High Throughput Sequencing, Bioinformatics & Computational Genomics (HTS-BCG)

Evolution or Revolution?

11th OIE Seminar, Saskatoon, Canada 2015
Fredrik Granberg

OIE Collaborating Centre - for Biotechnology-based Diagnosis of Infectious Diseases in Veterinary Medicine
Swedish University of Agricultural Sciences (SLU) & National Veterinary Institute (SVA), Uppsala, Sweden
About Us & What We Do
OIE Collaborating Centre
- for Biotechnology-based Diagnosis of Infectious Diseases in Veterinary Medicine
Uppsala, Sweden

The National Veterinary Institute (SVA)
- Routine diagnostics & research
- 575,000 samples analyzed/year
- BioSafety Level 3 laboratory

Veterinary and Animal Science Centre (VHC), 2014
- Around 50,000 m²
  - Teaching facilities
  - Research labs
  - Animal premises
- Sweden’s only University Animal Hospital
Expertise & Competences

Always trying to find important new subjects and fields...

- Gel-based classical PCR (since 1987)
- First four PCR kits (in 1990)
- Real-time PCR systems, TaqMan, PriProET, LUX, SYBR Green...
- Isothermal amplification platforms, RT-LAMP....
- Solid phase microarrays, padlock probes
- Liquid phase microarrays, Luminex
- Broad-range detection platforms
- Antigen, antibody “amplification” with proximity ligation
- Variation tolerant systems, VOCMA
- Novel PCR systems for rapid pathotyping of RNA viruses
- Full-genome sequencing
- High-Throughput Sequencing (HTS)
High-Throughput Sequencing
Evolution of sequencing technology

First Generation Sequencing
- *Sanger Sequencing* [1977]
  - Gel-based systems
  - Capillary sequencing

Next (or second) Generation Sequencing (NGS)
- *Massively parallel sequencing* [2006/2007]
  - 454 / Roche sequencing
  - Illumina (Solexa) sequencing
  - SOLiD systems
  - Ion Torrent sequencing

Third Generation Sequencing
- *Singel molecule sequencing* [2013]
  - PacBio RS II (Pacific Biosciences)
  - MinION (Oxford Nanopore)

Fourth Generation Sequencing?
Increase in sequencing capacity

Decrease in sequencing costs

Data from the NHGRI Genome Sequencing Program (GSP)
Exponential growth of information in GenBank

- High-Throughput Sequencing (HTS)
- Capillary sequencing
- Gel-based systems
HTS-based Applications
HTS-based diagnostics

• Primary diagnostics – Detection
  – Detection, identification & characterization of previously unidentified microorganisms
  – Molecular marker profiles directly from clinical samples

• Adjunct (secondary) diagnostics - Further characterization
  – Whole genome sequencing
  – Pathotyping or resistance typing information
Two major approaches

• **Unbiased (random) sequencing**
  – Metagenomics studies → Pathogen discovery
  – Whole genome sequencing

• **Targeted (amplicon) sequencing**
  – Characterization of known pathogens
  – 16S rRNA profiling of bacteria
  – Deep sequence coverage for minor population variants

---

Metagenomics

*"The genetic composition of entire communities of microbial organisms"*
Pros and Cons with HTS-based applications

HTS-based detection and characterization of pathogens

+ Unbiased (all genomes in a sample)
+ Cultivation-independent
+ Robust
+ Large number of samples (targeted sequencing)
+ Different type of samples
  - Biopsy
  - Blood
  - Feces
  - Swab
  - Isolates

- Time-demanding
- Still expensive (getting cheaper)
- A well-equipped laboratory
- Generates large amounts of data
- Random and Systematic error
- Advanced bioinformatics
- Validation and quality assurance for diagnostic use
General methodology for HTS-based applications

1. Sampling
2. Sample preparation
3. Pre-amplification
   - Random amplification
   - Targeted amplification
4. Preparation of sequencing library
5. High-throughput sequencing
6. Bioinformatics
7. Follow-up
1. Sampling

• Clinical material
  – Collect samples according to observed symptoms and applicable recommendations
  – Tissue/body fluid most likely to contain the pathogen
  – Safe transport & storage of samples (prevent degeneration)
  – Correct & complete documentation
  – Zoonotic diseases
    • “One world, One Health”
  – ’First responders’

• Cultivated material
  – Normal laboratory procedures
2. Sample Preparation

- Homogenization
- Filtration

- Enrichment
  - Ultracentrifugation

- Removal of host material
  - Filtration
  - Treat with nucleases (RNase & Dnase)

- Extraction (RNA och DNA)
3. Pre-Amplification

- Targeted amplification (Amplicon)
- Random amplification

MDA (multiple displacement amplification) by *Phi29* DNA polymerase

Sequence-independent, single-primer amplification (SISPA)
# 4 & 5. Library preparation & Sequencing

<table>
<thead>
<tr>
<th>Smaller bench-top sequencer</th>
<th>Capacity/Time</th>
<th>Full size sequencer</th>
<th>Capacity/Time</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MiSeq</strong></td>
<td>~6 Gb (2 x 300) 36 hours</td>
<td><strong>HiSeq</strong></td>
<td>~600 Gb (2 x 100) 11 days</td>
</tr>
<tr>
<td>Illumina</td>
<td></td>
<td>Illumina</td>
<td></td>
</tr>
<tr>
<td><strong>Ion Torrent</strong></td>
<td>~1 Gb (400bp) 4 hours</td>
<td><strong>Ion Proton</strong></td>
<td>~30 Gb (150bp) 8-10 hours</td>
</tr>
<tr>
<td>Life Technologies</td>
<td></td>
<td>Life Technologies</td>
<td></td>
</tr>
<tr>
<td><strong>454 Junior</strong></td>
<td>35 Mb (up to 400bp) 12 hours</td>
<td><strong>454 GS FLX+</strong></td>
<td>700 Mb (up to 1kb) 23 hours</td>
</tr>
<tr>
<td>Roche</td>
<td></td>
<td>Roche</td>
<td></td>
</tr>
<tr>
<td><strong>MinION</strong></td>
<td>(up to 10kb)</td>
<td><strong>PacBio RS II</strong></td>
<td>~300Mb (up to 12kb) 2 hours</td>
</tr>
<tr>
<td>Oxford Nanopore</td>
<td></td>
<td>Pacific Biosciences</td>
<td></td>
</tr>
</tbody>
</table>

Mb = mega base pairs: 1,000,000 bp; Gb = giga base pairs: 1,000,000,000 bp
6. Bioinformatics & Computational Genomics

Sequence data
- Quality control
- Assembly of reads into contigs
- Homology searches

- Bioinformatics filtering, sorting and classification pipeline
- Alignments and annotations
Finding traces of a viral or bacterial genome does not mean finding an infectious agent

**Indirect evidence - Correlated**
- PCR validation or antibody detection in matching case studies.
- Localization of antigen in affected tissue(s).
- Comparative study with healthy controls.

**Direct evidence - Causative**
- Reproducing the disease: Inoculation with an isolate, synthetic genome, or biological sample.

**Countermeasures**
- Rapid diagnostics
- Prevent spread
- Vaccine & antiviral drugs or Antibiotics
Recent Findings & Examples
Example: Schmallenberg Virus (SBV)

- First case 2011, Germany
- Detected by using an approach for HTS-based virus detection
- A previous unknown orthonbunyavirus
- Affects ruminants
  - Fever, diarrhoea and reduced milk production
  - Stillbirths and birth defects
- Follow-up: Combination of HTS-based screening and classical approaches, such as virus isolation and rapid characterization of the virus

The European spread in Jan 2013, FluTrackers.com and FLI
Our Findings & Current projects

1. Pigs with Postweaning Multisystemic Wasting Syndrome (PMWS)
2. Shaking Mink Syndrome
3. Honeybees with Colony Collapse Disorder (CCD) Symptom
4. Full-length sequencing of African swine fever virus (ASFV)
1. Pigs with Postweaning Multisystemic Wasting Syndrome (PMWS)

Lymph nodes from animals with PMWS
- Porcine Circovirus Type 2 (PCV-2) - a known contributing factor
- Torque Teno Virus (TTV) – associated
- A novel Porcine Parvovirus with genetic relationship to Bocaviruses

<table>
<thead>
<tr>
<th>Blast hit</th>
<th>% sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCV-2</td>
<td>99.545</td>
</tr>
<tr>
<td>Other viruses</td>
<td>Torque Teno virus: 0.107</td>
</tr>
<tr>
<td></td>
<td>Parvovirus: 0.076</td>
</tr>
<tr>
<td>Pig</td>
<td>0.015</td>
</tr>
<tr>
<td>Other</td>
<td>0.015</td>
</tr>
<tr>
<td>No hit</td>
<td>0.243</td>
</tr>
</tbody>
</table>
1. Pigs with Postweaning Multisystemic Wasting Syndrome (PMWS)

Comparative study with healthy controls: 36 animals with PMWS & 24 healthy

Co-infection by all three viruses:

- PMWS affected animals 71%
- unaffected pigs 33%

Novel Virus
Indirect evidence - Correlated

Blomström et al. 2009 and 2010 Virus Research
Our Findings & Current projects

1. Pigs with Postweaning Multisystemic Wasting Syndrome (PMWS)
2. Shaking Mink Syndrome
3. Honeybees with Colony Collapse Disorder (CCD) Symptom
4. Full-length sequencing of African swine fever virus (ASFV)
2. Shaking Mink Syndrome (SMS)

• Brain homogenates from minks affected by SMS were used to reproduce the disease in 3 healthy individuals.

• Conventional methods could not detect any infectious agent (*Gavier-Widen et al., 2004*).

• HTS-based metagenomic analysis:
  – **Novel astrovirus (AstV)**
2. Shaking Mink Syndrome (SMS)

- **PCR for astrovirus detection:**
  - Detected in naturally infected animals, but not in healthy minks.

- Associated with CNS diseases in various host species:
  - Astrovirus encephalitis in a boy with X-linked agammaglobulinemia *(Quan et al., 2010)*
  - Divergent Astrovirus Associated with Neurologic Disease in Cattle *(Li et al., 2013)*
Our Findings & Current projects

1. Pigs with Postweaning Multisystemic Wasting Syndrome (PMWS)
2. Shaking Mink Syndrome
3. Honeybees with Colony Collapse Disorder (CCD) Symptom
4. Full-length sequencing of African swine fever virus (ASFV)
3. Honeybee Colonies with CCD Symptoms

- Collected in 2010 from commercial hives in the Northern Spain. (Marina Vicente-Rubiano, Consuelo Rubio-Guerri, Deborah Kukielka & José Manuel Sánchez-Vizcaíno)
- Lack of vitality of adult worker honeybees and unusual depopulation.
- **Positive for Israeli Acute Paralysis Virus (IAPV)** by RT-PCR assay.
- IAPV has been linked to CCD - **Other contributing factors present**?

**Results of metagenomic analysis:**

<table>
<thead>
<tr>
<th>Virus</th>
<th>Virus Family/Taxa</th>
<th>Reads</th>
<th>Contigs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aphid lethal paralysis virus (ALPV)</td>
<td>Dicistroviridae</td>
<td>1371</td>
<td>16</td>
</tr>
<tr>
<td>Israel acute paralysis virus (IAPV)</td>
<td>Dicistroviridae</td>
<td>1017</td>
<td>7</td>
</tr>
<tr>
<td>Lake Sinai virus (LSV)</td>
<td>Unclassified</td>
<td>97</td>
<td>1</td>
</tr>
<tr>
<td>Turnip ringspot virus (TuRSV)</td>
<td>Secoviridae</td>
<td>1302</td>
<td>14</td>
</tr>
</tbody>
</table>

Granberg et al. 2013 PLoS ONE
3. Honeybee Colonies with CCD Symptoms

Israel acute paralysis virus (IAPV)
  – Similar to strains in France.

Aphid lethal paralysis virus (ALPV)
  – Only recently recognized to infect bees.

New strain of Lake Sinai virus (LSV)
  – Prevalent in the USA but not previously detected in other geographical areas

Turnip ringspot virus
  – An infectious viral agent of plants
  – Bees as vectors of pollen-borne viruses

First metagenomic study on a honeybee population outside of North America
Our Findings & Current projects

1. Pigs with Postweaning Multisystemic Wasting Syndrome (PMWS)
2. Shaking Mink Syndrome
3. Honeybees with Colony Collapse Disorder (CCD) Symptom
4. Full-length sequencing of African swine fever virus (ASFV)
4. ASFV - Selective PCR

Selective amplification: 15 primer pairs
Sequencing platform: MiSEQ

40% of the genome
4. Full genome comparison: ASFV in Sardinia
Summary

HTS-based approaches offer the possibility to identify any potential pathogen:

- Multifactorial diseases and co-infections
- Diseases with unknown etiology
- Vector organisms and reservoirs
- New and divergent pathogens (Identify and genetically characterize)
- Full genome sequence of multiple strains (including reference strains) of a single agent – correlate genotype to phenotypes - evolutionary dynamics

*Improved diagnosis of emerging or re-emerging diseases with known or unknown aetiology*

Challenges:

- HTS generates large amounts of data
- Storage and bioinformatics analysis
- Validate sequencing approaches for diagnostic use

HTS have the potential to open new scenarios for diagnosis, control and investigation of infectious diseases.
Acknowledgments

For the collaboration and for providing interesting samples:
Profs P. Wallgren, C. Fossum et al, Uppsala, Sweden
Dr. J. Benyeda, Mohács & Dr. Á. Bálint, Budapest, Hungary
Dr. Gian Mario De Mia & Claudia Torresi, IZS-UM, Perugia, Italy
Prof. Lars Erik Larsen, National Veterinary Institute, Technical University of Denmark
Dr. Charles Masembe, Makerere University Kampala, Uganda

For technical support, advices: Dr. M. Hakhverdyan and Dr. M. Leijon, Uppsala, Sweden

For assistance with the analysis pipeline: SLU Global Bioinformatics Centre, Uppsala, Sweden

For exchanging ideas in metagenomics and in veterinary medical infection-biology:
Prof. W.I. Lipkin, New York, USA
Prof. G.J. Viljoen, IAEA, Vienna, Austria
Prof. M.C. Horzinek, Bilthoven, The Netherlands
Dr. J.F. Valarcher, Uppsala, Sweden

For direct participation in this work and presentation:
Prof. Sándor Belák, Prof. Mikael Berg, Dr. Anne-Lie Blomström,
Dr. Maja Malmberg and Oskar Karlsson, Uppsala, Sweden
Collaborations

Genomic/Sequencing Technology Platforms

- SNP&SEQ Technology Platform
- Uppsala Genome Center

(UPPmax Next generation sequencing Cluster & Storage)

- Bioinformatics Resources
- The Kalkyl cluster, 2.784 powerful 64-bit processor cores

SLU Global Bioinformatics Centre

- Bioinformatics Resources
- Development of bioinformatics pipelines
Research Projects
Connected to High-Throughput Sequencing & Metagenomics

**AniBioThreat**
Bio-preparedness: Security/Safety/Bridging research

**PhD-project:** “Development of viral metagenomics for increased preparedness against infectious disease”

**Epi-SEQ**
(2015) Molecular epidemiology of epizootic diseases using next generation sequencing technology

- **FORMAS**
  - Viral metagenomics and bioinformatics as powerful tools in veterinary infection biology

- **BioBridges**
  - Next generation microfluidic pathogen detection platforms

- **RAPIDIA-FIELD**
  - Development of field test for rapid screening
THANK YOU FOR YOUR ATTENTION