GMI WG4 Proficiency Testing Survey - July 2013

Help to shape the future GMI Proficiency Testing Program

This short survey seeks to identify end users and their general capabilities that wish to participate in the GMI administered proficiency testing (PT) program for whole genome sequencing.

GMI is currently an informal global, visionary taskforce of scientists and other stakeholders who shares an aim of making novel genomic technologies and informatics tools available for improved global patient diagnostics, surveillance and research, by developing needs- and end-user-based data exchange and analysis tools for characterization of all microbial organisms and microbial communities.

A working group has been established to develop a PT system for inter-laboratory test performance to ensure harmonization and standardization in whole genome sequencing and data analysis, with the aim to produce comparable data for the GMI initiative. The intention is to include three optional components in the PT: I) DNA extraction, purification, sequencing and analysis II) sequencing and analysis of purified DNA, and III) analysis of datasets from NGS platforms. With this survey we are taking the important first step towards further developing the system.

As a participant in this survey, your needs and capacity are taken into consideration when creating the PT which by then will allow participants to ensure the quality of their DNA preparation, sequencing, and analysis (e.g. phylogeny) based on standardized testing.

If you have any questions or feedback for this survey, please contact Susanne Karlsmose (suska@food.dtu.dk) at the Technical University of Denmark.

Note: An asterisk (*) indicates a question that requires an answer.

GMI WG4 Proficiency Testing Survey - July 2013

End Users Section:

NOTE: Contact information submitted in this survey will only be used for initiating the GMI proficiency test.

A) About you and your organisation (if multiple groups or individuals from your organisation wish to participate in the proficiency test, each of them must fill out the online survey)

*1. Organisation name

*2. Which sector does your organisation belong to? (please select all that apply)

- [ ] Public Health
- [ ] Animal
- [ ] Plant/environment
- [ ] Food
- [ ] University
- [ ] Research
- [ ] Governmental
- [ ] Private

Comment

*3. What is your principle role? (please select all that apply)

- [ ] Clinician
- [ ] Hospital scientist/laboratory scientist/microbiologist
- [ ] Infection control practitioner/nurse
- [ ] Public Health professional/epidemiologist
- [ ] Researcher/academic
- [ ] Bioinformatician
- [ ] Postgraduate student
- [ ] Other

Other, please specify

*4. Point of contact for GMI proficiency testing

Contact person:	
Address:	
City:	
State/Province:	
Postal code:	
Country:	
Email:	
Phone number (incl. country code):	

*5. About your organisation #1
(please select one answer per row)

	Yes	No	Not relevant
5.1 Is your organisation able to ship and receive isolates (UN3373)?	◯	◯	◯
5.2 Is your organisation able to ship and receive DNA samples?	◯	◯	◯
5.3 Does your organisation have an FTP server to send and receive genomic data?	◯	◯	◯

*6. About your organisation #2
(please select one answer per row)

	Internally	Externally	Both	Not relevant
6.1 Does your organisation perform NGS internally or are samples shipped to an external NGS provider?	◯	◯	◯	◯
6.2 Does your organisation perform bioinformatics analysis of NGS data internally or externally?	◯	◯	◯	◯

GMI WG4 Proficiency Testing Survey - July 2013

End Users Section:

B) About your staff

7. Point of contact for NGS (practical sequencing)

Name:	
Address (if different from organsisation):	
City/Town:	
State/Province:	
ZIP/Postal Code:	
Country:	
Email Address:	
Phone Number:	

8. Point of contact for bioinformatics (data analysis)

Name:	
Address (if different from organisation):	
City/Town:	
State/Province:	
ZIP/Postal Code:	
Country:	
Email Address:	
Phone Number:	

GMI WG4 Proficiency Testing Survey - July 2013

End Users Section:

C) Capabilities

9. Please indicate which sequencing platform(s) you have access to and intend to use in the PT exercise (please select all that apply)

	We have access to this sequencing platform	We have this resource in-house	This resource is external to our organisation	We intend to use this platform in the PT exercise
Ion Torrent PGM	☐	☐	☐	☐
Ion Torrent Proton	☐	☐	☐	☐
GS Junior System (454)	☐	☐	☐	☐
Genome Sequencer FLX (454)	☐	☐	☐	☐
PacBio RS	☐	☐	☐	☐
PacBio RS II	☐	☐	☐	☐
HiScanSQ	☐	☐	☐	☐
HiSeq 1000	☐	☐	☐	☐
HISeq 1500	☐	☐	☐	☐
HiSeq 2000	☐	☐	☐	☐
HiSeq 2500	☐	☐	☐	☐
Genome Analyzer Iix	☐	☐	☐	☐
MiSeq Benchtop Sequencer	☐	☐	☐	☐
ABI SOLiD	☐	☐	☐	☐
other	☐	☐	☐	☐

If other, please specify

10. When sequencing a single bacterial 5MB genome at 20X coverage, what are the costs and how many do you annually sequence for each platform you use?
Note: 1) use the highest multiplexing capability for this size genome that your group routinely applies for whole genome variant analysis, and 2) only the platforms used needs to be addressed

	Cost when handled internally (including DNA prep and library, but excluding staff cost)	Number of genomes annually sequenced internally	Cost when handled externally (including DNA prep and library, but excluding staff cost)	Number of genomes annually sequenced externally
Ion Torrent PGM				
Ion Torrent Proton				
GS Junior System (454)				
Genome Sequencer FLX (454)				
PacBio RS				
PacBio RS II				
HiScanSQ				
HiSeq 1000				
HISeq 1500				
HiSeq 2000				
HiSeq 2500				
Genome Analyzer Iix				
MiSeq Benchtop Sequencer				
ABI SOLiD				
other				

If other, please specify

GMI WG4 Proficiency Testing Survey - July 2013

Target Organisms Section

A) Focal Pathogens:

11. Please list up to five MOST FREQUENTLY INVESTIGATED pathogens in your laboratory

1.
2.
3.
4.
5.

12. Please list up to five pathogens whose genomes have been MOST FREQUENTLY SEQUENCED and/or analysed by your team in the last 12 months

1.
2.
3.
4.
5.

GMI WG4 Proficiency Testing Survey - July 2013

Target Organisms Section

B) Criteria for selecting the target pathogens for GMI PT

*13. Why do you use NGS?

Please select the top five criteria in order of importance for the potential use of NGS to your team (indicate at least one criterion) (1 - most important; 5 - least important)

13.1 Identification of pathogens or taxonomical studies	
13.2 Discovery of new pathogens	
13.3 Monitoring of emerging clones/variants	
13.4 Microbiome and/or metagenomic analysis	
13.5 Characterization of antimicrobial drug resistance	
13.6 Characterization of virulence factors or virulence determinants	
13.7 High resolution clustering for outbreak investigations	
13.8 Tracing local or global transmission chains	
13.9 Sustaining the competitive edge of the laboratory	
13.10 Opportunities for building new collaborations	
13.11 Potential to answer important questions in microbial evolution	
13.12 Other, please specify	

14. If 'other' in the question above, please specify the potential use of NGS to your team

*15. How do you select pathogens?

Please select the top five criteria in order of importance for selecting pathogens for WGS-projects (indicate at least one criterion) (1 - most important; 5 - least important)

15.1 Utility for near real time prospective laboratory surveillance	
15.2 Availability of clinical and epidemiological meta-data for isolates	
15.3 Discovery of new biomarkers and drug targets	
15.4 Testing of WGS throughput and analysis capability of your team	
15.5 Availability of funds for the project	
15.6 Availability of quality reference sequences	
15.7 High impact on public health	
15.8 Compliance with biosecurity and biosafety regulations	
15.9 Provides a system within which to address novel scientific questions	
15.10 Other, please specify	

16. If 'other' in the question above, please specify the criteria for selecting pathogens for WGS-projects

***17. To what extent do you agree or disagree with the following statements about evaluation criteria for the GMI PT program? (please select one answer per row)**

	Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree
17.1 Assessment of the quality of WGS reads is a very important consideration	◯	◯	◯	◯	◯
17.2 Ability to integrate and accommodate sequence data from multiple vendor platforms is a very important consideration	◯	◯	◯	◯	◯
17.3 Capacity for de novo sequencing and genome assembly is a very important consideration	◯	◯	◯	◯	◯
17.4 Capacity for analysis of emerging biothreats is a very important consideration	◯	◯	◯	◯	◯
17.5 Accurate classification of existing frequently tested and globally relevant pathogens (e.g., foodborne Salmonella) is a very important consideration	◯	◯	◯	◯	◯
17.6 Quality of reference based assembly is a very important consideration	◯	◯	◯	◯	◯
17.7 Quality of annotation is a very important consideration	◯	◯	◯	◯	◯
17.8 Single nucleotide polymorphism (SNP) calls is a very important consideration	◯	◯	◯	◯	◯
17.9 Tree building is a very important consideration	◯	◯	◯	◯	◯

Target Organisms Section

C) Priority pathogens

18. For the PT-components including 1) DNA extraction, purification, sequencing and data analysis and 2) sequencing and analysis of purified DNA, which microorganisms (bacterial, viral, fungal or protozoan) would in your opinion be the most important to include? (please list up to five: 1 - most important; 5 - least important)

1.
2.
3.
4.
5.

19. For the PT-components including analysis of datasets from an NGS platform for testing of bioinformatics capabilities, which microorganisms (bacterial, viral, fungal or protozoan) would in your opinion be the most important to include? (please list up to five: 1 - most important; 5 - least important)

1.
2.
3.
4.
5.

20. If participating in a future PT, which number of strains per dispatch would be appropriate for you?
(please select one answer per row)

	1	2	3	4	Not applicable
24.1 Microorganisms (bacterial) for DNA purification and sequencing	◯	◯	◯	◯	◯
24.2 Microorganisms (viral) for DNA purification and sequencing	◯	◯	◯	◯	◯
24.3 Microorganisms (fungal) for DNA purification and sequencing	◯	◯	◯	◯	◯
24.4 Microorganisms (protozoan) for DNA purification and sequencing	◯	◯	◯	◯	◯
24.5 Samples of DNA for sequencing	◯	◯	◯	◯	◯
24.6 Datasets from NGS platforms for data analysis	◯	◯	◯	◯	◯

GMI WG4 Proficiency Testing Survey - July 2013

Quality Assessment Section

A) Objectives, sample preparation and targets

***21. Which of the following describe how you use/intend to use NGS data? (please select all that apply)**

- [ ] De novo sequencing
- [ ] Resequencing
- [ ] Metagenomics
- [ ] Amplicon
- [ ] RNA-seq
- [ ] Other

Other, please specify

***22. If you are currently performing NGS runs, how are you preparing your libraries? (please select all that apply)**

- [ ] Physical shearing
- [ ] Enzymatic shearing
- [ ] Transposon based fragmentation
- [ ] Not relevant (we are not performing practical NGS)

***23. Do you multiplex samples for NGS runs? (please select one answer)**

- ◯ Yes
- ◯ No
- ◯ Not relevant (we are not performing practical NGS)

24. To which coverage level do you usually sequence? (please select one answer per row)

	less than 10	11-30	31-60	over 60	Not applicable
Bacteria	◯	◯	◯	◯	◯
Virus	◯	◯	◯	◯	◯
Fungus	◯	◯	◯	◯	◯
Protozoa	◯	◯	◯	◯	◯

***25. Do you routinely include standard/reference material on NGS runs? (please select one answer)**

- ◯ Yes
- ◯ No
- ◯ Not relevant (we are not performing practical NGS)

***26. Which type(s) of markers/information do you intend do capture with NGS? (please select all that apply)**

- ☐ Physical shearing (size of the inserts and fragments used to construct libraries)
- ☐ Insertion and deletions (i.e. Indels)
- ☐ Copy # variants (changes in gene copy #, changes in # of tandem repeats)
- ☐ Locus specific variants (presence/absence of specific alleles)
- ☐ Rearrangements (structural changes that affect the order of loci)
- ☐ Mobile elements (phages and plasmids)
- ☐ Single nucleotide polymorphisms (SNPs)
- ☐ Other

Other, please specify

GMI WG4 Proficiency Testing Survey - July 2013

Quality Assessment Section

B) Post processing, QC/QA, and analyses

27. Do you trim low quality bases from NGS reads prior to downstream analyses?

If yes, please describe the criteria and software you use:

If no, why do you not trim?

*28. How important do you think quality filtering is to NGS analyses? (please select one answer)

- ◯ Very
- ◯ Somewhat
- ◯ Negligible
- ◯ Not sure

*29. Do you have established criteria for quality assessment and quality control of assemblies based on NGS data? (please select one answer)

- ◯ Yes
- ◯ No
- ◯ I do not assemble NGS data for analyses

30. If you answered 'yes' in Q29, which of the following criteria do you use? (please select all that apply)

- ☐ Number of bases
- ☐ Number of contigs
- ☐ N50
- ☐ Coverage
- ☐ % reads mapped to reference
- ☐ Annotation success
- ☐ Other

If other, please specify

31. For the parameters you check marked in Q30, which value have you applied and why?
(please base answers on a 5MB genome)

Number of bases	
Number of contigs	
N50	
Coverage	
% reads mapped to reference	
Annotation success	
Other, please specify	

32. How would you describe your group's use of software for NGS analysis?
(please select one answer)

- ◯ Others perform analyses
- ◯ Only use externally developed software
- ◯ Majority externally developed software but some in-house software is also used
- ◯ Majority internally developed software but some external software is also used

33. Which assembly software packages are you currently using for NGS data analysis?
(please select all that apply)

- ☐ ABySS
- ☐ ALLPATHS-LG
- ☐ CABOG
- ☐ CLC
- ☐ Edena
- ☐ Euler
- ☐ Mira
- ☐ MSR-CA
- ☐ Newbler
- ☐ SGA
- ☐ SOAPdenovo
- ☐ SSAKE
- ☐ VCAKE
- ☐ Velvet
- ☐ Other

If other, please specify

34. Which mapping software packages are you currently using for NGS data analysis? (please select all that apply)

- [ ] bfast
- [ ] Bowtie 1
- [ ] Bowtie 2
- [ ] BWA
- [ ] ELAND
- [ ] MAQ
- [ ] Novoalign
- [ ] SHRiMP
- [ ] ssaha2
- [ ] Other

If other, please specify

35. Which additional software packages are you currently using for NGS data analysis? In the text boxes below, please insert purpose, name and URL (e.g. detection of resistance genes, Resfinder, http://cge.cbs.dtu.dk/services/ResFinder/, open access)

1. Purpose, name, URL/in-house:
2. Purpose, name, URL/in-house:
3. Purpose, name, URL/in-house:
4. Purpose, name, URL/in-house:
5. Purpose, name, URL/in-house:
6. Purpose, name, URL/in-house:
7. Purpose, name, URL/in-house:
8. Purpose, name, URL/in-house:
9. Purpose, name, URL/in-house:
10. Purpose, name, URL/in-house: