

# Supporting information for:

## APE in the Wild: Automated Exploration of Proteomics Workflows in the bio.tools Registry

Vedran Kasalica,<sup>\*,†</sup> Veit Schwämmle,<sup>‡</sup> Magnus Palmblad,<sup>¶</sup> Jon Ison,<sup>§</sup> and  
Anna-Lena Lamprecht<sup>\*,†</sup>

<sup>†</sup>*Dept. of Information and Computing Sciences, Utrecht University, The Netherlands*

<sup>‡</sup>*Dept. of Biochemistry and Molecular Biology, University of Southern Denmark, Denmark*

<sup>¶</sup>*Center for Proteomics and Metabolomics, Leiden University Medical Center, The Netherlands*

<sup>§</sup>*Institut Français de Bioinformatique, CNRS, France*

E-mail: v.kasalica@uu.nl; a.l.lamprecht@uu.nl

### Table of Contents:

- *Table S1*: List of up to 100 solutions generated for each run for Use Case 1 - (supplementary files)
- *Table S2*: List of up to 100 solutions generated for each run for Use Case 2 - (supplementary files)
- *Table S3*: List of up to 100 solutions generated for each run for Use Case 3 - (supplementary files)

- *Table S4*: List of up to 100 solutions generated for each run for Use Case 4 - (supplementary files)
- *Table S5*: Spreadsheet contains general information regarding the exploration runs. Columns aggregate solutions found for each run based on their length - page S-3
- *Table S6*: Quality evaluation of the first 20 workflows per each run, for Use Case 1 - (supplementary files)
- *Table S7*: Quality evaluation of the first 20 workflows per each run, for Use Case 2 - (supplementary files)
- *Table S8*: Quality evaluation of the first 20 workflows per each run, for Use Case 3 - (supplementary files)
- *Table S9*: Quality evaluation of the first 20 workflows per each run, for Use Case 4 - (supplementary files)
- *Table S10*: Summary of the evaluation results - (supplementary files)
- *Figure S1*: Distribution of the workflow quality scores using a box plot - page S-4
- *File S1*: Tool annotations provided for the original domain - (supplementary files)
- *File S2*: Tool annotations provided for the extended domain - (supplementary files)
- *File S1*: Tool annotations provided for the full domain - (supplementary files)

Table S 5: Numbers of workflows obtained in the experiments.

Length	Original		Extended		Full bio.tools	
	I/O	I/O+C	I/O	I/O+C	I/O	I/O+C
<b>Use Case #1</b>						
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	1	0	12	1	15	2
4	6	1	100	38	100	45
5	25	8		100		100
6	96	39				
7	100	100				
<b>Use Case #2</b>						
1	0	0	1	0	1	0
2	0	0	11	0	12	0
3	0	0	100	0	100	0
			...			
8	0	0		0		0
			...			
20	0	0		0		
<b>Use Case #3</b>						
1	0	0	1	0	1	0
2	1	0	47	0	54	0
3	7	0	100	0	100	0
4	31	1		7		8
5	100	5		100		100
6		14				
7		33				
8		72				
9		100				
<b>Use Case #4</b>						
1	0	0	1	0	1	0
2	1	0	15	0	23	0
3	3	0	100	2	100	3
4	8	1		75		85
5	24	6		100		100
6	81	26				
7	100	100				

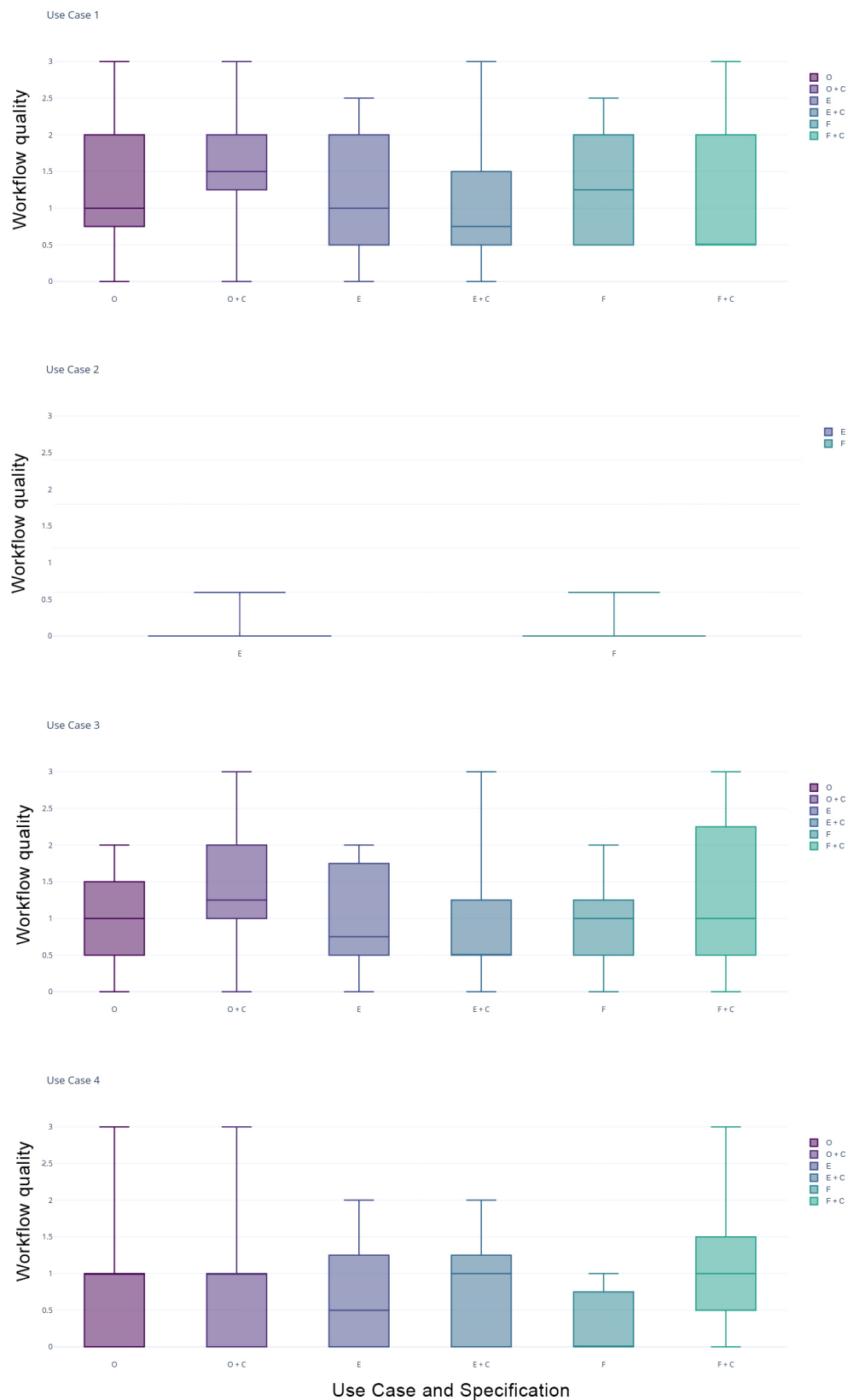


Figure S 1: Workflow quality evaluation