, reflecting current and future developments, and provided a platform for open discussion and technical exchange relating to sequencing technologies, data processing, assembly and analysis of data for a variety RNA viruses, with a particular focus on quality and interpretation of results and the significance and use of NGS data of different viruses for public and animal health.

Specific aspects addressed included: limitations of NGS in relation to diversity of viruses within a population (including mixed genotypes/subtypes), and of sequences within a virus population; linkage of sequences (markers) within a virus population, especially relevant to segmented viruses; inherent errors in different systems and processes (quality assurance); and quantitative assessment and statistical significance of minority variants.

The success of the workshop was reflected in some excellent feedback, in particular as regards its timeliness, coverage of the different sequencing platforms and analysis pipelines, and variety of RNA viruses discussed. Many participants saw this workshop as a prelude to a subsequent hands-on workshop with more in-depth bioinformatics training.

Generous financial support for the workshop was provided by 4 companies and a grant from APHL.

The expert discussion will form the basis of a report, to be submitted to a special issue of *Viruses* devoted to NGS of viruses, providing guidance on the generation and interpretation of NGS data on viruses.

Programme

DAY 1 – WEDNESDAY 20TH MAY

8:00 – 9:00	**Registration**
9:00 – 9:15	**Welcome/Objectives** **Sylvie van der Werf and Alan Hay**
SESSION 1:	**Objectives of NGS/Deep Sequencing – Possibilities & Limitations** **Chairs: John McCauley and Peter Walker**
9:15 – 9:40	The Evolution and Epidemiology of Virus Epidemics from Genome Sequencing Andrew Rambaut, Institute of Evolutionary Biology, Edinburgh, UK
9:40 – 10:05	Virus Discovery Ron Fouchier, Erasmus MC, Rotterdam, The Netherlands
10:05 – 10:30	Outbreak Detection and Investigation with NGS Martin Beer, Friedrich Loeffler Institute, Isle of Riems, Germany
10:30 – 10:45	Discussion
10:45 – 11:15	**Tea/Coffee Break**
SESSION 2:	**Sequencing Technologies - Current & Future Developments** **Chairs: David Wentworth and Nicholas Loman**
11:15 – 11:40	Platforms and Pipelines (Now and in the Future) David Wentworth, CDC, Atlanta, GA, USA
11:40 – 11:55	Respiratory Virus RNA Detection with RNA-Seq Using Capture Technology Gary Schroth, Illumina, San Diego, CA, USA
11:55 – 12:10	Bacterial and Viral Sequencing with Ion Torrent NGS Technology Mathieu Boimard, Thermo Fisher Scientific, Saint Aubin, France
12:10 – 12:20	From Sample to Sequence-Ready with the Access Array™ System and New Applications in Single-Cell Genomics Cyprien Dulac, Fluidigm Europe B.V, France
12:20 – 12:30	Discussion
12:30 – 13:30	**Lunch**

SESSION 3:	**Sample Preparation, Data processing, Assembly, Analysis** **Chairs: Marco Vignuzzi and Elodie Ghedin**
13:30 – 13:55	Sample Preparation & Analysis Saskia Smits, Erasmus MC, Rotterdam, The Netherlands
13:55 – 14:20	Data Processing, Assembly, Analysis Simon Watson, Sanger Institute, Hinxton, UK
14:20 – 14:35	Consensus Iterator: a Hybrid Algorithm for Consensus Sequence Determination in the Presence of Noisy Short-Read Shotgun Sequence Data Robert Carter, St. Jude Children's Research Hospital, Memphis, TN, USA
14:35 – 14:55	Optimising the Reconstruction of Whole Genomes, for HIV and Other Diverse Viruses, from NGS Data Chris Wymant, Imperial College London, UK
14:55 – 15:10	Discussion
15:10 – 15:30	**Tea/Coffee Break**
SESSION 4:	**Examples of Different Virus Pipelines** **Chairs: Ron Fouchier and Martin Beer**
15:30 – 15:50	RSV and Other viruses Matthew Cotten, Sanger Institute, Hinxton, UK
15:50 – 16:10	MERS Coronavirus Sequences Detected in Dromedary Camels from a Single Farm Leo Poon, The University of Hong Kong, Hong Kong, SAR China
16:10 – 16:30	Intrinsic Genetic Diversity of Rabies Virus and Host Adaptation Hervé Bourhy, Institut Pasteur, Paris, France
16:30 – 16:50	Prospective, Real-Time Nanopore Sequencing for Ebola Genomic Epidemiology under Outbreak Conditions Nicholas Loman, Birmingham University, Birmingham, UK
16:50 – 17:00	Discussion
17:00 – 17:30	**General Discussion** **Chairs: David Wentworth and Elodie Ghedin**
19:30	**Dinner at Restaurant "L'entrepôt"**

DAY 2 - THURSDAY 21ST MAY

SESSION 4:	**Examples of Different Virus Pipelines – Continued** **Chairs: Gavin Smith and Hervé Bourhy**
9:00 – 9:20	Are Arthropods at the Heart of Virus Evolution? Yong-Zhen Zhang, China-CDC, Beijing, China
9:20 – 9:40	Virus Identification in Biological Samples by NGS: Validation and Examples of Use in Clinical Cases Marc Eloit, Institut Pasteur, Paris, France
09:40 – 09:55	Full-Length HIV-1 env Deep Sequencing in a Donor with Broadly Neutralizing V1/V2 Antibodies Melissa Laird, Pacific Biosciences, Menlo Park, CA, USA
09:55 – 10:10	European Mobile Lab NGS Pipeline David Matthews, Bristol University, Bristol, UK
10:10 – 10:30	Discussion
10:30 – 11:00	**Tea/Coffee Break**
SESSION 5:	**Advances in Understanding of Viral Infections from NGS data** **Chairs: Andrew Rambaut and Maria Zambon**
11:00 – 11:25	Going Viral: NGS Platform Integration for Analysis of Virus Diversity and Population Dynamics Elodie Ghedin, New York University, New York, NY, USA
11:25 – 11:50	Data to Knowledge: Data Integration and Big Data Maria Giovanni, NIAID, Rockville, MD, USA
11:50 – 12:05	Lessons from Developing a New Influenza Annotation Pipeline Benjamin Turner, QIAGEN Custom Informatics Solutions, Hilden, Germany
12:05 – 12:20	Discussion
12:20 – 13:20	**Lunch**

SESSION 5:	**Advances in Understanding of Viral Infections from NGS data – Continued** **Chairs: Monica Galiano and Nancy Cox**
13:20 – 13:40	Metagenomic Detection of Immunogenic Parasites in Bat Feces Gavin Smith, Duke NUS, Singapore
13:40 – 13:55	Reconstruction of an Empirical Fitness Landscape Reveals the Mutational Robustness and Evolvability of RNA Viruses Marco Vignuzzi, Institut Pasteur, Paris, France
13:55 – 14:10	The Intrinsic Heterogeneity of Human Influenza A Viruses Evaluated by Deep Sequencing the Virus Directly in Nasal Swabs Cyril Barbezange, Institut Pasteur, Paris, France
14:10 – 14:25	Testing for Drug Resistance in Influenza: Taking It to Another Level? Larisa Gubareva, Influenza Division, CDC, Atlanta, GA, USA
14:25 – 14:35	Use of Next Generation Sequencing for the Detection of Antiviral Resistant Influenza Viruses Aeron Hurt, WHO CC, Melbourne, Australia
14:35 – 14:55	Whole-Genome Deep Sequencing of Longitudinal Samples from HIV-1 Patients Followed from Early into Chronic Infection Richard Neher, Max-Planck-Institute for Developmental Biology, Tubingen, Germany
14:55 – 15:10	Discussion
15:10 – 15:30	**Tea/Coffee Break**
SESSION 6:	**Significance and Communication of Information** **Chairs: Sylvie van der Werf and Maria Giovanni**
15:30 – 15:55	The Public Health Importance of Timely Sharing of Sequence Data Nancy Cox, GISAID Scientific Advisory Council
15:55 – 16:20	Public Health Perspective (Surveillance; Zoonotic Risk) Maria Zambon, Public Health England, London, UK
16:20 – 16:45	Applications of NGS in Animal Health Peter Walker, CSIRO, Geelong, Victoria, Australia
16:45 – 17:00	Discussion
17:00 – 17:30	**General Discussion (Conclusions)** **Chairs: Andrew Rambaut and Maria Giovanni**
17:30	**Close of Workshop**

ORGANISING COMMITTEE

Martin Beer	Friedrich-Loeffler-Institut, Riems, Germany
Ron Fouchier	Erasmus MC, Rotterdam, The Netherlands
Monica Galiano	PHE, London, UK
Maria Giovanni	NIAID, Maryland, USA
Alan Hay (Co-Chair)	The Francis Crick Institute, Mill Hill Laboratory, London, UK
Aeron Hurt	WHO CC, Melbourne, Australia
Paul Kellam	The Sanger Institute, Hinxton, UK
Philippe Lemey	Rega Institute, Leuven, Belgium
John McCauley	The Francis Crick Institute, Mill Hill Laboratory, London, UK
Andrew Rambaut	Institute of Evolutionary Biology, Edinburgh, UK
Gavin Smith	Duke NUS, Singapore
David Spiro	NIAID, Maryland, USA
Sylvie van der Werf (Co-Chair)	Institut Pasteur, Paris, France
Marco Vignuzzi	Institut Pasteur, Paris, France
Richard Webby	St Jude Children's Research Hospital, Memphis, USA
David Wentworth	CDC, Atlanta, USA

Collated Workshop Evaluation

16 Respondents	Very good	Good	Average	Poor
Day 1 - Wednesday 20th May				
SESSION 1: 09:00 – 10:45: ***Objectives of NGS/Deep Sequencing – Possibilities & Limitations*** ***Chairs: John McCauley and Peter Walker***	**53 %**	**47 %**	**0 %**	**0%**
SESSION 2: 11.15 – 12.30: ***Sequencing Technologies - Current & Future Developments*** ***Chairs: David Wentworth and Nicholas Loman***	**33 %**	**40 %**	**27 %**	**0%**
SESSION 3: 13.30-15.10: ***Sample Preparation, Data processing, Assembly, Analysis*** ***Chairs: Marco Vignuzzi and Elodie Ghedin***	**60 %**	**27 %**	**13 %**	**0%**
SESSION 4: 15.30 – 17.30 ***Examples of Different Virus Pipelines*** ***Chairs: Ron Fouchier and Martin Beer***	**33 %**	**60 %**	**7 %**	**0%**
General Discussion ***Chairs: David Wentworth and Elodie Ghedin***	**27 %**	**47 %**	**26 %**	**0%**
DAY 2 - Thursday 21st May				
SESSION 4 continued: 09.00 – 10.30: ***Examples of Different Virus Pipelines*** ***Chairs: Gavin Smith and Hervé Bourhy***	**38 %**	**56 %**	**6 %**	**0%**
SESSION 5: 11.00 – 12.20: ***Advances in Understanding of Viral Infections from NGS data*** ***Chairs: Andrew Rambaut and Maria Zambon***	**38 %**	**62 %**	**0 %**	**0%**
SESSION 5 continued: 13.20 – 14.55: ***Advances in Understanding of Viral Infections from NGS data*** ***Chairs: Monica Galiano and Nancy Cox***	**25 %**	**63 %**	**12 %**	**0%**
SESSION 6: 15.30 – 17.30: ***Significance and Communication of Information*** ***Chairs: Sylvie van der Werf and Maria Giovanni***	**36 %**	**36 %**	**14 %**	**14%**

Any other specific comments?

- Very interesting and useful workshop. Especially liked Elodie Gedhin's talk, but all talks were very good. Excellent workshop

- Although organised by isirv, topics were not restricted to respiratory viruses. This was absolutely the right approach for this meeting, given the technical nature of the workshop. It was good to have views from a range of applications

- The programme promised (in my opinion) more detailed descriptions of methodology, especially sample preparation. Most speakers only briefly touched the subject of sample preparation. In our experience the main difficulty with working with viruses, particularly clinical samples, is in obtaining enough viral material to allow detection. More information regarding how others have solved such problems would have been appreciated. The presentations were more aimed at a "conference-style" meeting presenting results, rather than a "workshop-style" meeting focusing on troubleshooting and sharing experiences.

- Lack of bioinformaticians

- The workshop was very good to provide a basic understanding of the capacity of NGS in virus studies- as exemplified by the talks by the different investigators on their experiences. However, a more advanced workshop could also be organized- one with a more hands-on, technical approach on the bioinformatics side of analysis.

- It was perhaps challenging to balance the discussion of data and scientific content with the in-depth discussion of methods and approaches that is necessary for a workshop. On balance I felt the workshop was successful, but not outstanding. The main weakness was that speakers in general provided overviews of their workflows, pipelines, and analyses without giving important details. Discussions were helpful, but also were not particularly focused. I think a future workshop could be much improved by posing a series of questions for each session that could help focus talks and discussion points. For example, what are specific challenges for sample preparation when sequencing known and unknown viruses, and how can they be overcome? What are the strengths and weaknesses of the major methods for library preparation, the major instrument platforms, and the major data analysis strategies for particular applications? Can we define a set of best practices for specific applications of NGS? Some of these questions were addressed in an ad-hoc fashion by particular speakers, but it would have been helpful to have sessions more explicitly organized and focused around these sorts of questions. I realize this is challenging because of the heterogeneity of approaches and applications for NGS and associated technologies, and again overall I felt the workshop was efficient and organized. However, I think increased focus and organization along these lines would really improve future versions of this workshop.

  A few additional minor suggestions:

  1. Including a session on the importance of communicating information to the public was an excellent idea. I particularly welcomed Dr. Cox's points on the need for frameworks for cooperation among scientists and clinicians in the developing and developed worlds. Unfortunately I felt that the session was long on case studies and short on generalizable recommendations for how scientists and clinicians should best collaborate and communicate NGS studies to the public and with their colleagues. Again a focus on developing recommendations for best practices would have been welcome.
  2. Without a formal poster session I felt that poster presenters were given little opportunity to present their data and interact with colleagues about their work. In the future I would either have a dedicated poster session with more posters (perhaps during / after a lunch break?), or else abolish posters altogether.

Would you attend a similar Expert Workshop again in the future? 100%- Yes / 0%- No

Explanatory answers;

- Yes, but only if there is a more technical component on it.
- Yes, I would attend and/or send people from my laboratory if I felt the sessions would be more focused on detailed discussion of either data or methods and best practices.

Would you recommend the Workshop to your colleagues? 93% -Yes / 7% - No

Explanatory answers;

- Yes, if they are interested in obtaining a broad overview of NGS.
- No, I would probably not recommend the workshop in its current form to colleagues with experience in NGS.

What aspects of the Workshop did you like most? And least?

Most:

- As we do not have much hands-on NGS yet, I really liked the talks about the differences between the different NGS platforms.
- The sessions were well organised, and generally ran according to the timetable. The auditorium was good, though there were some issues with the sound during some of the presentations.
- Interesting topics, good presentations.
- Regular breaks; no more than 4 talks per session and open discussion at end of each session. Very constructive! Talks all in series with no parallel sessions. All talks a good length, not too long.
- Several commented on having plenty of time for discussion and a friendly atmosphere.
- Meeting the other investigators that work on similar problems.
- I liked most the high level of competence and experience of the speakers
- The variation in the talks/topics and that there were so many excellent speakers
- I mostly liked the aspects of: comparing different platforms, everything related to influenza and specifically resistance testing. It was good to get a broad overview of how other groups use NGS.
- Sample preparation and assembly workflow
- Many aspects of NGS were covered. Good overview of the field at this moment. It was very good that there were no parallel sessions, so no need to choose.
- The food provided for the lunches was tasty!
- Head-to-head comparison of different platforms
- Applications of technology that were virus specific

Least:

- The food provided for the lunches was tasty, but very difficult to eat. There was a lack of tables to put plates down on, it is not easy to eat big pieces of meat (big enough to necessitate cutting into smaller bits) whilst standing and balancing plate and glass.
- I had preferred if some of the speakers had more time to present their data.
- The buffet lunch was poor for French standards.
- Although it was nice to read the posters, there should have been a fixed time when the people who made the posters should have stood there so that you could ask questions about the poster.
- Apps and software for analysis are still inaccessible for biologists who are unable to code in R. After hearing a talk on analysis pipelines, it is still hard to return home and try them on your data.

Do you have any suggestions for improving such a Workshop?

- More troubleshooting session rather than just presenting success stories. More focus on methodology.
- Apart from improving the lunch facilities and catering, I don't think there is anything that could be improved.
- A technical training on bioinformatics analysis. Perhaps speakers from outside of the microbiology field that may have some innovative ideas on NGS utility.
- A few more student presentations?
- An idea could be to couple the workshop with a one-day hands on tutorial on NGS pipelines and analyses,

where you could perhaps bring your own data and get comments from others/tutors.

- More information about which quality the data should have to be trusted, like which minimum coverage/average coverage to use for surveillance, minimum coverage to use for clinical diagnostics (influenza NA resistance), How minor variants (and to what percentage of the whole population) should be reported to databases (GISAID).
- Add more time for discussion with other participants (during lunch/coffee breaks etc). The seated dinner did not allow for meeting different people.
- Morning coffee/refreshments on arrival on day 2 as well as day 1.
- A demo/webex of exactly how to tackle analysis would be helpful

Is there anything you would like to see at a future Workshop that was not included?

- More troubleshooting and methodology.
- Novel bioinformatics methods that have potential in the field of Viral NGS analysis.
- A technical training on bioinformatics analysis.
- Proposals for standardised surveillance approaches of virological diseases in the public held domain (medical or veterinary)
- Some more practical related sessions.
- I would've liked to see presentations on sample and library preparation. Additional presentations on data curating and analysis pipelines would be nice.
- Solutions for storing data are going to be important in the future.
- It would be good to involve communications experts in these meetings because uploading or transferring data is still a problem. There are isolated good examples e.g. Mission Rabies use of mobile phone technology and the cloud to upload data (in use for Rabies Veterinary Vaccine field trials).

Many thanks