

SUPPLEMENTARY MATERIAL

Deep Neural Networks Outperform Human Expert's Capacity in Characterizing Bioleaching Bacterial Biofilm Composition

Antoine Buetti-Dinh^{*a,b}, Vanni Galli^{*c}, Sören Bellenberg^{*d}, Olga Ilie^{a,b}, Malte Herold^e,
Stephan Christel^f, Mariia Boretska^d, Igor V. Pivkin^{a,b}, Paul Wilmes^e, Wolfgang Sand^{d,g,h},
Mario Veraⁱ, Mark Dopson^f

^a*Institute of Computational Science, Faculty of Informatics, Università della Svizzera italiana, Lugano, Switzerland*

^b*Swiss Institute of Bioinformatics, Lausanne, Switzerland*

^c*Institute for Information Systems and Networking, University of Applied Sciences of Southern Switzerland, Manno, Switzerland*

^d*Fakultät für Chemie, Biofilm Centre, Universität Duisburg-Essen, Essen, Germany*

^e*Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Belvaux, Luxembourg*

^f*Centre for Ecology and Evolution in Microbial Model Systems, Linnaeus University, Kalmar, Sweden*

^g*College of Environmental Science and Engineering, Donghua University, Shanghai, People's Republic of China*

^h*Mining Academy and Technical University Freiberg, Freiberg, Germany*

ⁱ*Institute for Biological and Medical Engineering. Schools of Engineering, Medicine & Biological Sciences. Department of Hydraulic & Environmental Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile*

*Equal contribution

Contents

1	Deep Learning Resampling	2
2	Human Expert's Performance	3
3	Negative Control	5
4	Example of AL Images	5
5	Deep Learning Performance <i>vs.</i> Amount of Training Data	6
6	TensorFlow code and microscopy images	7
7	Test For Humans	8

1. Deep Learning Resampling

Run	Correct Prediction Rate
1	90.20%
2	89.80%
3	93.10%
4	87.30%
5	90.20%
6	88.90%
7	86.30%
8	90.10%
9	89.50%
10	87.40%
	89.28%

Figure S 1: Resampling. Different division of the images into training and testing sets is carried out in ten different runs and averaged (bold).

2. Human Expert's Performance

		PREDICTED CLASS					correct prediction rate
		A	L	SAL	SL	S	
ACTUAL CLASS	A	2	0	0	0	0	100%
	L	0	1	0	0	0	50%
	SAL	0	0	1	1	0	50%
	SL	0	0	0	2	0	100%
	S	0	0	0	0	2	100%
							80.0%

Figure S 2: Best human experts' performance in predicting the species composition of bacterial biofilms (see "Human Expert 12" in Fig. S3). The matrices indicate the share of images correctly deduced in the diagonal line (shaded grey) and categories the misclassified images were assigned are shown in the horizontal plane.

Correct Answer	Human Expert 1	Human Expert 2	Human Expert 3	Human Expert 4	
AXX	A	A	SAX	AXX	
SXX	S	S	SXX	SXX	
SAL	SL	SL	SAL	LSX	
LXX	A	L	LXX	ALX	
AXX	SA	A	AXX	AXX	
SLX	SA	SA	SLX	ALS	
SXX	S	S	SXX	SXX	
SLX	SL	SL	SAX	ALS	
LXX	A	L	SXX	ALX	
SAL	SA	SA	LXX	SAX	
Accuracy		0.4	0.7	0.6	0.5

Correct Answer	Human Expert 5	Human Expert 6	Human Expert 7	Human Expert 8	
AXX	AXX	SAX	ALX	AXX	
SXX	SXX	SXX	SXX	SAX	
SAL	LSX	SAL	ALX	SAL	
LXX	ALX	LXX	ASX	SLX	
AXX	LXX	AXX	AXX	AXX	
SLX	ALS	SAL	LSX	SLX	
SXX	SXX	SXX	LSX	SXX	
SLX	LSX	SLX	ASX	SLX	
LXX	ALX	LXX	ALX	SAL	
SAL	SLX	SAX	ASL	LXX	
Accuracy		0.4	0.7	0.4	0.7

Correct Answer	Human Expert 9	Human Expert 10	Human Expert 11	Human Expert 12	
AXX	AXX	AXX	A	A	
SXX	SXX	SXX	S	S	
SAL	LAX	LSX	AS	SL	
LXX	LXX	LXX	LS	AL	
AXX	AXX	AXX	A	A	
SLX	SAL	SAL	LS	SL	
SXX	SXX	SXX	S	S	
SLX	SAL	SAX	LS	SL	
LXX	LXX	LXX	LC=LA=LS?	L	
SAL	SAX	SAX	LS	SLA	
Accuracy		0.6	0.6	0.6	0.8

Correct Answer	Human Expert 13	Human Expert 14	Human Expert 15	Human Expert 16	
AXX	A	AXX	AXX	SAX	
SXX	S	SXX	SXX	SXX	
SAL	AL	LSX	SLX	LAX	
LXX	L	ALX	ALX	LXX	
AXX	A	LXX	LXX	AXX	
SLX	SL	SLA	SLX	SLA	
SXX	S	SXX	SAX/SLX	SXX	
SLX	SAL	SLX	SAX/SLX	SAX	
LXX	L	LXX	ALX	LXX	
SAL	SA	SAX	SAX	LAX	
Accuracy		0.7	0.5	0.3	0.5

Correct Answer	Human Expert 17	Human Expert 18	Human Expert 19	Human Expert 20	
AXX	A	ASX	LXX	LXX	
SXX	S	SXX	SXX	SXX	
SAL	L	LSX	SAX	SAX	
LXX	L	AXX	LXX	SAL	
AXX	A	LXX	SAL	LAX	
SLX	AL	ALS	SAL	LSX	
SXX	S	SXX	SXX	SXX	
SLX	AL	ALS	SAL	LXX	
LXX	L	ALX	LXX	SXX	
SAL	A	SLX	SLX	SLX	
Accuracy		0.6	0.2	0.4	0.3

Figure S 3: Performance of the individual human experts in the custom-built double-blind test. Green highlights indicate correct answer.

3. Negative Control

AXX	16.00%
LXX	12.00%
SAL	21.00%
SAX	17.00%
SLX	15.00%
SXX	19.00%
Expected	16.67%

Figure S 4: Negative control. Performance of deep neural networks when subjected to samples devoid of bacteria, and consisting only on chalcopyrite grains.

4. Example of AL Images

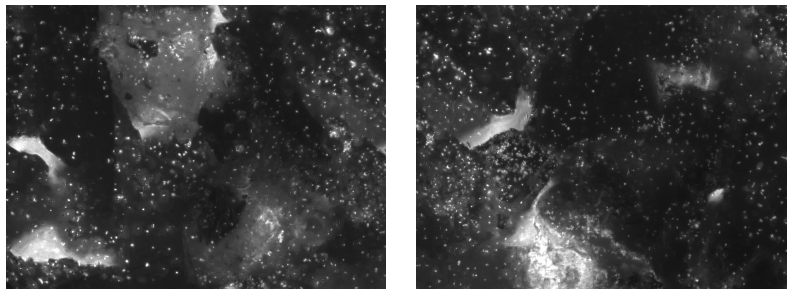


Figure S 5: Example of EFM images representing the AL biofilm category.

5. Deep Learning Performance *vs.* Amount of Training Data

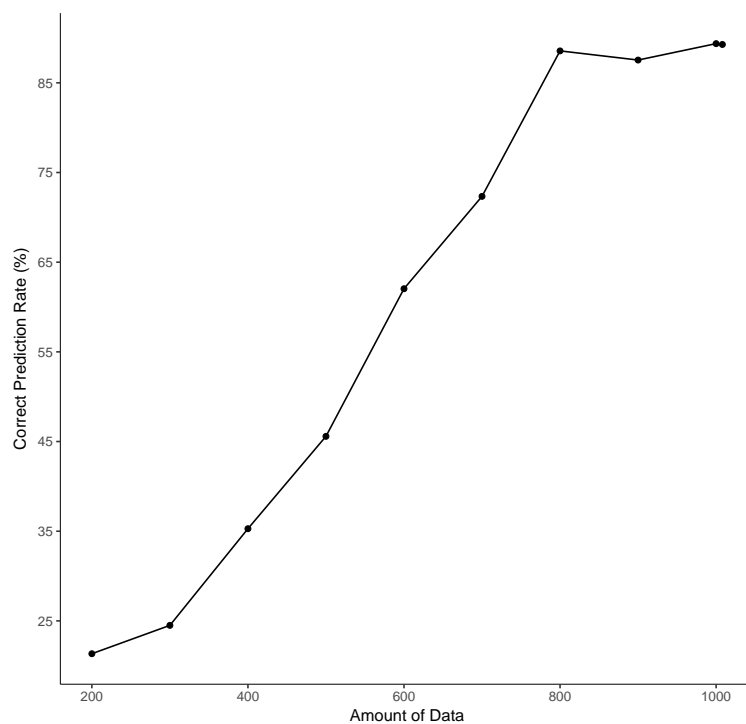


Figure S 6: Deep learning performance *vs.* amount of training data. The algorithm performance was tested in its accuracy expressed in percentage of images accurately classified as a function of the number of images used for training. Data augmentation was applied by duplicating training images. The correct prediction rate was obtained by averaging the outcome over the different image categories.

6. TensorFlow code and microscopy images

TensorFlow code describing the convolutional neural network (CNN) used in this study, as well as the images used for CNN training and testing are available in the FAIRDOMHub repository (DOI: [10.15490/fairdomhub.1.investigation.281.1](https://doi.org/10.15490/fairdomhub.1.investigation.281.1)).

7. Test For Humans

Guess the Microbes !

In the following pages you can see microscopy images of microbial biofilms. The biofilms are composed of bacteria (the bright dots) that feed on mineral particles (dark areas).

After looking at some labelled examples of the different biofilms formed by different bacterial mixes (Fig.1-6), will you be able to correctly guess unlabelled images in the last two pages (Fig.7-8) ?

BACKGROUND

With this test, we would like to study the performance of a computer program to correctly classify microscopy images, based on the bacterial composition of the represented biofilm. The method relies on convolutional neural networks (CNNs, a type of artificial neural networks used in deep learning), which were previously trained on ~500 images.

In particular, this test will allow us to compare the performance of deep learning *vs.* the human capacity in distinguishing biofilm composition based on microscopy images.

Spoiler alert! The computer program turned out to be quite good, reaching ~90% accuracy in guessing the correct image type, can you do better ?

1 TRAINING

The following images are representative of the biofilms formed by the different bacterial mixes. There are three species:

- ★ *Acidithiobacillus caldus*
- ★ *Leptospirillum ferriphilum*
- ★ *Sulfobacillus thermosulfidooxidans*

labelled as **A**, **L** and **S**, respectively.

Images represent biofilms of mixtures composed between 1-3 of these bacteria. For example, the mixture ALX contains the first two bacteria listed above, and not the third (indicated as “X”). Similarly, LXX indicates that the biofilm mixture is only composed of L. And so on...

Optionally, you can look at more such images in the folders named with the mixture names, see the folder `more_images` (you can download it at <https://goo.gl/Yc6vcT> (~200 MB)).

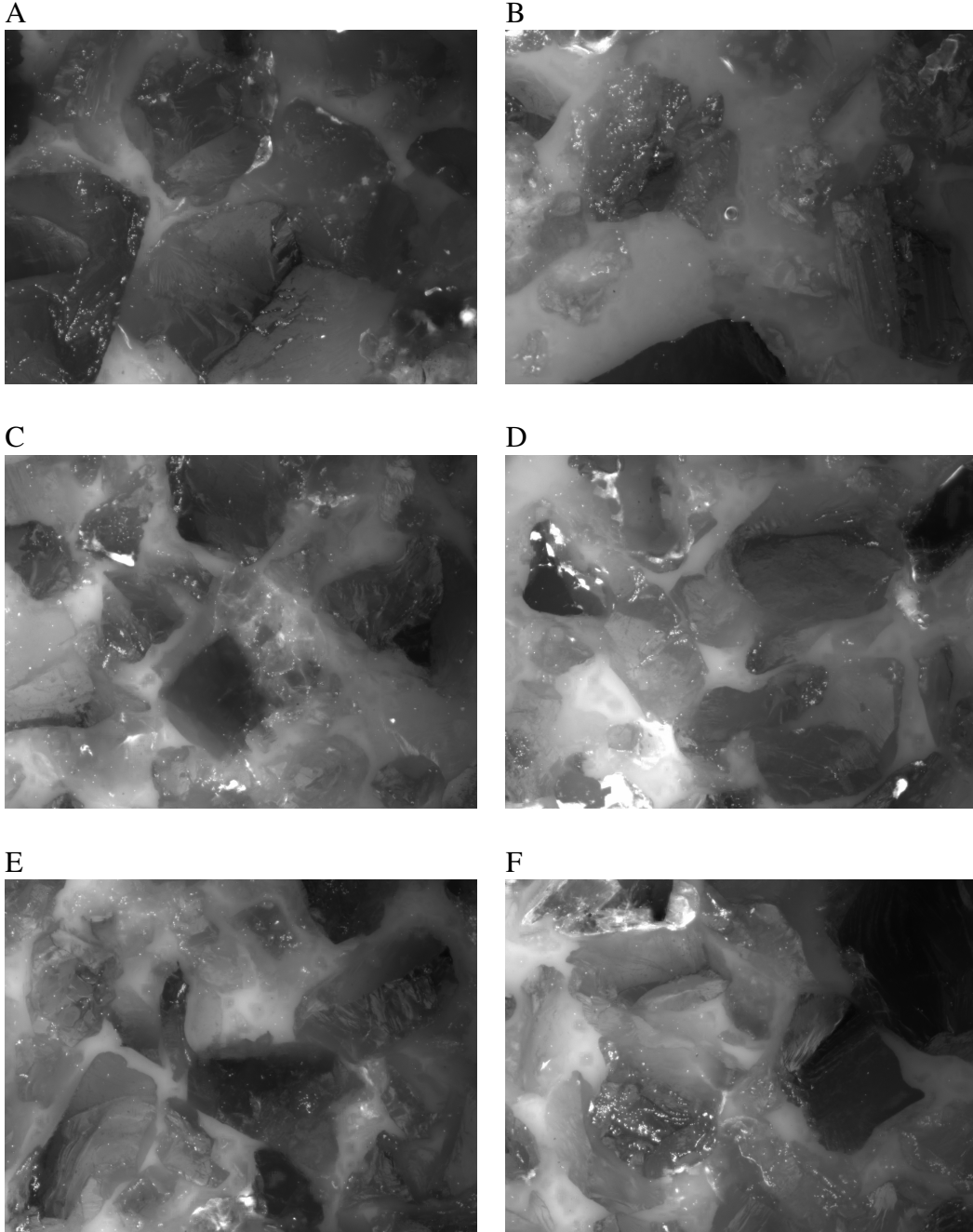


Figure 1: AXX. More images can be found in the folder `more_images/AXX`.

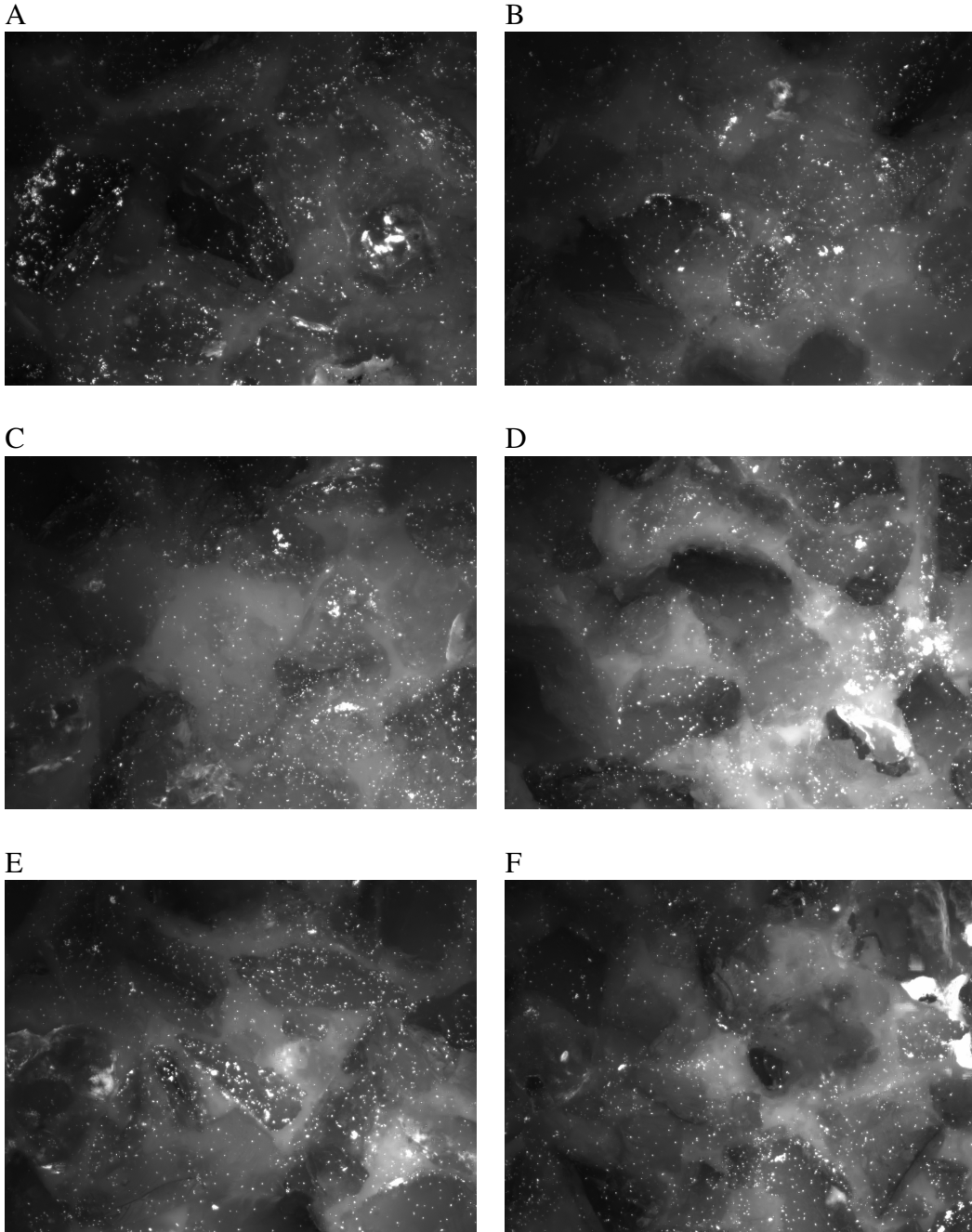


Figure 2: LXX. More images can be found in the folder `more_images/LXX`.

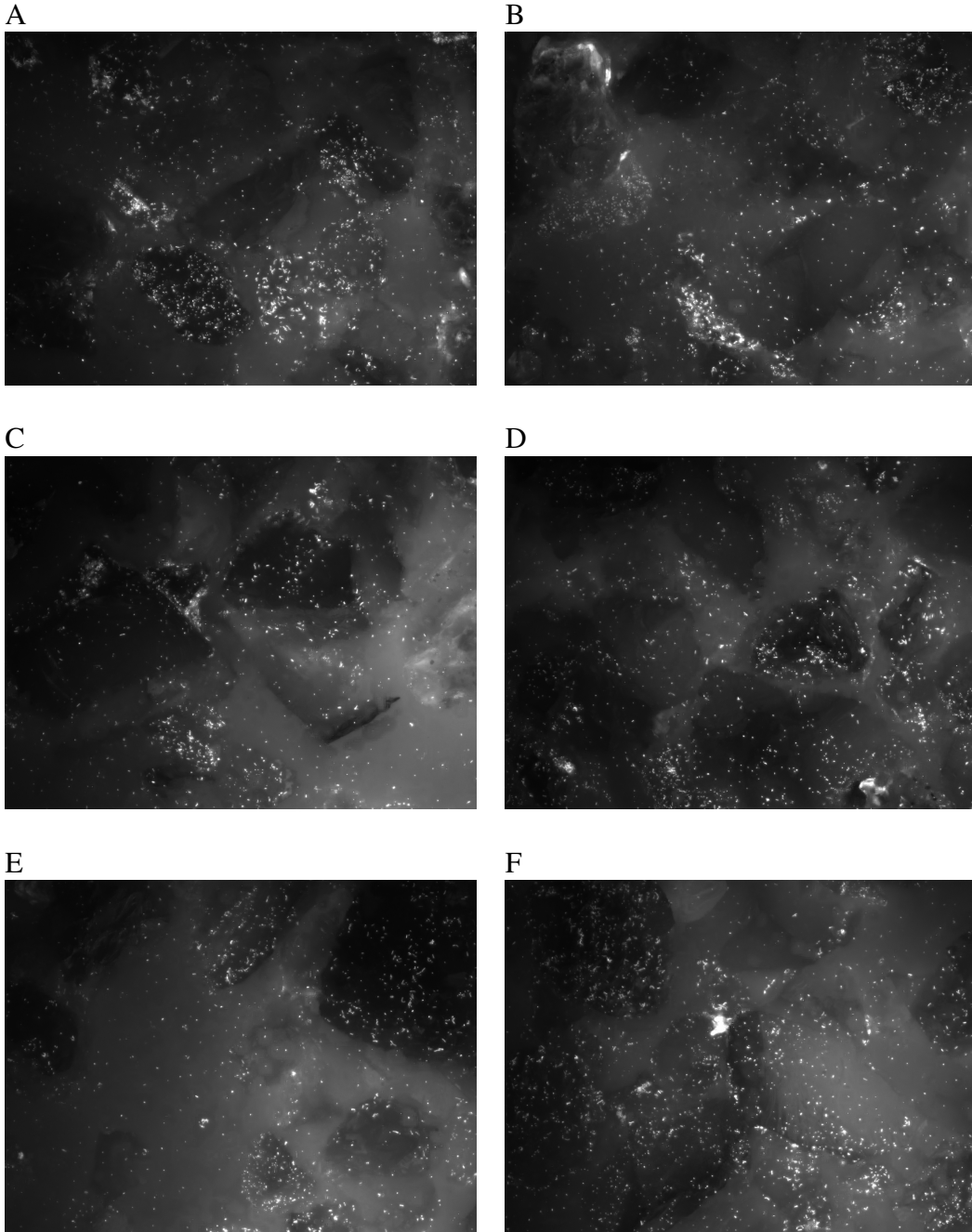


Figure 3: SAL. More images can be found in the folder `more_images/SAL`.

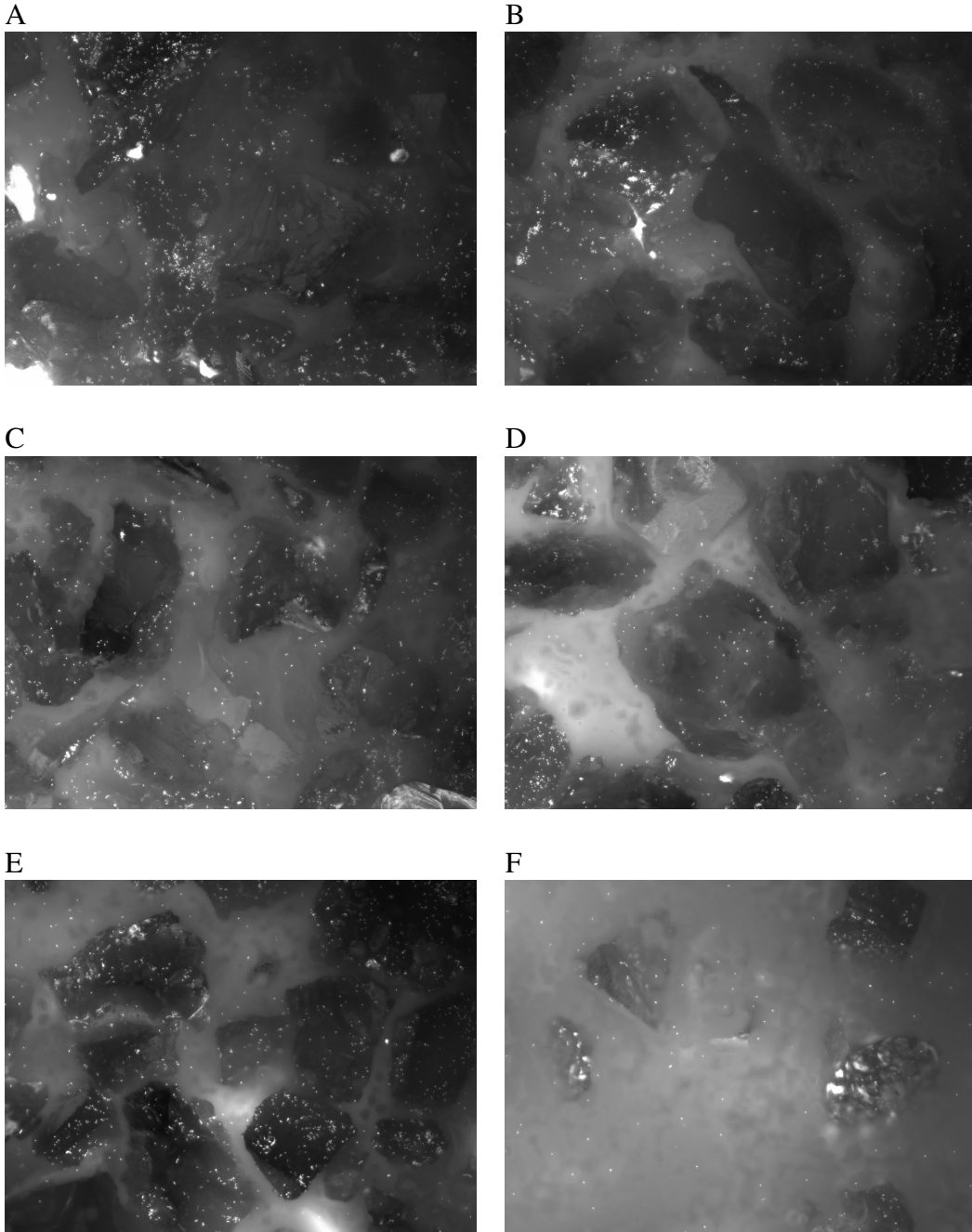


Figure 4: SAX. More images can be found in the folder `more_images/SAX`.

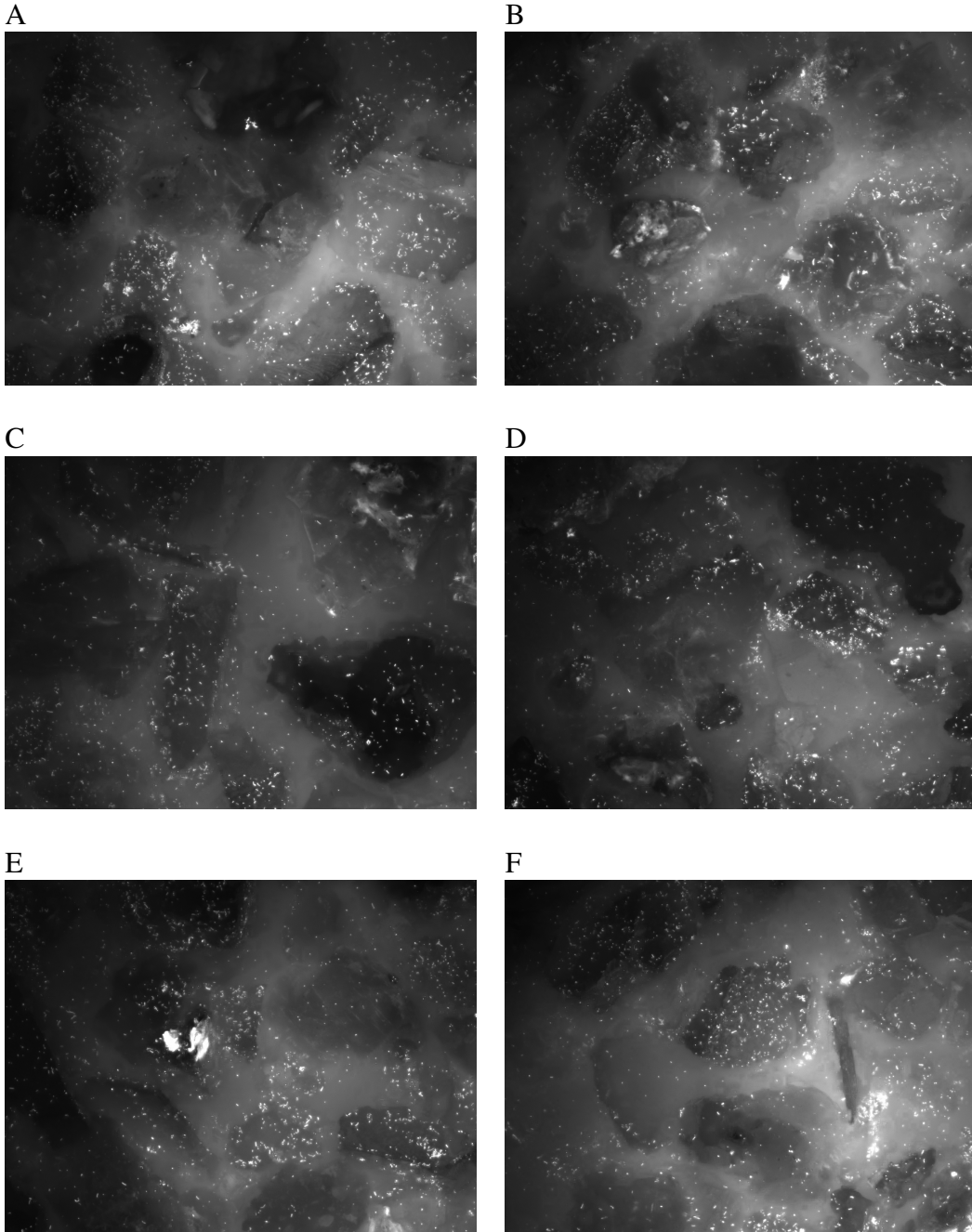


Figure 5: SLX. More images can be found in the folder `more_images/SLX`.

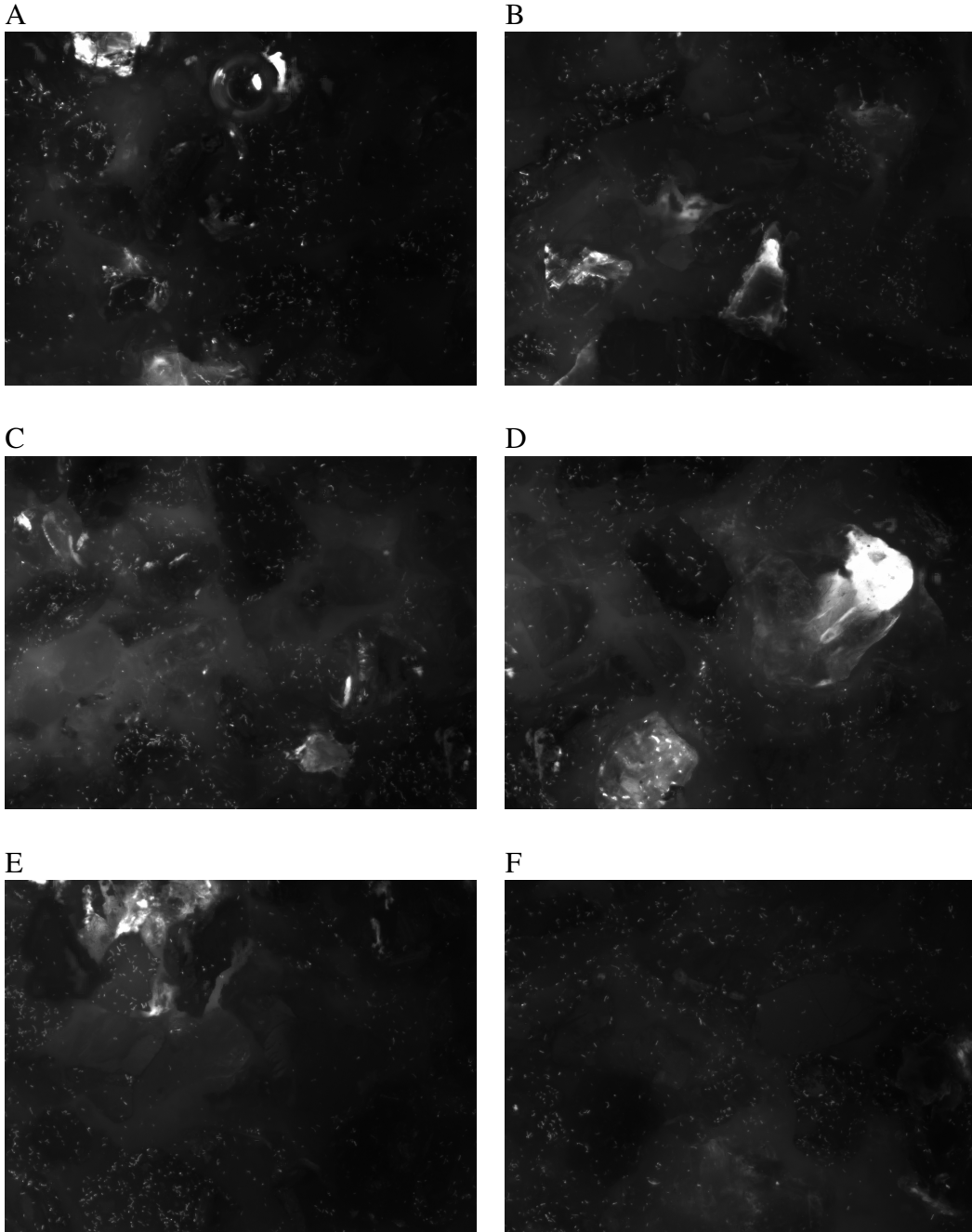


Figure 6: SXX. More images can be found in the folder `more_images/SXX`.

2 TEST

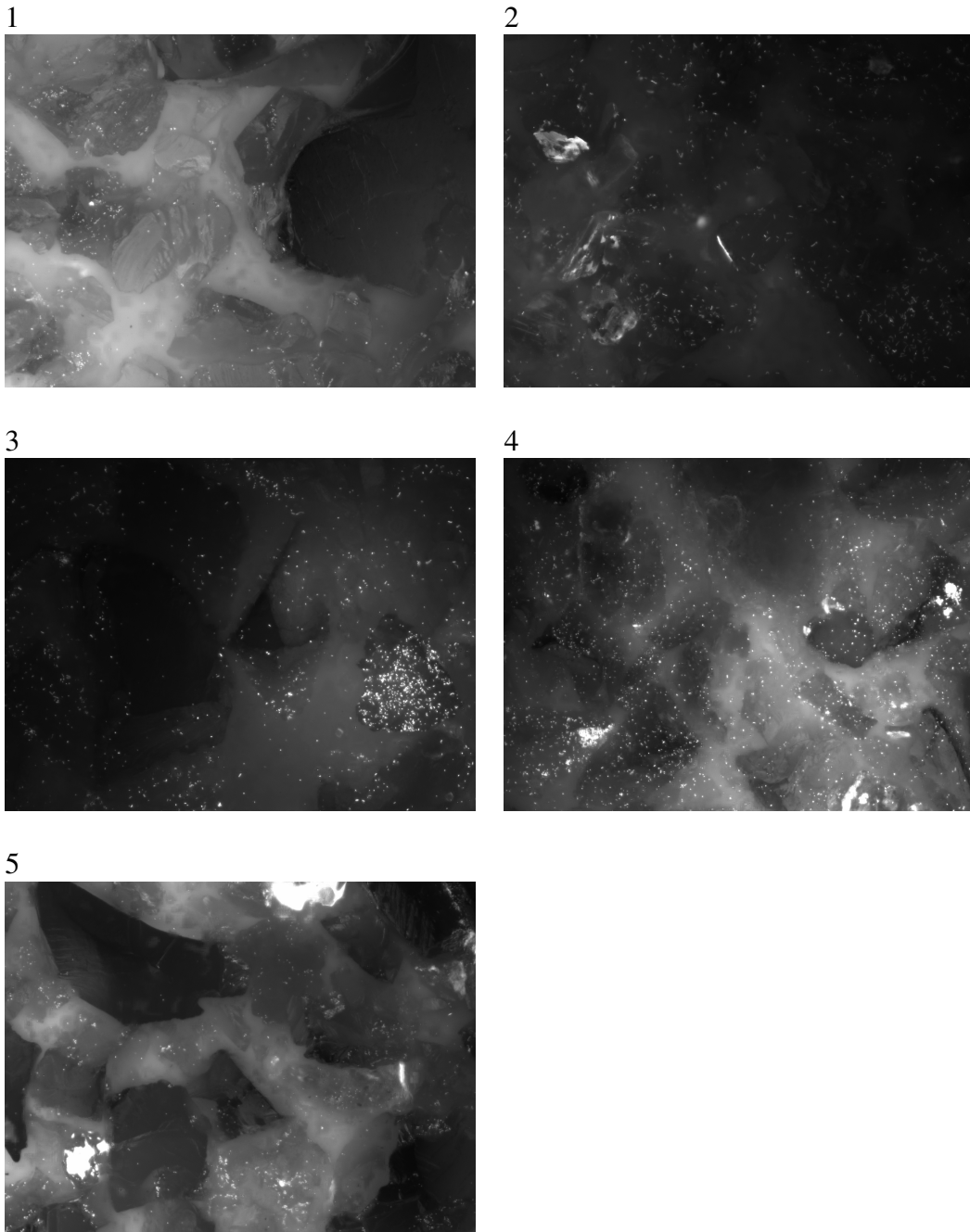


Figure 7: Test (part 1)

6



7



8



9



10

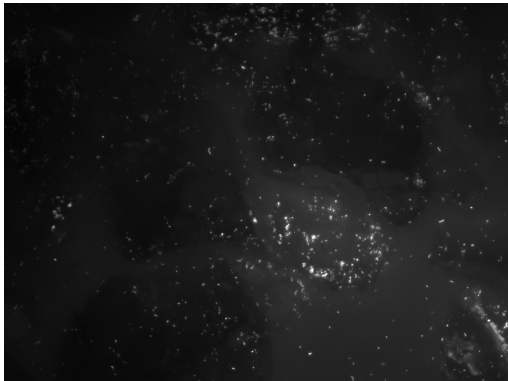


Figure 8: Test (part 2): Please write the name of what you think is the bacterial mixture as your answer. Some features in particular were determinant for your choice ? You can send the answers to antoine.buetti@lnu.se. **THANK YOU!**