







PROTOCOL for GMI Proficiency Test, 2015

1	OVERVIEW AND OBJECTIVES	1
2	INTRODUCTION	2
3	OUTLINE OF THE GMI PT	3
3.1	Shipping, receipt and storage of bacterial strains3	
3.2	Using FTP to transfer files3	
3.3	Supplied test material3	
3.4	Procedure and analysis of test material4	
4	DISCUSSION FORUM	5
5	REPORTING OF RESULTS AND EVALUATION	5
	APPENDICES	
		_

HISTORY OF CHANGES; version 2
Reference for *S. enterica* ST00025 corrected to STA00025

1 OVERVIEW AND OBJECTIVES

The proficiency test, 2015, consists of three general parts:

- 1a. DNA extraction, purification, library-preparation, and whole-genome-sequencing from **live cultures**
- 1b. Whole-genome-sequencing of pre-prepared DNA
- 2. Phylogenetic/clustering analysis of three fastq datasets

The main **objective** of this proficiency test is to quantify differences among laboratories in order to facilitate the development of reliable laboratory results of consistently good quality within the area of DNA preparation, sequencing, and analysis (e.g. phylogeny). This ensures that the discrepancies and differences among laboratories are known and will contribute to the standardization of whole genome sequencing and data analysis, with the aim to produce comparable data for the GMI initiative. A further objective is to assess and improve the uploaded data to databases such as NCBI, EBI and DDBJ.



The GMI proficiency test 2015 is supported by COMPARE, which has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 643476. In addition, the GMI PT 2015 is supported by the GenomeTrakr Network and Microbiologics[®].

2 INTRODUCTION

GMI is a global, visionary taskforce of scientists and other stakeholders who share an aim of applying novel genomic technologies and informatics tools to improve 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.

The GMI working group 4 (WG4) steered by the US FDA, Microbiologics, and Technical University of Denmark has prepared this proficiency test (PT). The PT consists of three parts, each of which are optional, and include assessing (1a) the laboratory's DNA preparation and sequencing procedures, (1b) the laboratory's sequencing output, and (2) the laboratory's procedure to identify variant sites within whole genome sequence data and cluster and distinguish samples based on those variants.

The proficiency test focuses on *Salmonella enterica*, *Escherichia coli* strain and *Staphylococcus aureus*, and allows for sign-up for each species separately. Note that item 1a and item 1b are parallel; i.e. when signing up for 1a for one species, the participation in 1b is connected.

The three items consist of

- 1a) DNA extraction, purification, library-preparation, and whole-genome-sequencing of six bacterial cultures: two *Salmonella enterica* strains, two *Escherichia coli* strain and two *Staphylococcus aureus* strains. Participants will be requested to upload reads to an ftp-site and **optionally** also identify the Multi Locus Sequence Type (MLST) of the strains as well as the resistance genes present in the strains if that is something that is routinely done within the laboratory.
- 1b) Whole-genome-sequencing of pre-prepared DNA delivered by GMI Working Group 4 of the same six bacterial strains mentioned in clause 1a.
- 2) Variant detection and phylogenetic/clustering analysis of three datasets each including fastq data from circa 20 genomes of *S. enterica*, *E. coli* and *S. aureus*. Note: If performing a reference based approach for variant detection, the reference applied for the analysis must be the species specific references indicated below (see 3.4.2).



Institutes/organizations which signed up to participate will receive the PT-material (bacterial strains, DNA and/or the login for download of datasets) according to the registered sign-up information.

3 OUTLINE OF THE GMI PT

3.1 Shipping, receipt and storage of bacterial strains

In August 2015, around 100 laboratories located worldwide will receive a parcel containing the two *Salmonella enterica* strains, two *E. coli* strains and two *S. aureus* strains together with corresponding purified DNA (according to the registered sign-up information). All bacterial strains and DNA are shipped as UN3373, Biological substance category B. Those who signed up for item 2 (phylogenetic analysis) will receive information and login for downloading the three datasets.

Please confirm receipt of the parcel through the confirmation form enclosed in the shipment.

The bacterial strains are shipped lyophilised as KwikStik's (see below for additional info on handling). On arrival, the KwikStik's must be refrigerated until handling in the laboratory.

The bacterial DNA is shipped as dried samples using a DNA stabilizing agent (DNAstable® *Plus*, Biomatrica). On arrival, either rehydrate your sample and store the liquid samples at room temperature in closed tubes, to prevent evaporation. Or store the dried samples in either

- (a) a dry storage cabinet at room temperature (15-25°C or 59-77°F) or
- (b) a heat-sealed, moisture-barrier bag along with a silica gel desiccant pack.

3.2 Using FTP to transfer files

For download of fastq files for item 2 and for upload of results, an ftp-server is used. The proficiency test organizer will provide each participant with username and login for this purpose. The ftp-site which will be used for this purpose is cgebase.cbs.dtu.dk. For information on how to transfer files, please see Appendix 1.

3.3 Supplied test material

3.3.1 Item 1a; Bacterial cultures

The procedure for reconstitution of the bacterial cultures should follow the manufacturer's procedures as presented in the instructional video or the written instructions on their website (see http://microbiologics.com/s.nl/sc.7/category.98564/.f or http://microbiologics.com/Support-Center/KWIK-STIK-trade).

MSDS for KwikStik are found here: http://microbiologics.com/Support-Center/Lyophilized-Microorganism-Preparations.



The bacterial cultures supplied have been sequenced multiple times and the genomes have been closed. Therefore, the PT-organizers encourage participants to maintain these bacterial strains in their strain collection and apply them as part of future internal quality control.

3.3.2 Item 1b; DNA

The supplied DNA has been stabilized by DNA Stable®plus (http://www.biomatrica.com/media/dnastable%20Plus/3004-0112.pdf). Each vial contains a minimum of 2 ug DNA. Before use, the samples should be re-suspended in 60-100 µl water or aqueous buffer and mixed by gentle pipetting or vortexing for 10 min (according to above mentioned protocol). Rehydrated samples can be stored at room temperature and used directly in downstream application.

3.3.3 Item 2; Fastq data set

Three datasets, one for each of *S. enterica*, *E. coli* and *S. aureus*, will be available for download from the ftp-site 'cgebase.cbs.dtu.dk'. Login to the ftp-site will be provided directly to each participant. Each dataset will consist of the original fastq files (i.e., whole genome sequence data) from circa 20 samples for phylogenetic cluster analysis based on a tool of the laboratory's own choice; SNP-calling, gene-by-gene, etc.

3.4 Procedure and analysis of test material

3.4.1 Item 1a and 1b; Bacterial cultures and DNA

Subculture the bacterial strains on a relevant growth medium of the laboratory's own choice and incubate. Following incubation and assessment of purity of the bacterial cultures, perform DNA extraction and whole-genome-sequencing according to the laboratory's standard procedure.

For the purified PT-DNA received, perform whole-genome-sequencing according to the laboratory's standard procedure.

For both bacterial cultures and DNA (items 1a and 1b), register relevant information related to the methods applied via https://www.surveymonkey.com/r/PT 2015 bacterial cultures and DNA (also see Appendix 2). Appendix 2 also describes the requested results when analyzing the sequences as regards the detected antimicrobial resistance genes and as regards the Multi Locus Sequence Type of the bacterial strain.

3.4.2 Item 2; Fastq data set

The three fastq datasets should be downloaded from the ftp-site. They are organized into three different .zip archives appropriately labeled with the taxon they represent. Within each archive



the participant will find the paired-end reads. The objective associated with this dataset is to assess the variability of laboratories in the clusters identified through the analysis of next-generation sequencing data. As such, the participant should employ their preferred method for constructing a matrix (e.g., gene, SNP, presence/absence, etc.) and for clustering samples (e.g., distance-, maximum-likelihood-, Bayesian-based).

If performing a reference based approach for variant detection, the reference applied for the analysis must be: STA00025 (*S. enterica*), EC002143 (*E. coli*) and SAH596 (*S. aureus*).

4 DISCUSSION FORUM

A web-based discussion forum is available for participants in the GMI PT 2015, allowing for individual sign-up and discussion with other PT-participants in relation to issues relating to the analysis for the present PT. Appendix 4 presents detailed information on the PT discussion forum.

5 REPORTING OF RESULTS AND EVALUATION

For all items (1a, 1b and 2), the results should be captured and entered into the Internet-based survey (https://www.surveymonkey.com/r/PT 2015 FASTQ dataset). See also Appendix 2 and 3.

5.1 Procedure and analysis of test material

5.1.1 Item 1a and 1b; Bacterial cultures and DNA

Results for item 1a and 1b must be submitted as a batch-upload. The web-interface of the batch upload interface (https://cge.cbs.dtu.dk/services/ringtrials/) provides a possibility to upload several isolates in a single submission. The interface is divided into five steps, and a progress bar presents the overview of the submitted files.

Step 1; Download the Excel Metadata template to your computer.

<u>Step 2</u>; Fill in the required fields with all the relevant information (metadata) about the isolates, the associated WGS file names, sequencing platform and sequencing type used to generate the data, etc. The second tab of the spreadsheet has extended description of the required metadata. Note that **sample name** should be the **same as label-name of the sample**, e.g. **GMI15-003-BACT** or **GMI15-003-DNA**.

<u>Step 3</u>; When the spreadsheet is properly filled out, upload the file to the web-interface by clicking on the button "Upload Metadata File" and selecting the file from the file browser.

After this, the spreadsheet will be validated to check if the metadata has the correct format.



A valid spreadsheet will be validated showing the message 'Excel template uploaded correctly'.

If the spreadsheet contains invalid metadata, an error message will appear at the top of the uploader (above step 1) displaying the number of errors found. To expand the error messages, click on the "plus" (+) sign to the right of the messages. To fix the errors, correct the errors in the original spreadsheet and upload the updated spreadsheet.

<u>Step 4</u>; Go via the batch upload interface to upload the individual WGS files that were previously included as metadata in the spreadsheet. An additional validation step will check if the provided files can be found in the spreadsheet.

<u>Step 5</u>; Click on the Submit button to upload the files. The progress will be displayed in the Upload Progress bar. It is important to keep the window opened until the upload is completed. After that, the web-interface will automatically submit the job to the server.

If the job is submitted correctly, you will get an on-screen confirmation message. Additionally, you will be suggested to type in your e-mail to have the results sent to your inbox after the job is finished.

Via the Internet-based survey

(<u>https://www.surveymonkey.com/r/PT 2015 bacterial cultures and DNA</u>; see also Appendix 2), answers should be submitted to the questions related to the analysed bacterial cultures and DNA.

5.1.2 Item 2; Fastq data set

Specifically, up to two types of files should be submitted:

For each dataset:

- 1. The DNA sequence matrix used for clustering should be in fasta format and have that as the file extension
 - The matrix should only contain those samples provided through the ftp site
 - Syntax for the names of samples in the matrix should be *only* the prefix preceding the first underscore in the file name. For example, **STA00025_1.fastq** should be named **STA00025** in the matrix. Note all capital letters.
 - The file should be named as follows LabID_Taxon.fasta (e.g., LAB1_ST.FASTA, LAB1_EC.FASTA, or LAB1_SA.FASTA).
- 2. The clusters themselves in newick format with the .tre as the file extension
 - The tree should only contain those samples provided through the ftp site
 - Syntax for the names of samples in the matrix should be *only* the prefix preceding the first underscore in the file name. For example, **STA00025_1.fastq** should be named **STA00025** in the matrix. Note all capital letters.
 - The file should be named as follows LabID_Taxon.tre (e.g., LAB1_ST.TRE, LAB1_EC.TRE, or LAB1_SA.TRE).



Via the Internet-based survey (https://www.surveymonkey.com/r/PT 2015 FASTQ dataset; see also Appendix 3), answers should be submitted to the questions related to the Fastq data set section.

5.2 Evaluation of results

For both bacterial cultures and DNA (items 1a and 1b), the submitted sequence data (fastq-files) will be evaluated according to the following specific quality markers: e.g. read length (bp), N50 (bp), total number of contigs and total length of sequence (bp) including percentage of reference genome covered. In addition, the PT-organizers will assemble the submitted reads and compare these assemblies 1) towards the relevant closed genome to assess the sequence error rate and coverage of the scaffold and 2) between the obtained sequences in items 1a and 1b.

Assessment of the submitted results from the analysis of the fastq datasets (item 2) is based on two criteria: 1) the concordance among laboratories in their answers to the questions in the SurveyMonkey (Appendix 3) and 2) the concordance between participants' in the information content contained in the matrix and the relationships among samples from the clustering analyses (i.e., the topology).

For the evaluation of the results, no official GMI quality threshold is currently available and therefore no acceptance limit has been defined for this proficiency test.

5.3 Deadline for submission of results

Results must be submitted electronically **no later than October 9**th **2015**. Immediately after this date, the survey will be closed and results submitted to the Internet-based survey, via the batch-upload and to the ftp-site will be evaluated. Delayed submission of results will not be accepted.

5.4 Analysis and publication of results

Individual results will be anonymized, and only the PT-organizers will have access to your laboratory's results. Each participating laboratory will receive an individual summary of the obtained performance. An overall report summarizing the results will be published and subsequently in a peer-reviewed publication. Authors and co-authors of the publications will be those who have contributed to the preparation and execution of the proficiency test. Due to the anonymity of results, the individual participating laboratories will not be acknowledged in the publications.

We are looking forward to receiving your results.

If you have any questions or concerns, please do not hesitate to contact us:



Issues related to the dry-lab fastq datasets, please contact:

James Pettengill
U.S. Food and Drug Administration
Center for Food Safety and Applied Nutrition
CPK1 RM2D019
5100 Paint Branch Parkway
College Park, MD 20740, US

Tel: +1 240-402-1992

E-mail: <u>James.Pettengill@fda.hhs.gov</u>

In relation to other issues, e.g. organizational issues, please contact the EQAS Coordinator:

Susanne Karlsmose Pedersen National Food Institute, Technical University of Denmark Søltofts Plads, Building 221, room 238, DK-2800 Kgs. Lyngby, DENMARK

Tel: +45 3588 6601

E-mail: suska@food.dtu.dk

___ ___









PROTOCOL for GMI Proficiency Test, 2015 - APPENDICES

Appendix 1: Using FTP to transfer files

Appendix 2: Overview of Internet-based survey - bacterial cultures and DNA

Appendix 3: Overview of Internet-based survey - FASTQ dataset

Appendix 4: GMI Proficiency Test Forum guide









Appendix 1

Using FTP to transfer files

FTP is an acronym for File Transfer Protocol and is used to transfer files between computers on a network. To access the folder for upload or download of files, do as described below.

Obtain access to upload or download files by using the relevant login provided by the proficiency test organizer.

Using a Windows-computer:

Open the Documents folder, and type 'ftp://cgebase.cbs.dtu.dk/' in the Address bar. Enter you username and password, click "Log on".

Using a Mac-computer:

FileZilla FTP client:

- Download and install FileZilla (https://filezilla-project.org/)
- Host:cgebase.cbs.dtu.dk
- Type username and password
- Connect

Or

Finder Mac application:

- In the Finder, choose Go > "Connect to Server," and wait for the pop-up window to show up.
- Specify server address ftp://cgebase.cbs.dtu.dk and click "Connect"
- In the new pop-up window enter you username and password, click "Connect"

Introduction

This survey seeks to capture info on participants' sequence procedures and specifications in relation to the bacterial cultures and DNA tested as part of the GMI Proficiency Test (PT) 2015.

The survey consists of six sections, collecting information on

- 1. User Information and Sample Storage
- 2. Bacterial Culture; DNA Isolation, Handling and Processing
- 3. Received DNA; Handling and Processing
- 4. Sequencing
- 5. Analysis of sequences; MLST and antimicrobial resistance genes
- 6. Submitted datafiles

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

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

GMI is a global, visionary taskforce of scientists and other stakeholders who share an aim of applying novel genomic technologies and informatics tools to improve 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.

The GMI proficiency test 2015 is supported by COMPARE, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643476. In addition, the GMI PT 2015 is supported by the GenomeTrakr Network and Microbiologics®.

*1. Institute name / Organization	name
≭ 2. Department name	
*3. Name of person responsible	for the handling of the PT-material
or runne or person responsible	

	ling of the PT-material (date for upload of sequence
data)	
	DD MM YYYY
Date PT-material received	/ / /
Date processing the bacterial cultures started	1 1
Date processing the bacterial cultures completed	
Date processing the DNA started	1 1
Date processing the DNA completed (upload of	1 1

sequence data)

*!	5.		•	S	S	3	t	0	ra	19	је	C	or	d	ti	01	15	0	f t	th	e	ba	JC	te	ri	al	Cl	ıltı
pro																												
(ple	e	a	3	\$	5	S	e) (56	į	e	:t	OI	1e	a	ns	SV	ve	r)									
0		-8	8	3(3(0	°C	2																				
0		-2	2	2(2(0	°C	2																				
0		4	4°	۰	°(C)																					
0		F	₹(c	C	0	or	n '	te	np	er	atu	re															
0		٨	۷c	lc	С)	s	toı	ra	ge	tin	ne																
0		٧	Ν	/(/ (е	e d	bit	n	ot	re	cei	/e l	oac	teri	al	cul	ltur	es	for	thi	is P	РΤ					
0		C	Эt	t	t	th	ne	r																				
If ot	the	e	r,	٠,	,		pl	ea	ıse	e d	efi	ne																
*6	6.	·	•	Ş	S	S	it	0	ra	30	le	C	OI	ıd	iti	0	ns	S C	of '	th	ıe	D	N	A i	in	tł	he	tir
(ple																												
0		-8	8	3(3 (0	°C	2																				
0		-2	2	2(2(0	°C	2																				
0		4	1°	۰	°(C)																					
0		F	₹(c	c	0	or	n	te	np	er	atu	re															
0											tin																	
0							ne																					
If ot	the	ıeı	r.				pl	ea	ısı	e d	efi	ne																
				_	_	_																						
								_																				

BACTERIAL CULTURES received

*7. How were the bacteria	l cultures cultivated [as a decimal separator, please use full
stop (.)]:	
7.1 - Type of agar media/liquid broth:	
7.2 - Incubation time (hours):	
7.3 - Incubation temperature (°C):	
*8. For the Gram-negative	bacterial cultures; DNA extraction procedure (enter 'NA' if
not relevant):	,
8.1 - If manual extraction; kit used, full name:	
8.2 - If manual extraction; catalogue number of kit:	
8.3 - If manual extraction, modifications to kit protocol:	
8.4 - If automatic extraction; robot used:	
8.5 - If automatic extraction; specific protocol:	
8.6 - If automatic extraction; modifications to protocol:	
*9. For the Gram-positive	bacterial cultures; DNA extraction procedure (enter 'NA' if
not relevant):	
9.1 - If manual extraction; kit used, full name:	
9.2 - If manual extraction; catalogue number of kit:	
9.3 - If manual extraction, modifications to kit protocol:	
9.4 - If automatic extraction; robot used:	
9.5 - If automatic extraction; specific protocol:	
9.6 - If automatic extraction; modifications to protocol:	
10. For bacterial cultures, D	DNA concentration (ng/μl) prior to library preparation was
measured on (please select	t one answer)
O Qubit	
○ Nanodrop	
O DNA concentration not measured	
O Other	
If other, please specify:	

	ntration (ng/µl) [as a decimal separator, please us
[-)]	
11.1 GMI15-001-BACT (Salmonella)	
11.2 GMI15-002-BACT (Salmonella)	
11.3 GMI15-003-BACT (E. coli)	
11.4 GMI15-004-BACT (E. coli)	
11.5 GMI15-005-BACT (S. aureus)	
11.6 GMI15-006-BACT (S. aureus)	
12. Total DNA amount (micro	ram) [as a decimal separator, please use full stop
12.1 GMI15-001-BACT (Salmonella)	
12.2 GMI15-002-BACT (Salmonella)	
12.3 GMI15-003-BACT (E. coli)	
12.4 GMI15-004-BACT (E. coli)	
12.5 GMI15-005-BACT (S. aureus)	
12.6 GMI15-006-BACT (S. aureus) 13. For bacterial cultures, DN to library preparation was me	A quality (e.g. RIN, 260/280 ratio and/or 260/230 r
12.6 GMI15-006-BACT (S. aureus) 13. For bacterial cultures, DN to library preparation was me	
12.6 GMI15-006-BACT (S. aureus) 13. For bacterial cultures, DN to library preparation was me Bioanalyser Nanodrop DNA quality not measured	
12.6 GMI15-006-BACT (S. aureus) 13. For bacterial cultures, DN to library preparation was me	
12.6 GMI15-006-BACT (S. aureus) 13. For bacterial cultures, DN to library preparation was me	
12.6 GMI15-006-BACT (S. aureus) 13. For bacterial cultures, DN to library preparation was me Bioanalyser Nanodrop DNA quality not measured Other If other, please specify:	
12.6 GMI15-006-BACT (S. aureus) 13. For bacterial cultures, DN to library preparation was me Bioanalyser Nanodrop DNA quality not measured Other If other, please specify:	asured on (please select one answer)
12.6 GMI15-006-BACT (S. aureus) 13. For bacterial cultures, DN to library preparation was me Bioanalyser Nanodrop DNA quality not measured Other If other, please specify:	asured on (please select one answer)
12.6 GMI15-006-BACT (S. aureus) 13. For bacterial cultures, DN to library preparation was me Bioanalyser Nanodrop DNA quality not measured Other If other, please specify: 14. Measure of DNA quality (euse full stop (.)]	asured on (please select one answer)
12.6 GMI15-006-BACT (S. aureus) 13. For bacterial cultures, DN to library preparation was me C Bioanalyser C Nanodrop C DNA quality not measured C Other If other, please specify: 14. Measure of DNA quality (equipment) use full stop (.)]	asured on (please select one answer)
12.6 GMI15-006-BACT (S. aureus) 13. For bacterial cultures, DN to library preparation was me	asured on (please select one answer)
12.6 GMI15-006-BACT (S. aureus) 13. For bacterial cultures, DN to library preparation was me Bioanalyser Nanodrop DNA quality not measured Other If other, please specify: 14. Measure of DNA quality (euse full stop (.)] 14.1 GMI15-001-BACT (Salmonella) 14.2 GMI15-002-BACT (Salmonella) 14.3 GMI15-003-BACT (E. coli)	asured on (please select one answer)

GMI Proficiency Testing 201	5 - bacterial cultures and DNA
	ality (260/230 ratio) [as a decimal separator, please
use full stop (.)]	
15.1 GMI15-001-BACT (Salmonella)	
15.2 GMI15-002-BACT (Salmonella)	
15.3 GMI15-003-BACT (E. coli)	
15.4 GMI15-004-BACT (E. coli)	
15.5 GMI15-005-BACT (S. aureus)	
15.6 GMI15-006-BACT (S. aureus)	

DNA received

. For the DNA rece	ved, DNA conce	ntratio
easured on (please	select one answ	er)
Qubit		
Nanodrop		
DNA concentration not mea	sured	
Other		
ther, please specify:		
. Measure of DNA	oncentration (no	յ/μl) [a։
GMI15-001-DNA (Salmonella)	
2 GMI15-002-DNA (Salmonella)	
B GMI15-003-DNA (E. coli)		
GMI15-004-DNA (E. coli)		
GMI15-005-DNA (S. aureus)		
G GMI15-006-DNA (S. aureus)		
. Total DNA amoun	(microgram) [as	a dec
GMI15-001-DNA (Salmonella)	
2 GMI15-002-DNA (Salmonella)	
B GMI15-003-DNA (E. coli)		
GMI15-004-DNA (E. coli)		
5 GMI15-005-DNA (S. aureus)		
G GMI15-006-DNA (S. aureus)		
. For the DNA recei	ved, DNA quality	(e.g. F
library preparation	was measured o	n (ple
Bioanalyser		
Nanodrop		
DNA quality not measured		
Other		
uici, picase specily.		
	Qubit Nanodrop DNA concentration not mease Other ther, please specify: Measure of DNA c GMI15-001-DNA (Salmonella GMI15-002-DNA (Salmonella GMI15-003-DNA (E. coli) GMI15-005-DNA (S. aureus) GMI15-001-DNA (Salmonella GMI15-001-DNA (Salmonella GMI15-006-DNA (S. aureus) GMI15-001-DNA (Salmonella GMI15-001-DNA (Salmonella GMI15-002-DNA (Salmonella GMI15-003-DNA (E. coli) GMI15-003-DNA (E. coli) GMI15-004-DNA (E. coli) GMI15-006-DNA (S. aureus)	Nanodrop DNA concentration not measured Other ther, please specify: Measure of DNA concentration (ng GMI15-001-DNA (Salmonella) GMI15-002-DNA (Salmonella) GMI15-003-DNA (E. coli) GMI15-005-DNA (S. aureus) Total DNA amount (microgram) [as GMI15-001-DNA (Salmonella) GMI15-003-DNA (E. coli) GMI15-003-DNA (S. aureus) Total DNA amount (microgram) [as GMI15-001-DNA (Salmonella) GMI15-003-DNA (E. coli) GMI15-003-DNA (E. coli) GMI15-005-DNA (S. aureus) For the DNA received, DNA quality library preparation was measured of Bioanalyser Nanodrop DNA quality not measured Other

GMI Proficiency Testing 2015 - bacterial cultures and DNA						
20. Measure of DNA quality (e.g	g. RIN or 260/280 ratio) [as a decimal separator, please					
use full stop (.)]						
20.1 GMI15-001-DNA (Salmonella)						
20.2 GMI15-002-DNA (Salmonella)						
20.3 GMI15-003-DNA (E. coli)						
20.4 GMI15-004-DNA (E. coli)						
20.5 GMI15-005-DNA (S. aureus)						
20.6 GMI15-006-DNA (S. aureus)						
21. If relevant: measure of DNA	quality (260/230 ratio) [as a decimal separator, please					
use full stop (.)]	, (, (,,,,					
21.1 GMI15-001-DNA (Salmonella)						
21.2 GMI15-002-DNA (Salmonella)						
21.3 GMI15-003-DNA (E. coli)						
21.4 GMI15-004-DNA (E. coli)						
21.5 GMI15-005-DNA (S. aureus)						
21.6 GMI15-006-DNA (S. aureus)						

SEQUENCING

22. What protocol was used to prepare the sample library for sequencing? For commercial kits please provide the full kit name, item number, and lot number if possible. For noncommercial kits please provide a citation for the protocol, or submit a summary of the protocol. Please note any deviations from the kit or cited protocol

				r citea proto
For	ommercial kits; full kit name:			
For	ommercial kits; catalogue number:			
For	ommercial kits; lot number:			
For	oncommercial kits; citation for the protocol:			
For	oncommercial kits; summary of the protocol:			
Devi	ations from the kit or cited protocol			
*	23. Please indicate the sequenc	ing platform yo	ou used in the prof	ficiency test
(pl	ease select one answer)			
0	Ion Torrent PGM			
0	Ion Torrent Proton			
0	Genome Sequencer Junior System (454)			
0	Genome Sequencer FLX System (454)			
0	Genome Sequencer FLX+ System (454)			
0	PacBio RS			
0	PacBio RS II			
0	HiScanSQ			
0	HiSeq 1000			
0	HISeq 1500			
0	HiSeq 2000			
0	HiSeq 2500			
0	Genome Analyzer lix			
0	MiSeq			
0	MiSeq Dx			
0	MiSeq FGx			
0	ABI SOLID			
0	other			
	ner, please specify			

GMI Proficiency Testing 2015 - bacterial cultures and DNA 24. Sequencing details #1 (please select one answer) Single-end Paired-end O Not relevant 25. Sequencing details #2: For the sequencing, the read length (bp) was set to be (expected read length) *26. Reads trimmed before upload (please select one answer) Yes O No If trimmed, which tool was applied (in the text box below, please insert name and URL/link (if possible)) 27. For the analysis of the sequences from the bacterial cultures and the corresponding DNA in the proficiency test, assembly is not requried. If, however, you were to assemble your sequences, which assembly tool would you apply? in the text box below, please insert name and URL (e.g. Velvet, https://www.ebi.ac.uk/~zerbino/velvet/, open access) Assembly tool:

ANALYSIS of sequences

<u>-</u> ,	o characterize or differentiate isolates (please select
all that apply)?	
MLST	
Allele-based	
Gene-by-gene-based	
SNP-based	
None	
Other (please specify)	
29. If you determined the MLST-type	e of the sequenced DNA, how was the analysis
performed (please select one answ	er)?
MLST-analysis was performed on raw reads	
MLST-analysis was performed on contigs	
MLST-analysis was not performed	
•	resistance (AMR) genes present in the sequenced
DNA, how was the analysis perform	led (please select one answer)?
 Analysis for AMR-genes was performed on raw read 	s
Analysis for AMR-genes was performed on contigs	
Analysis for AMR-genes was not performed	
31. For the DNA from the received b	acterial culture, if MLST-analysis was performed
	hich MLST-type does the isolate belong to?
31.1 GMI15-001-BACT (Salmonella)	
31.2 GMI15-002-BACT (Salmonella)	
31.3 GMI15-003-BACT (E. coli)	
31.4 GMI15-004-BACT (E. coli)	
31.5 GMI15-005-BACT (S. aureus)	
31.6 GMI15-006-BACT (S. aureus)	

32. For the DNA from the received b	acterial culture, if MLST-analysis was performed
based on the sequence analysis, wh	nich alleles characterize the isolate?
32.1 GMI15-001-BACT (Salmonella)	
32.2 GMI15-002-BACT (Salmonella)	
32.3 GMI15-003-BACT (E. coli)	
32.4 GMI15-004-BACT (E. coli)	
32.5 GMI15-005-BACT (S. aureus)	
32.6 GMI15-006-BACT (S. aureus)	
33. For the DNA from the received b	acterial culture, if analysis for antimicrobial
resistance genes was performed ba	ased on the sequence analysis, which antimicrobial
resistance genes does the isolate h	arbour (please list the genes according to the
following order of antimicrobial cla	sses: Aminocyclitols, aminoglycosides, ß-lactams,
fluoroquinolones, glycopeptides, lii	ncosamides, macrolides, oxazolidones, phenicols,
pleuromutilins, polypeptide antibio	tics, quinolones, streptogramins, sulfonamides,
tetracyclines, trimethoprim, other)?	
33.1 GMI15-001-BACT (Salmonella)	
33.2 GMI15-002-BACT (Salmonella)	
33.3 GMI15-003-BACT (E. coli)	
33.4 GMI15-004-BACT (E. coli)	
33.5 GMI15-005-BACT (S. aureus)	
33.6 GMI15-006-BACT (S. aureus)	
34. For the received DNA. if MLST-a	nalysis was performed based on the sequence
analysis, which MLST-type does the	
34.1 GMI15-001-DNA (Salmonella)	
34.2 GMI15-002-DNA (Salmonella)	
34.3 GMI15-003-DNA (E. coli)	
34.4 GMI15-004-DNA (E. coli)	
34.5 GMI15-005-DNA (S. aureus)	
34.6 GMI15-006-DNA (S. aureus)	
35. For the received DNA, if MLST-a	nalysis was performed based on the sequence
analysis, which alleles characterize	-
35.1 GMI15-001-BACT (Salmonella)	
35.2 GMI15-002-BACT (Salmonella)	
35.3 GMI15-003-BACT (E. coli)	
35.4 GMI15-004-BACT (E. coli)	
35.5 GMI15-005-BACT (S. aureus)	
35.6 GMI15-006-BACT (S. aureus)	

36. For the received DNA, if analysis for antimicrobial resistance genes was performed

based on the sequ	uence analysis, w	hich antimicrobia	l resistance genes	does the isolate
harbour (please li	ist the genes acco	ording to the follow	ving order of antin	nicrobial classes:
,	minoglycosides, f	•	, , , ,	- '
lincosamides, ma	acrolides, oxazolid	lones, phenicols, _l	pleuromutilins, po	lypeptide
, <u>-</u>	lones, streptograr	nins, sulfonamides	s, tetracyclines, tr	imethoprim,
other)?				
36.1 GMI15-001-DNA (Salm	ionella)			
36.2 GMI15-002-DNA (Salm	ionella)			
36.3 GMI15-003-DNA (E. co	oli)			
36.4 GMI15-004-DNA (E. co	oli)			
36.5 GMI15-005-DNA (S. au	reus)			
36.6 GMI15-006-DNA (S. au	reus)			
37. For the detect	tion of the Multi L	ocus Sequence Ty	pe, which tool did	l you apply? in
	w, please insert n	•	• •	
Typing), http://cge	e.cbs.dtu.dk/servi	ces/MLST/, open a	iccess)	-
Tool for detection of			·	
MLST:				
38. For the detect	tion of the resista	nce genes harbou	red in the seqence	s, which tool did
you apply? in the	text box below, p	lease insert name	and URL (e.g. Re	sFinder,
http://cge.cbs.dtu	ı.dk/services/Resf	inder/, open acce	ss)	
Tool for detection of				
resistance genes:				

SUBMITTED datafiles

*39. The obtained non-assembled sequence data have been uploaded for bacterial cultures and DNA following the description of batch-upload in the PTprotocol, for

	Yes	No
Salmonella	0	0
E. coli	\circ	\circ
S. aureus	0	O
Comments:		

Info:

For both bacterial cultures and DNA, the submitted sequence data (fastq-files) will be evaluated according to the following specific quality markers, e.g. read length (bp), N50 (bp), total number of contigs and total length of sequence (bp) including percentage of reference genome covered. In addition, the PT-organizers will assemble the submitted reads and compare these assemblies 1) towards the relevant closed genome to assess the sequence error rate and coverage of the scaffold and 2) between the obtained sequences for both the bacterial cultures and DNA.

GMI Proficiency Testing 2015 - FASTQ dataset

Introduction

This survey seeks to capture info in relation to the fastq data set component of the GMI Proficiency test (PT) 2015.

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

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

GMI is a global, visionary taskforce of scientists and other stakeholders who share an aim of applying novel genomic technologies and informatics tools to improve 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.

The GMI proficiency test 2015 is supported by COMPARE, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643476. In addition, the GMI PT 2015 is supported by the GenomeTrakr Network and Microbiologics®.

≭1. Institute na	me / Organization name
*2. Departmen	t name
≭3. Name (s) of	person(s) responsible for the analysis

FASTQ data set

- 4. Were reads quality filtered before conducting the analysis? (please select one answer)
 - Yes
 - O No
- 5. If reads were quality filtered, please provide the name of the program
- 6. For variant detection, which of the following did you use: (please select one answer)
 - C A reference based approach
 - De novo assemblies
 - A combination of both

	de novo assemblies)			
*8. If you use a reference-based approach, which tools do you use for mapping and variant detection (please insert name and URL; e.g. Bowtie2, http://bowtie-bio.sourceforge.net/bowtie2/index.shtml; VarScan, http://varscan.sourceforge.net)? (Enter 'NA' if you do not use a reference-based approach)				
*9. What kind of methodo	ology for phylogeny construction did you apply?			
C SNPs				
Methodology other than SNPs				
If methodology other than SNPs (please sp	pecify):			
applying SNPs, go to question 10, not applying SNPs, go to question 13 10. Which quality criteria dinimum coverage to defin	lid you use for SNP calling? (e.g. % of mapped reads and			
minimum coverage to defin	e variant).			
1 C Typhimurium				
.1 - S. Typhimurium				
.1 - S. Typhimurium .2 - E. coli .3 - S. aureus				
.2 - E. coli .3 - S. aureus	uga far SNDa filtarings			
.2 - E. coli	use for SNPs filtering:			
.2 - E. coli .3 - S. aureus 11. Which criteria did you u .1 - Filter SNPs with excess coverage (i.e.				
.2 - E. coli .3 - S. aureus I1. Which criteria did you us1 - Filter SNPs with excess coverage (i.e. epetitive regions): .2 - Did you filter SNPs occurring in a luster (a.k.a. pruning) (indicate 'yes' or 'no') .3 - Which definition of the cluster did you				
.2 - E. coli .3 - S. aureus I1. Which criteria did you u .1 - Filter SNPs with excess coverage (i.e. epetitive regions): .2 - Did you filter SNPs occurring in a luster (a.k.a. pruning) (indicate 'yes' or 'no') .3 - Which definition of the cluster did you se (i.e. ≥3 SNPs in 1000 base pairs (bp):				
.2 - E. coli .3 - S. aureus I1. Which criteria did you us1 - Filter SNPs with excess coverage (i.e. epetitive regions): .2 - Did you filter SNPs occurring in a luster (a.k.a. pruning) (indicate 'yes' or 'no') .3 - Which definition of the cluster did you				
.2 - E. coli .3 - S. aureus 11. Which criteria did you use1 - Filter SNPs with excess coverage (i.e. epetitive regions): .2 - Did you filter SNPs occurring in a luster (a.k.a. pruning) (indicate 'yes' or 'no') .3 - Which definition of the cluster did you se (i.e. ≥3 SNPs in 1000 base pairs (bp): .4 - Other, please specify:				
.2 - E. coli .3 - S. aureus I1. Which criteria did you u .1 - Filter SNPs with excess coverage (i.e. epetitive regions): .2 - Did you filter SNPs occurring in a luster (a.k.a. pruning) (indicate 'yes' or 'no') .3 - Which definition of the cluster did you se (i.e. ≥3 SNPs in 1000 base pairs (bp): .4 - Other, please specify:				
.2 - E. coli .3 - S. aureus 11. Which criteria did you use1 - Filter SNPs with excess coverage (i.e. epetitive regions): .2 - Did you filter SNPs occurring in a luster (a.k.a. pruning) (indicate 'yes' or 'no') .3 - Which definition of the cluster did you se (i.e. ≥3 SNPs in 1000 base pairs (bp): .4 - Other, please specify:				

GMI Proficiency Testing 2015 - FASTQ dataset

14. If you do assemblies, do you calculate the number of contigs (please select one
answer)
C Yes
O No
C We don't perform assemblies
15. If you do assemblies, do you filter out contigs below a certain size (please select
one answer)
C Yes
C No
C We don't perform assemblies
If yes, indicate minimum size
16. If you do assemblies, do you calculate N50 (please select one answer)
C Yes
C No
C We don't perform assemblies
17. If you perform assemblies, do you calculate the size of the chromosome (please
select one answer)
C Yes
C No
○ We don't perform assemblies
18. Do you calculate coverage as a quality metric? (please select one answer)
C Yes
O No
st19. Did you check for contamination and/or verify the species? If so, which tool did you apply?
In the text box below, please insert name and URL (e.g. KmerFinder 1.2, http://cge.cbs.dtu.dk/services/KmerFinder)
(Enter 'NA' if you do not check for contamination and/or verify the species)

GMI Proficiency Testing 2015 - FASTQ dataset 20. If you did check, could you verify the species? □ We did not attempt to verify species Yes, for all S. Typhimurium Yes, for all E. coli Yes, for all S, aureus No, for some S. Typhimurium No, for some E. coli No, for some S. aureus If no, please indicate why 21. Can you call a Multi Locus Sequence Type (MLST) (please select one answer) Yes, we do this using a BLAST (or BLAST-like) approach based on assemblies Yes, we do this through a mapping based approach (e.g. SRST) 0 We are not interested in MLST Comments **SUBMITTED datafiles** Please carefully follow the instructions regarding the naming of submitted files and the samples that should be included in them! Thank you. *22. The following files have been submitted to the ftp-site: S. Typhimurium E. coli dataset S. aureus dataset dataset A fasta formatted DNA sequence matrix that was used for clustering (e.g., a fasta file extension) A newick formatted file with the clusters themselves (.tre file extension) (the format can be otabianed through the R package APE or using FigTree's "Export Trees" option) Comments: The number and identity of samples in each uploaded file should match exactly those that were included in the original data (i.e., 20, 21, 24 sequences of S. Typhimurium, E. coli and S. aureus, respectively). Within each file the samples should be named as the prefix within the original fastq file and the file should be named ParticipantID_TaxonGroup.appropriateFileExtenation (e.g., LAB1_Ecoli.fasta). See the instructions for additional information on the samples to be included and the naming of them.

GMI Proficiency Test Forum guide

Global Microbial Identifier



Content

Sign up to forum platform	3
Request to join GMI PT forum group	5
Introduce yourself to other GMI PT participants	<i>7</i>
Post a message on GMI PT forum	10
Troubleshooting	12

The purpose of the *GMI Profiency Test* forum is to facilitate the exchange of information and experiences among GMI PT participants.

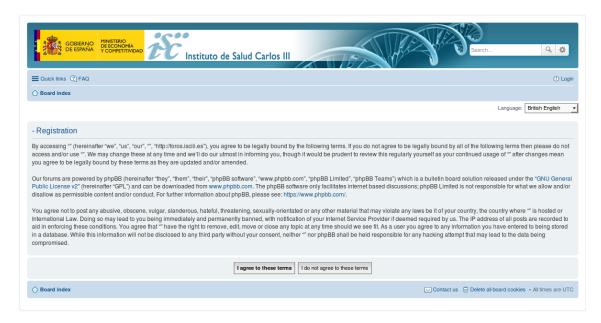
Sign up to forum platform

GMI Profiency Test forum is hosted by the Instituto de Salud Carlos III (Spanish National Health Institute Carlos III). The first step is to register and create an account, which is described below.

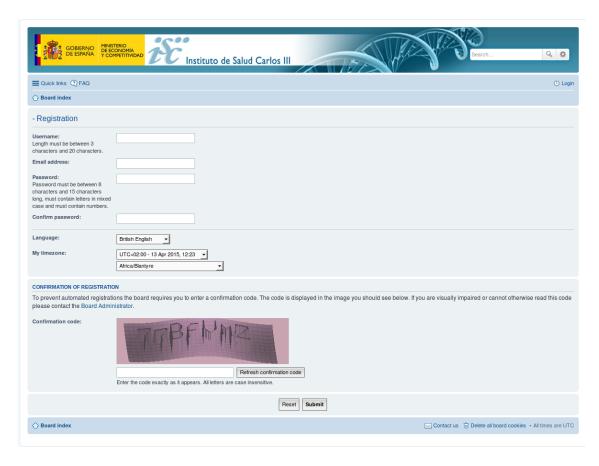
Go to the GMI PT forum web page (https://foros.isciii.es/viewforum.php?f=7) and press the **Register** button.



Read carefully forum terms and press I agree to these terms to accept agreement.



Enter your username (with a length between 3 and 20 characters), email address and set up a password (must be between 8 and 15 characters long, must contain letters in mixed case and numbers). Additionally you could change your time zone. Enter confirmation code and press *Submit*.



A welcome email will be sent to your email address. Click on *Return to the index page* and log in using your selected username and password.



Request to join GMI PT forum group

Joining as a member of GMI PT forum requires to be accepted by *Micro_Bio_GMI* group. Membership is restricted to *GMI Proficiency Test* participants. This section details the procedure to request that you be added to that group.

Click on a link located at the upper right corner of the screen that is labeled with your username. Click on *User Control Panel* and select *Usergroups* tab.



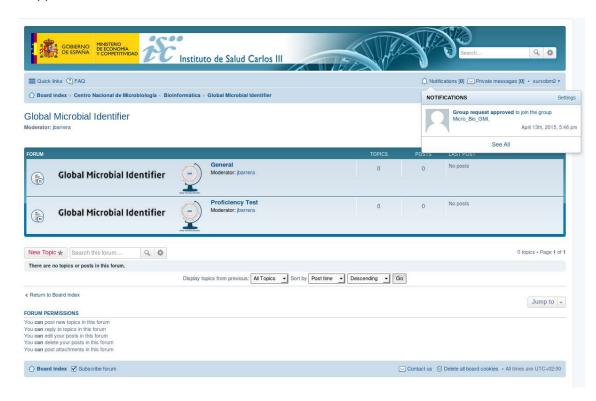
Clik on *Micro_Bio_GMI* radio button. Select *Join selected* in *Select* dropdown list and press *Submit* button.



Confirm request to GMI PT forum group.



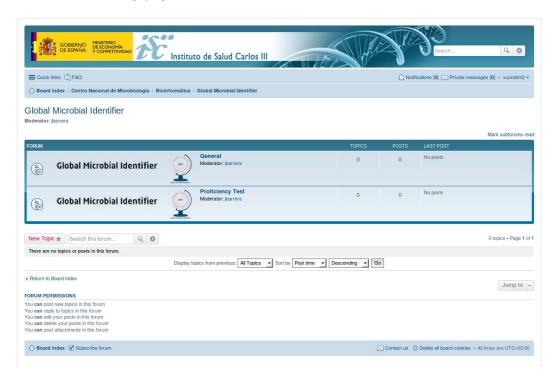
Your request must be approved by GMI PT forum moderator. It should take not more than a few days. You will see a notification in the upper right corner of the screen when your request is approved.



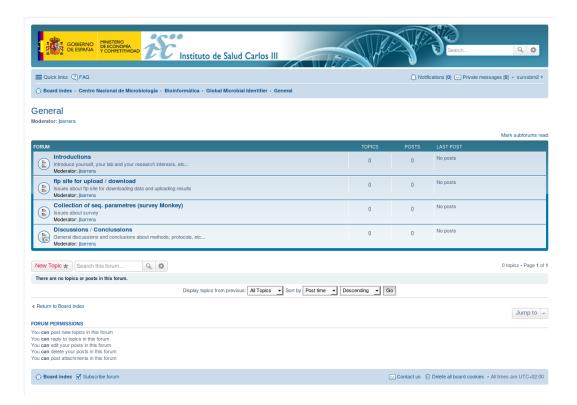
Introduce yourself to other GMI PT participants

The first time you log into the forum after gaining admission to the *Micro_Bio_GMI* group were accepted you should introduce yourself. This is not mandatory but is recommended to facilitate communication among all GMI PT participants.

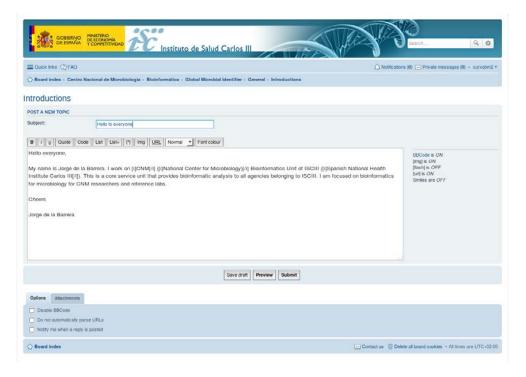
Click on *General* category link (if you do not see *Global Microbial Identifier* forum categories, see Troubleshooting, page 12).



Click on *Introductions* category link.



Write a post introducing yourself and your lab and press *Submit* to publish it.



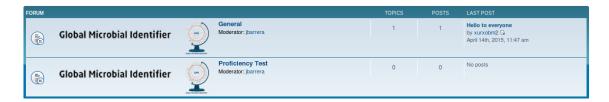
Once post is published it can be read by all GMI PT participants within *Introductions* forum category.



Post a message on GMI PT forum

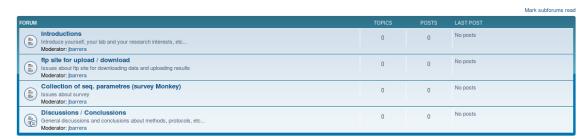
GMI PT forum is organized in categories to improve readability. When you want to write a message you should choose the right forum category to post on. This section details GMI PT forum categories.

Forum categories are organized in two branches: *General* and *Proficiency Test*. No messages can be posted at the main level.



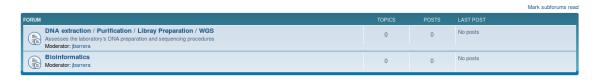
Introduction posts, issues about ftp site or survey and discussions/conclusions must be posted in the suitable category under *General* branch.

General



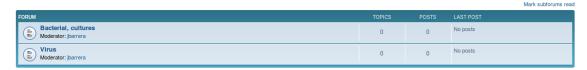
Profiency Test branch encompass categories related to the PT itself.

Proficiency Test



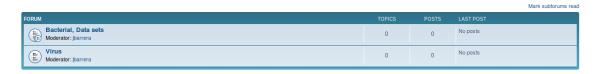
<u>Wet lab</u> and whole genome sequencing issues must be posted in the suitable category under DNA extraction / Purification / Library Preparation / WGS category.

DNA extraction / Purification / Libray Preparation / WGS



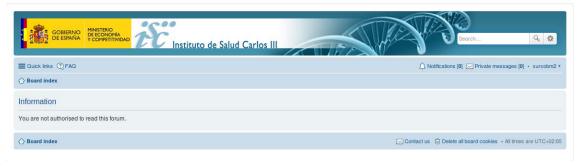
<u>Dry lab</u> issues must be posted in the suitable category under *Bioinformatics*.

Bioinformatics



Troubleshooting

Problem: You are not authorized to read this forum



This message is shown if you have not registered.

Solution: Register as user following steps described in this guide.

Problem: Global Microbial Identifier forum categories are not shown



GMT PT forum is hosted by corporate forum site of *Instituto de Salud Carlos III*. Therefore, GMI PT forum is not the root forum. To access directly to GMI PT forum you should browse through *Bioinformática -> Global Microbial Identifier* or use URL https://foros.isciii.es/viewforum.php?f=7