How bioinformatics impact plant movement across border lines

Ioannis E. Tzanetakis
What it takes to provide producers with quality propagation material

• A breeding selection may be evaluated for 7-10 years in the open field

• Breeding fields tend to be pathogen ‘magnets’

• Clonally propagated crops: 500,000,000 plants of a particular cultivar (e.g. ‘Fuji’ apple) come from a single breeding selection

• All 500,000,000 plants can be infected with the same systemic pathogen as the mother plant

• So, to avoid the propagation of pathogens we do...
Virus Elimination

• Testing

• Virus eradication

• Testing

• Maintenance of Foundation Material

• Distribution to Propagators

• Repeat testing at intervals including tests for new viruses
Diagnostics
Laboratory and Greenhouse Testing
Virus eradication

- Prepare plants for treatment
- Heat, cryo- or chemo- therapy
- Meristem isolation
Virus eradication

- Prepare plants for treatment

- Heat, cryo- or chemo- therapy

- Meristem isolation

- Grow whole plant in tissue culture and in greenhouse

- Testing to confirm virus elimination was successful and repeat testing at intervals including tests for new viruses

- 2-4 years!!!!!
High throughput sequencing in detection and discovery of systemic plant pathogens

- Emergence of new diseases
- Efficiency of high throughput sequencing
- Analysis feasible through high performance computing
- Automated tools needed for non-professionals
Library preparation

- sRNA
- **total RNA (rRNA -)**
- dsRNA
- particle-enriched material
Platform outputs

- sRNA
- total RNA (rRNA -)
- dsRNA
- particle-enriched material

<table>
<thead>
<tr>
<th>Platform</th>
<th>Name</th>
<th>Run-time</th>
<th>Max length/read</th>
<th>Output (Gb)/run</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illumina</td>
<td>MiSeq</td>
<td>5-55 h</td>
<td>2 x 300 bp</td>
<td>0.3-13</td>
</tr>
<tr>
<td></td>
<td>HiSeq</td>
<td>10 h-11 d</td>
<td>2 x 150 bp</td>
<td>15-500</td>
</tr>
<tr>
<td>NextSeq</td>
<td></td>
<td>11-30 h</td>
<td>2 x 150 bp</td>
<td>19-120</td>
</tr>
<tr>
<td>HiSeqX</td>
<td></td>
<td>3 d</td>
<td>2 x 150 bp</td>
<td>1,800</td>
</tr>
<tr>
<td>NovaSeq</td>
<td></td>
<td>48 h</td>
<td>2 x 150 bp</td>
<td><strong>6,000</strong></td>
</tr>
<tr>
<td>Ion Torrent</td>
<td>PGM</td>
<td>3-7 h</td>
<td>400 bp</td>
<td>0.09-1.9</td>
</tr>
<tr>
<td></td>
<td>Proton</td>
<td>4-6 h</td>
<td>500 bp</td>
<td>12-88</td>
</tr>
<tr>
<td></td>
<td>S5</td>
<td>2.5-4 h</td>
<td>400 bp</td>
<td>2-16</td>
</tr>
<tr>
<td></td>
<td>RSII</td>
<td>2 h</td>
<td>3,000 bp</td>
<td>0.09</td>
</tr>
<tr>
<td>PacBio</td>
<td>Sequel</td>
<td>0.5-6 h</td>
<td><strong>20,000 bp</strong></td>
<td>0.08-1.25</td>
</tr>
<tr>
<td>Oxford Nanopore</td>
<td>MinION</td>
<td>1 min-48 h</td>
<td>10,000 bp$</td>
<td>44</td>
</tr>
</tbody>
</table>

Adapted from *Int. J. Mol. Sci.* 2017, 18, 308
# Bioinformatics pipelines

<table>
<thead>
<tr>
<th>Pipeline</th>
<th>Detection of known viruses</th>
<th>New virus discovery</th>
<th>Remarks</th>
<th>Reference</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linux based:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIP</td>
<td>yes</td>
<td>yes</td>
<td>could be used for plant virus detection</td>
<td>Lin et al., 2016</td>
<td><a href="https://github.com/keylabivdc/VIP">https://github.com/keylabivdc/VIP</a></td>
</tr>
<tr>
<td>VirusHunter</td>
<td>yes</td>
<td>yes</td>
<td>could be used for plant virus detection</td>
<td>Zhao et al., 2013</td>
<td><a href="http://www.ibridgenetwork.org/wustl/virushunter">http://www.ibridgenetwork.org/wustl/virushunter</a></td>
</tr>
<tr>
<td>VirusSeeker</td>
<td>yes</td>
<td>yes</td>
<td>could be used for plant virus detection</td>
<td>Zhao et al., 2017</td>
<td><a href="https://wupathlabs.wustl.edu/virusseeker">https://wupathlabs.wustl.edu/virusseeker</a></td>
</tr>
<tr>
<td>VirusFinder 2 with VERSE algorithm</td>
<td>yes</td>
<td>yes</td>
<td>could be used for plant virus detection</td>
<td>Wang et al., 2015</td>
<td><a href="https://bioinfo.uth.edu/VirusFinder/">https://bioinfo.uth.edu/VirusFinder/</a></td>
</tr>
<tr>
<td>Virtool</td>
<td>yes</td>
<td>yes</td>
<td>tested for plant virus detection</td>
<td>Rott et. Al., 2017</td>
<td><a href="https://www.virtool.ca/">https://www.virtool.ca/</a></td>
</tr>
<tr>
<td>Truffle</td>
<td>yes</td>
<td>no</td>
<td>tested for plant virus detection; e-probe based detection; has GUI</td>
<td>Visser et al., 2016</td>
<td><a href="https://sourceforge.net/projects/truffle/">https://sourceforge.net/projects/truffle/</a></td>
</tr>
<tr>
<td><strong>Web-based GUI:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VirusDetect</td>
<td>yes</td>
<td>yes</td>
<td>tested for plant virus detection</td>
<td>Zheng et al., 2017</td>
<td><a href="http://virusdetect.feilab.net/cgi-bin/virusdetect/index.cgi">http://virusdetect.feilab.net/cgi-bin/virusdetect/index.cgi</a></td>
</tr>
<tr>
<td>VSD toolkit</td>
<td>yes</td>
<td>yes</td>
<td>tested for plant virus detection; needs open source internet-based analytical environment called Yabi (Hunter et al., 2017)</td>
<td>Barrero et al., 2017</td>
<td><a href="https://github.com/muccg/yabi">Source code for Yabi: https://github.com/muccg/yabi</a></td>
</tr>
<tr>
<td>VirusTAP</td>
<td>yes</td>
<td>no</td>
<td>could be used for plant virus detection</td>
<td>Yamashita et al., 2016</td>
<td><a href="https://gph.niid.go.jp/cgi-bin/virustap/index.cgi">https://gph.niid.go.jp/cgi-bin/virustap/index.cgi</a></td>
</tr>
</tbody>
</table>

Villamor et al., *Phytopathology*, in press
**Case study: Elderberry diseases**

<table>
<thead>
<tr>
<th></th>
<th>Plant species</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>Sambucus nigra</em></td>
<td>Golden</td>
</tr>
<tr>
<td>2</td>
<td><em>S. nigra</em> subsp. <em>canadensis</em></td>
<td>Hwy O</td>
</tr>
<tr>
<td>3</td>
<td><em>S. nigra</em> subsp. <em>canadensis</em></td>
<td>Maxima</td>
</tr>
<tr>
<td>4</td>
<td><em>Sambucus racemosa</em></td>
<td>G-17321</td>
</tr>
<tr>
<td>5</td>
<td><em>S. racemosa</em> subsp. <em>sibirica</em></td>
<td>Tashkent #2149</td>
</tr>
</tbody>
</table>
Virfind pipeline

Input files:
- fasta
- fastq
- sff

Output files:
- Mapped_reads_host.fna
- Blastn_NON_VIRUS_reads.fna
- Blastn_NON_VIRUS_report.tab
- Blastn_VIRUS_reads.fna
- Blastn_VIRUS_report.tab
- Blastx_VIRUS_reads.fna
- Blastx_VIRUS_report.tab
- Reads_with_NO_Blastn_NO_Blastx.fna
  
- translate to amino acid
- Reads_with_NO_Blastn_NO_Blastx.faa
  
- conserved domain search
- Conserved_domain_search_report.txt

Assembly

<table>
<thead>
<tr>
<th>Sample</th>
<th>Carlavirus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Golden</td>
<td>A B C</td>
</tr>
<tr>
<td>Hwy O</td>
<td>D</td>
</tr>
<tr>
<td>Maxima</td>
<td>A* B* C D</td>
</tr>
<tr>
<td>G-17321</td>
<td>C D</td>
</tr>
<tr>
<td>Tashkent #2149</td>
<td>C D</td>
</tr>
</tbody>
</table>
VIRUSES SHOW SIGNIFICANT INTER HOST DIVERSITIES – IMPLICATIONS FOR DIAGNOSTICS

<table>
<thead>
<tr>
<th>Sample</th>
<th>Carlavirus</th>
<th>Raw reads</th>
<th>Mismatches (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Golden</strong></td>
<td>A</td>
<td>4472</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>3259</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1579</td>
<td>1.6</td>
</tr>
<tr>
<td><strong>Hwy O</strong></td>
<td>D</td>
<td>33597</td>
<td>3.3</td>
</tr>
<tr>
<td><strong>Maxima</strong></td>
<td>A*</td>
<td>486</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>B*</td>
<td>907</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1678</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>647</td>
<td>4.0</td>
</tr>
<tr>
<td><strong>G-17321</strong></td>
<td>C</td>
<td>1620</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>981</td>
<td>3.9</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>235</td>
<td>1.9</td>
</tr>
<tr>
<td><strong>Tashkent #2149</strong></td>
<td>C</td>
<td>7140</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>780</td>
<td>4.5</td>
</tr>
</tbody>
</table>

**Maxima A***

GGTGTTTGCCCTCAAGCAGGTGGCCGTTCAAAACCTTGAGGTTCCCCAACCATTAATCTCGTGACCCGATTGAGGGG
..GA..CC..T.G.............AACC.....TC.AGA........GGTT..CC.T..ACT..TA...........A..**

**Golden A**

GGTGTTTGCCCTCAAGCAGGTGGCCGTTCAAAACCTTGAGGTTCCCCAACCATTAATCTCGTGACCCGATTGAGGGG
..GA..CC..T.G.............AACC.....TC.AGA........GGTT..CC.T..ACT..TA...........A..**

**Maxima B**

CTTCTCTACTGAAAACCGAGGGCTTACGTCATTGCTTTTCGAAAAGTGAGGTTAAAGAGCTGCTCAGAGATCACGTGCTGGA

**Golden B**

CTTCTCTACTGAAAACCGAGGGCTTACGTCATTGCTTTTCGAAAAGTGAGGTTAAAGAGCTGCTCAGAGATCACGTGCTGGA

Hosts for reference virus genomes

PCR diagnostics: isolates undetectable
Advantages

- Neutral tool – no biological bias
- Cost-effective
- User-friendly, requires little training and supervision
- Automated, able to process 24 or more barcoded samples in a single lane
- Adaptable: Able to discover any type of pathogen - Custom-made code for all pathogens of an individual host
Yet, we are missing a lot...

- Database depth and computational power
- Sample preparation strategies and construction of cDNA libraries
- Comparison of sequencing platforms suitability for reliable diagnostics
- Validation of complete work-flows (from extraction to data analysis) and comparison with existing technologies (Al Rwahnih et al. 2015, Rott et al., 2017)
- Assessment of economic benefits, implementation costs & savings to industry
- Development of standard operating procedures (SOPs) for harmonization and implementation – ring tests
HTS and its (potential) impact in plant movement across borders
Certification at the era of agricultural globalization

- Harmonization of standards using information from multiple programs (EPPO, NAPPO, etc)
Example: Berry certification Standards

A. Introduction
B. General Provisions
C. Scope
D. Common Definitions, Abbreviations and Acronyms
E. Program Responsibilities
   a. G1 Blueberry Plants
   b. Responsibilities of G2 Nurseries
   c. Responsibilities of G3 and G4 Nurseries
   d. Responsibilities of Participating State Departments of Agriculture
F. Eligibility Requirements
   a. G1 Block
   b. G2 Block
   c. G3 Block
   d. G4 Block
   e. Propagation by Tissue Culture
G. References and Websites for National Blueberry Certification Guidelines
Appendix  A. National Vaccinium Certification Pathogen List and recommended testing procedures

Appendix B. Site Requirements for National Vaccinium Certification Standard for Blueberries
Appendix C. Nursery Sanitation and Pest Management Guidelines for National Vaccinium Certification Standard
Appendix D. Sampling Protocols for National Vaccinium Certification Standard
Appendix E. Visual Inspection Guidelines for National Vaccinium Certification Standard
Appendix F. Blueberry Pests of Concern in Nursery Stock, Their Symptoms, Reported Distribution, Diagnostic Tests, and Nutrient Deficiency Symptoms
Appendix G. Protocols for Testing for National Vaccinium Certification Standards
Appendix H. Containerized Plants
Appendix I. Pest Management Plan
Appendix J. Nursery Field Map and Inventory
Appendix K. Documentation, Identification and Tagging
Appendix L. Forms and Labels
HTS and 21st Century Plant Certification

- Nursery certification guidelines:
  - Input from stakeholders, regulators and scientists
  - Developed based on current scientific knowledge
  - Systems-based approach that works and is manageable (as determined by pilot studies in several US states)

![Diagram showing the HTS certification levels: G1, G2, G3, G4, with details on material testing, propagation, and source for further propagation, leading to certification levels such as Elite, Nuclear, Pre-Elite, Foundation, Mother, Extra Super Elite, Pre-Basic, etc.](image)
Why? Regulatory agencies need to approve

Provisional release 6 months vs 4 years

Candidate G1 plant

- Negative for target pathogens
  - Next generation sequencing
  - Positive for target pathogens
  - Thermal therapy, meristem tip-culture, regeneration whole plants
  - Laboratory testing for NGS-detected targeted pathogen(s)

Bioassays

G1 plant enters certification scheme

Gergerich et al., 2015. Plant Disease
Idealized journey for plants to producers

1. International cultivar/selection
   - Controlled import permit
   - Treatment to eliminate pathogens

2. Clean Plant Center
   - Pathogen testing
     - Pathogen present?
       - Yes: Treatment to eliminate pathogens
       - No: Protected culture
         - Regular retesting

3. Distribution and maintenance in nurseries
   - In-state sales monitored by certification agency

4. Phytosanitary requirements met - Interstate & International shipments

Gergerich et al., 2015. Plant Disease
Acknowledgments