SeqAn Usage Statistics

Svenja Mehringer

24 April 2017

This is an automatically generated report that evaluates the current SeqAn user statistics. The data is taken from the file `../logfiles/seqan.logs` preprocessed by a script `global_seqan.R`. If you want to evaluate this report on another file, please change the variable `seqan_log_file` in the script `global_seqan.R`.

Timeframe

2016-11-22 to 2017-05-10

The Data

The collected infromation contains the following columns:

|	names	desc
1	calls	The number of calls.
2	date	The IP address
3	ip	The Date of the call
4	app	The application name
5	version	The application version
6	cpu	The Bitsystem. Either 32 or 64.
7	os	The operating system (Windows, Linux or Mac OS)
8	who	A Parameter only used in OpenMS. (Here: SeqAn)
9	clustered	0 or 1, depending on wether the call was from a cluster.
10	country	(freeip.net) The country derived from the ip adress
11	region	(freeip.net) The region derived from the ip adress
12	city	(freeip.net) The city derived from the ip adress
13	zipcode	(freeip.net) The zipcode derived from the ip adress
14	latitude	(freeip.net) The latidude derived from the ip adress
15	longitude	(freeip.net) The longitude derived from the ip adress

Table 1: The column names of the collected user information.

Some preprocessing details (`global_seqan.R`):

- All entries marked as clusters are removed from the data set.
- A seperate table keeps track of the amount of cluster uses.
- The date column is transform to a date-format of R.

SeqAn

To display the general usage of Seqan, the plot shows the cumulative usage of all application summed together over the given time period.

Cumulative Sum of All Calls of SeqAn
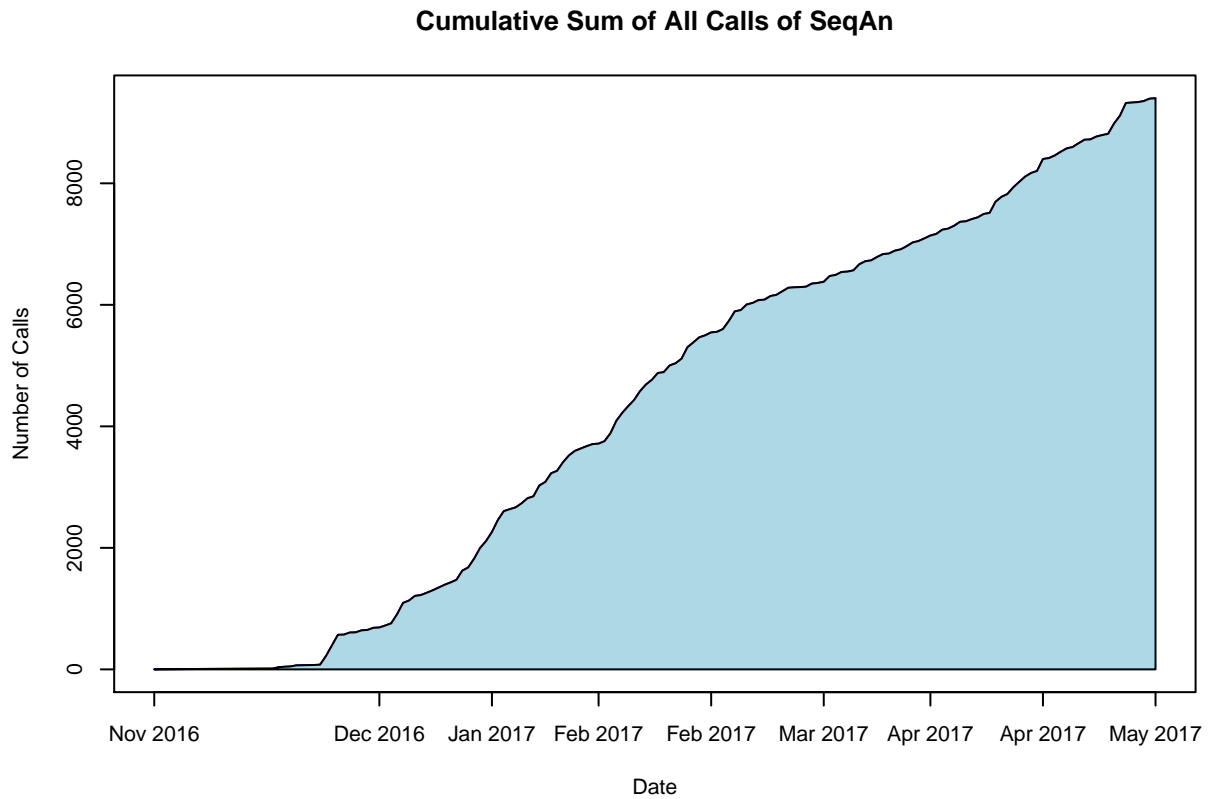
SeqAn Worldmap

Distribution of Countries and Cities per SeqAn Application

	#Countries	Country with most calls	#Cities	City with most calls
razers3	19	United States	63	Ashburn
yara_mapper	10	United States	22	Ashburn
sak	9	United States	29	Ashburn
mason_simulator	8	United States	24	Ashburn
yara_indexer	7	United States	18	Ashburn
seqan_tcoffee	7	United States	10	Ashburn
mason_variator	7	United States	13	Ashburn
micro_razers	6	United States	8	Ashburn
mason_genome	6	United States	12	Ashburn
stellar	5	United States	14	Ashburn
razers	5	United States	8	Ashburn
gustaf	5	United States	12	Ashburn
pair_align	4	United States	6	Ashburn
mason_materializer	4	United States	12	Ashburn
fiona	4	United States	8	Ashburn
alf	4	Germany	8	Berlin
tree_recon	3	United States	5	Ashburn
splazers	3	United States	5	Ashburn
seqcons2	3	United States	5	Ashburn
samcat	3	United States	4	Ashburn
sam2matrix	3	United States	4	Ashburn
rabema_prepare_sam	3	United States	7	Ashburn
rabema_evaluate	3	United States	8	Ashburn
rabema_build_gold_standard	3	United States	7	Ashburn
mason_methylation	3	Germany	8	Berlin
mason_frag_sequencing	3	Germany	9	Berlin
gustaf_mate_joining	3	United States	7	Ashburn
dfi	3	United States	4	Ashburn
snp_store	2	United States	4	Ashburn
sgip	2	United States	3	Ashburn
roi_feature_projection	2	United States	3	Ashburn
mason_splicing	2	Germany	5	Berlin
insegt	2	United States	4	Ashburn
four2three	2	Germany	2	Berlin
fiona_illumina	2	United States	2	Ashburn

	#Countries	Country with most calls	#Cities	City with most calls
compute_gain	2	United States	2	Ashburn
casbar	2	United States	3	Ashburn
bisar	2	United States	3	Ashburn
bam2roi	2	United States	3	Ashburn
roi_plot_thumbnails	1	Germany	1	Berlin
rep_sep	1	Japan	1	
razers3_quality2prob	1	New Zealand	1	Auckland
fx_fastq_stats	1	Germany	1	Berlin
fx_bam_coverage	1	Germany	2	Berlin

Daily and Weekly usage of SeqAn Applications

	Total Sum	Average per Day	Average per Week
alf	104	0.6153846	4.3076923
bam2roi	153	0.9053254	6.3372781
bisar	198	1.1715976	8.2011834
casbar	197	1.1656805	8.1597633
compute_gain	178	1.0532544	7.3727811
dfi	196	1.1597633	8.1183432
fiona	191	1.1301775	7.9112426
fiona_illumina	146	0.8639053	6.0473373
four2three	77	0.4556213	3.1893491
fx_bam_coverage	82	0.4852071	3.3964497
fx_fastq_stats	46	0.2721893	1.9053254
gustaf	189	1.1183432	7.8284024
gustaf_mate_joining	165	0.9763314	6.8343195
insegt	195	1.1538462	8.0769231
mason_frag_sequencing	94	0.5562130	3.8934911
mason_genome	217	1.2840237	8.9881657
mason_materializer	221	1.3076923	9.1538462
mason_methylation	49	0.2899408	2.0295858
mason_simulator	328	1.9408284	13.5857988
mason_splicing	45	0.2662722	1.8639053
mason_variator	232	1.3727811	9.6094675
micro_razers	155	0.9171598	6.4201183
pair_align	152	0.8994083	6.2958580
rabema_build_gold_standard	206	1.2189349	8.5325444
rabema_evaluate	212	1.2544379	8.7810651
rabema_prepare_sam	206	1.2189349	8.5325444
razers	170	1.0059172	7.0414201
razers3	934	5.5266272	38.6863905
razers3_quality2prob	1	0.0059172	0.0414201
rep_sep	1	0.0059172	0.0414201
roi_feature_projection	148	0.8757396	6.1301775
roi_plot_thumbnails	30	0.1775148	1.2426036
sak	247	1.4615385	10.2307692
sam2matrix	192	1.1360947	7.9526627
samcat	198	1.1715976	8.2011834

	Total Sum	Average per Day	Average per Week
seqan_tcoffee	212	1.2544379	8.7810651
seqcons2	121	0.7159763	5.0118343
sgip	147	0.8698225	6.0887574
snp_store	196	1.1597633	8.1183432
splazers	150	0.8875740	6.2130178
stellar	253	1.4970414	10.4792899
tree_recon	194	1.1479290	8.0355030
yara_indexer	242	1.4319527	10.0236686
yara_mapper	1027	6.0769231	42.5384615

Top 10 Applications

In the following the data was aggregated by application. The first figure pictures the sum of all calls of each application, while the second shows the number of unique users. The figures differentiate between individual and cluster calls/users. The term 'cluster calls/users' refers to IP addresses that were previously identified as part of a cluster, but left in the dataset as a representative for the corresponding cluster.

Top 10 Applications based on overall calls
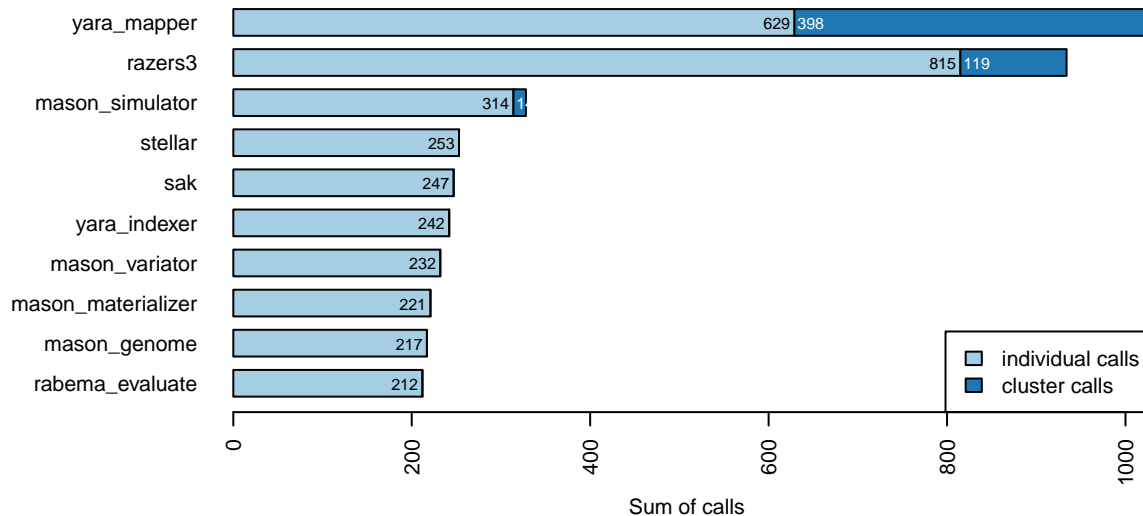
Top 10 Applications based on overall users
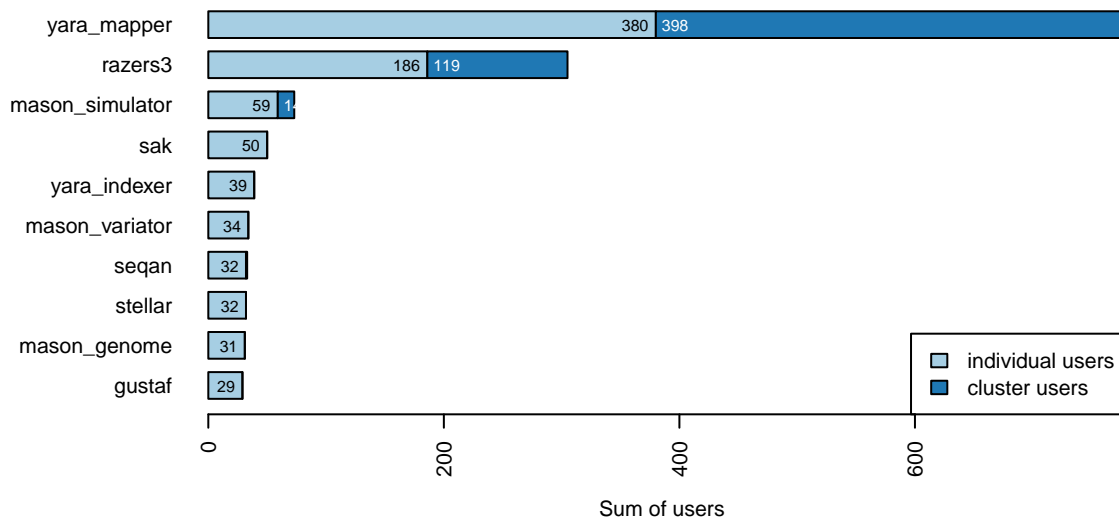
Operating System

For an overview over the distribution of different operating systems the data was aggregated by column 'os'. As before, the first figure pictures the sum of all calls of each operating system, while the second shows the number of unique users.

**Distribution of Operating Systems
over 9400 overall Calls**

Linux 94.3%

Windows 0.1%
MacOS 5.6%
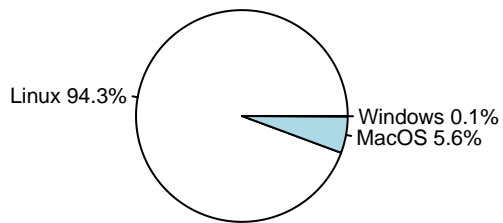
**Distribution of Operating Systems
over 1306 Unique Users**

Linux 94.6%

Windows 0.5%
MacOS 4.9%
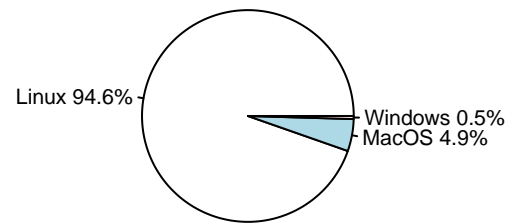
Unique User Statistic

The bar plots on the left display the unique new users. New Unique users means that each IP Adress tracked only appears once in this statistic at the time point of it's first occurence. The line plot on the right displays the same information as a line instead of bars. The red line on the left plot shows the cumulative sum of the new unique users.

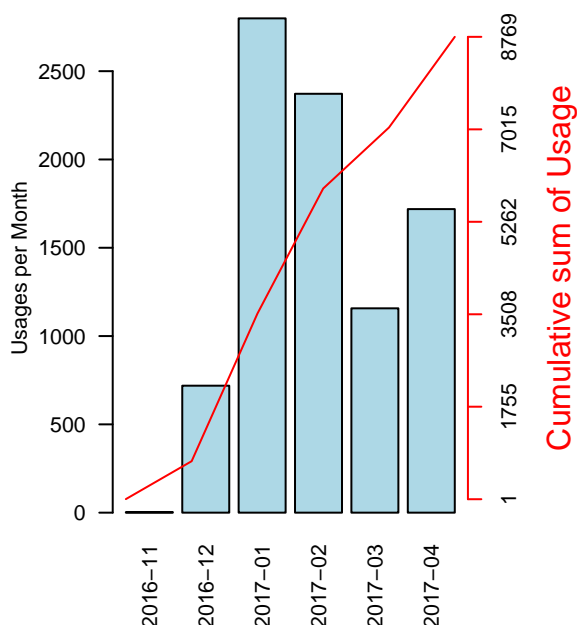**New (unique) SeqAn User Count per Month**
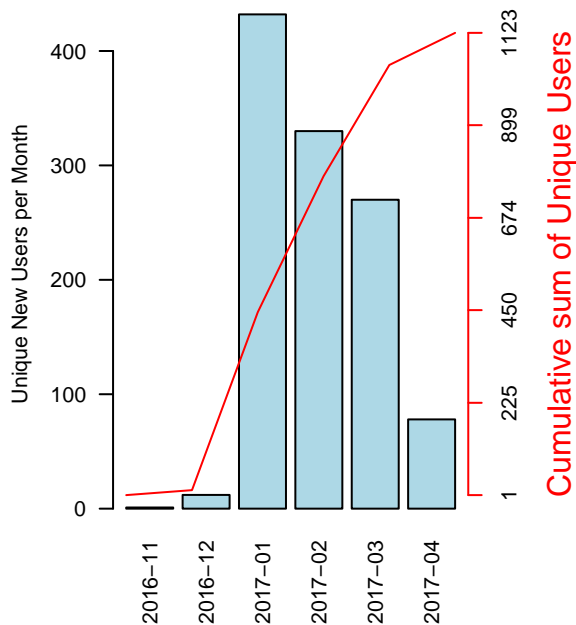
New (unique) SeqAn User Count per Month
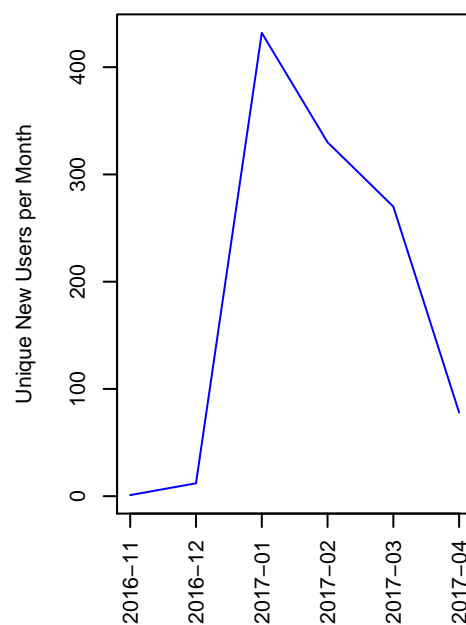
For comparison the following plots show the toal counts of usages (calls) per month, accumulated and as a line plot.

SeqAn Usage Count per Month

SeqAn Usage Count per Month