Trackable and scalable LC-MS metabolomics data processing using asari

Shuzhao Li*, Amnah Siddiqa, Maheshwor Thapa, Shujian Zheng
Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA

*Corresponding author, E-mail: shuzhao.li@jax.org

Metabolomics holds the promise to measure and quantify small molecules comprehensively in biological systems, and LC-MS (liquid chromatography coupled mass spectrometry) has become the leading technology in the field. Significant challenges still exist in the computational processing of data from LC-MS metabolomic experiments into metabolite features, including provenance and reproducibility of the current software tools. We present here asari, a new open-source software tool for LC-MS metabolomics data processing. Asari is designed with a set of new algorithmic framework and data structures, and all steps are explicitly trackable. It offers substantial improvement of computational performance over current tools, and is highly scalable.

In LC-MS metabolomics, a sample is scanned by mass spectrometer consecutively during the chromatography, generating a time series of spectra, each containing a list of ions with mass to charge ratio (m/z) and intensity values. The goal of data processing is to report a quantitative value per metabolite feature per sample, which is a proxy of biological concentration. Multiple software tools have been developed for LC-MS metabolomics data processing over the years, and the most widely used are XCMS and MZmine (Smith et al, 2006, Katajamaa et al, 2006, Pluskal et al, 2010, Du et al, 2020, Yu et al, 2013, Melamud et al, 2010, Rurik et al, 2020). XCMS is also wrapped into numerous workflows and is the main choice in cloud environments (Tautenhahn et al, 2012, Delabriere et al, 2021, Pang et al, 2021). Most these software tools follow a similar framework: building ion chromatogram, detection of elution peaks, alignment of retention time in liquid chromatography, and correspondence of peaks across samples. The design was optimized when instrument resolution was limited and sample numbers were small, not taking advantage of the ultrahigh resolution of modern instruments and the statistical patterns in larger data. The correspondence step is error prone because a feature may only be present as a high-quality peak in a subset of samples, and the computational problem is complicated by missing data, low-quality data, m/z alignment and retention time alignment.

Asari uses a concept of "composite map” to look for peak patterns in cumulative data (**Figure 1A**). A specific form of a metabolite is observed in LC-MS as an elution peak. When the same metabolites exist in multiple biological samples, such peaks are seen recurrently in nearly identical m/z and similar elution profiles. When these corresponding signals from each sample are superimposed and summed up, the observed peaks become representative of all samples (**Figure 1B**). The "composite map" is a complete list of these composite chromatograms. With this approach, peak detection is no longer required on individual samples. It can be done on the composite map, then the peak area is looked up in each individual sample and reported as feature intensity values (**Figure 1A**). This leads to a significant performance gain, by not repeating the computational cost of peak detection on all individual samples. Because the composite map has higher signals than any individual sample, the quality of peak detection is often improved. Even a peak is only present in a single sample, it will be detected and reported in asari, which is important to applications such as personalized medicine and exposomics.

The implementation of composite map is facilitated by a set of transparent data structures. Mass tracks are extracted ion chromatograms spanning the full range of LC, therefore each mass track has a unique m/z within a sample (**Figure S1**). A MassGrid records the alignment of mass tracks across samples. A feature is defined at experiment level, and elution peaks are defined at sample level. A metabolite may have multiple degenerate features due to isotopes, adducts, neutral loss and fragments, which are grouped by an "empirical compound”. An empirical compound is a computational unit for a tentative metabolite, since the experimental measurement may not separate compounds of identical mass (isomers). Asari explicitly links mass track, peak, feature and empirical compound, so that each processing step can be traced and verified. These data structures are exported as JSON or text tables. An interactive dashboard can be launched after data are processed, to allow users to visually inspect data and feature quality easily (**Figures S2, S3**).

The ability to verify feature quality is a priority in asari. Besides peak shape and signal-to-noise ratio (SNR), we have implemented a set of selectivity metrics: mSelectivity is how distinct are m/z measurements (**Figure 1C**), and cSelectivity is how distinct are chromatograhic elution peaks (**Figure 1D**). A derivative of mSelectivity is dSelectivity, applied to how distinct are database records. In feature tables generated by asari, the values of SNR, cSelectivity and peak shape are usually sufficient to judge the quality of LC-MS features.

We demonstrate the results of asari on four datasets generated in our lab (HZV029, MT02, SZ22, BM21) and three public datasets (SLAW as described in Delabriere et al, 2021 as LargeQE, ST001667 and ST001237). They are compared to XCMS, the current leading software. The HZV029 dataset contains 268 data files, from two QC samples that were analyzed repeatedly over 17 batches. The number of features detected in a LC-MS metabolomics experiment is dependent on how parameters allow low-quality peaks to be counted (Myer et al, 2017). Therefore, the comparison first focuses on features of high intensity, and the majority of XCMS features are found in asari (912 out of 1091, **Figure 2A** left). When the data are further filtered by 40% presence across files, all but 19 features from XCMS are found in the result by asari (**Figure 2A** right). Investigation of these 19 features revealed that 10 were present in asari features that did not pass the average height of 1E6, and the remaining 9 features were not deemed of good quality (see Methods). The intensity values of the common features are in good agreement (**Figure 2B**). Besides these data from Orbitrap platforms, similar agreement is seen in Q-TOF data (**Figure S4**).

The MT02 dataset contains the widely used human plasma reference sample NIST SRM 1950. The overall features in this sample detected by both asari and XCMS have consistent values (**Figure 2C**). To establish the true positive features, we referred to the previously reported metabolites in this sample (Simon-Manso et al, 2013), and curated a list of features that were manually verified in raw data (**Table S1**). Both asari and XCMS successfully detected all these 39 “ground truth” features (**Figure 2D**). In the SZ22 dataset, ground truth was established by credentialing in E. coli (similar to Mahieu et al, 2014). A subset of E. coli metabolites were labeled by ^{13}C isotope during the cell culture, and they were selected by elevated ^{13}C/^{12}C ratio and manual inspection of raw data (**Table S2**). Asari successfully detected 71 out of 74 of these credentialed features (**Figure 2E**). Two of the missed features were of low intensity and one of incomplete elution peak. These data indicate that the feature detection by asari is at least on par with XCMS performance.

Reproducibility of feature quantification (also called semi- or relative quantification, to distinguish from targeted methods) is largely driven by experimental variations, while the processing software plays a partial role. Because the HZV029 dataset contains many repeated measurements of the same material, we calculated their pairwise Pearson correlations between samples (**Figure 2F**) and coefficient of variation of features (**Figure 2G**) as metrics of reproducibility. When features are binned by the asari quality metrics of SNR, peak shape and cSelectivity, the top features show better reproducibility (**Figure 2F**). XCMS performed not as well in these metrics of reproducibility, likely due to more missing values (not shown). Of note, the more important contributions to reproducibility by asari reside in its trackable steps, few parameters, transparently linked data structures and the visual dashboard where users can easily verify results. In asari, the only parameter requiring user attention is the mass precision (default at 5 part per million). This eliminates many reproducibility problems in complicated parameter setting in other tools.

To further investigate the performance in quantification, we designed an experiment where human plasma and vegetable juice were mixed by varying ratios (BM21 dataset, **Figure 3A**). Therefore, a subset of features are expected to have their peak areas correlated with the mixing ratio. Overall, 8,222 features were detected by both XCMS and asari in the BM21 dataset, whereas asari has better quantification as indicated by more features with correlation coefficient > 0.9 (**Figure 3B**).

To test the computational efficiency, multiple datasets were processed by both asari and XCMS, and asari provides significant improvement of CPU time over XCMS by 1~2 orders of magnitude (**Figure 3C**). When tested on the SLAW dataset using varying sample numbers, the CPU time and memory use is mostly a linear function of sample numbers (**Figure 3D, E**). The results indicate that the performance gap between XCMS and asari widens for larger studies. XCMS can also become more complicated if it goes beyond simple workflows or large studies are processed (Delabriere et al, 2021). The full SLAW dataset of > 2,000 samples was processed by XCMS in the previous study on a cluster node of 15 CPU cores in 7~12 hours. Now it takes asari ~1 hour on a regular laptop computer.

In summary, the development of asari has significantly contributed to the reproducible data in metabolomics, by a full set of linked and transparent data structures in all processing steps. This allows developers to trace, debug and optimize the process into the future. The end users can navigate and verify features by interactive visualization of extracted ion chromatograms in asari dashboard. Asari has delivered a new generation of computational performance, which is necessary for the future growth of metabolomics. Asari has been mostly tested on Orbitrap platforms. Community involvement will be important to cover the diverse platforms and methods in metabolomics. Asari is free and open-source, and its modular design enables easy reuse of the code for many tasks in computational metabolomics.

Figures

Figure 1. Algorithmic designs and quality metrics in asari.

A) Asari takes centroid mzML files as input, and build chromatograms for each as mass tracks. To prioritize modern mass resolution, m/z alignment is performed first to form a MassGrid, aided by isotopic landmarks. The retention time (RT) alignment is based on LOWESS regression, using a subset of high-quality elution peaks. Elution peak detection is performed on the composite mass tracks, and feature table is generated by looking up the corresponding peak areas in each individual sample. Annotation groups degenerate features into empirical compounds, and reference databases are used to match the m/z values in empirical compounds.

B) The "composite map" is a representation of data from all samples, by adding up the signals in corresponding mass tracks after RT alignment.

C) Illustration of mSelectivity (y-axis) as a function of neighboring m/z values. Each dot represents a m/z feature, and its mSelectivity value depends on the horizontal distance to neighbor features. The error in matching m/z values is modeled as a gaussian distribution dependent on mass precision, and mSelectivity is low when a feature has neighbors with close m/z values.

D) Chromatographic peak selectivity (cSelectivity) is defined by the fraction of the data points in all peaks above 1/2 this peak height and all data points above 1/2 this peak height. cSelectivity is 1 when the chromatogram has no noise above the half height of any peak.

Figure 2. Evaluation of asari feature detection and reproducibility.

A) Overlap between asari and XCMS on HZV029 dataset. Similar parameters were applied to both software tools: min intensity 1000, 5 ppm mass accuracy. In XCMS, centwave window is set at (1, 30), min peak height at 1E6. Because asari has no minimal peak height requirement on individual samples, the features are filtered by average peak height above 1E6, which is more stringent and results in fewer features. The common (matched within 5 ppm and 10 seconds) and unique numbers of features are shown on the left. When further filtered by the presence in at least 40% of samples, the common and unique numbers of features are shown on the right. The common numbers differ between two tools because of decisions in peak splitting or merging.

B-C) Scatter plot of the $\log_2$ peak areas of common features between the two tools. B) corresponds to the right panel in A) on a random sample in HZV029. C) corresponds to a NIST SRM 1950 reference sample. R value shown is correlation coefficient in Pearson correlation.

D-E) Detected features on ground truth datasets in NIST SRM 1950 reference sample (D) and credentialed E. coli samples (E).

F) Of HZV029 asari features, 4,746 have SNR > 1E3, among which 1,187 are denoted as medium quality for peak shape < 0.95. A set of 1,005 features with SNR > 1E4, peak shape > 0.95 and cSelectivity > 0.99 are denoted as top quality. The heatmaps show the reproducibility of randomly selected 32 Qstd samples, colored by their Pearson correlation coefficients.

G) Reproducibility across 17 batches is shown by the distribution of coefficients of variation of the top features and medium features in all 184 Qstd samples. The feature data are not normalized or batch corrected.

Figure 3. Evaluation of quantification and computational performance.

A) Design of the BM21 dataset, by varying mix ratios between human plasma and vegetable juice. A well quantified metabolite is expected to show good correlation between the mixing ratios and the reported peak areas, as exemplified by the feature on top (m/z 189.1232, 159 seconds). Asari calculates peak area differently from XCMS, resulting in higher values in Orbitrap data.

B) Overall quantification results in the BM21 dataset, shown as feature numbers binned by Pearson correlation coefficients between peak areas and sample mixing ratios.

C) Computational performance in user CPU time (equivalent to single core) by asari and XCMS on different datasets (sample numbers show in parentheses on X-axis). Y-axis is in $\log_{10}$ scale. The annotation step is included in asari not in XCMS.

D-E) CPU time and wall clock time (D) and memory (E) used by asari and XCMS on the SLAW dataset using varying number of samples.

Supplements

Figure S1 Illustration of mass tracks in a single sample. The region has 7 mass tracks marked by green boxes spanning horizontally, each of a unique m/z value. A peak is detected from the track indicated by the yellow arrow.

Figure S2. Screen shot of asari Dashboard: feature browser.

Figure S3. Screen shot of asari Dashboard, view of a mass track.

Figure S4. Consistency of feature peak areas shown on an Agilent Q-TOF dataset (ST001667).

Table S1. Manually verified true features in NIST SRM 1950.
Potentially redundant isomers are colored. Since the goal here is not metabolite identification, but to test if software detects the presence of a real feature, the isomers are not distinguished in experimental data.

Table S2. Manually verified true features in credentialed E. coli samples.

References:

Delabriere, A., Warmer, P., Brennsteiner, V. and Zamboni, N., 2021. SLAW: A Scalable and Self-Optimizing Processing Workflow for Untargeted LC-MS. Analytical Chemistry, 93(45), pp.15024-15032.

Du, X., Smirnov, A., Pluskal, T., Jia, W. and Sumner, S., 2020. Metabolomics data Preprocessing using ADAP and MZmine 2. In Computational Methods and Data Analysis for Metabolomics (pp. 25-48). Humana, New York, NY.

Katajamaa, M., Miettinen, J. and Orešič, M., 2006. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics, 22(5), pp.634-636.

Li, S., Sullivan, N.L., Rouphael, N., Yu, T., Banton, S., Maddur, M.S., McCausland, M., Chiu, C., Canniff, J., Dubey, S. and Liu, K., 2017. Metabolic phenotypes of response to vaccination in humans. Cell, 169(5), pp.862-877.

Mahieu, N.G., Huang, X., Chen, Y.J. and Patti, G.J., 2014. Credentialing features: a platform to benchmark and optimize untargeted metabolomic methods. Analytical Chemistry, 86(19), pp.9583-9589.

Melamud, E., Vastag, L. and Rabinowitz, J.D., 2010. Metabolomic analysis and visualization engine for LC− MS data. Analytical chemistry, 82(23), pp.9818-9826.

Myers, O.D., Sumner, S.J., Li, S., Barnes, S. and Du, X., 2017. Detailed investigation and comparison of the XCMS and MZmine 2 chromatogram construction and chromatographic peak detection methods for preprocessing mass spectrometry metabolomics data. Analytical Chemistry, 89(17), pp.8689-8695.

Pang, Z., Chong, J., Zhou, G., de Lima Morais, D.A., Chang, L., Barrette, M., Gauthier, C., Jacques, P.É., Li, S. and Xia, J., 2021. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic acids research, 49(W1), pp.W388-W396.

Pluskal, T., Castillo, S., Villar-Briones, A. and Orešič, M., 2010. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC bioinformatics, 11(1), pp.1-11.

Rurik, M., Alka, O., Aicheler, F. and Kohlbacher, O., 2020. Metabolomics data processing using OpenMS. Computational Methods and Data Analysis for Metabolomics, pp.49-60.

Simon-Manso, Y., Lowenthal, M.S., Kilpatrick, L.E., Sampson, M.L., Telu, K.H., Rudnick, P.A., Mallard, W.G., Bearden, D.W., Schock, T.B., Tchekhovskoi, D.V. and Blonder, N., 2013. Metabolite profiling of a NIST Standard Reference Material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources. Analytical chemistry, 85(24), pp.11725-11731.

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R. and Siuzdak, G., 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Analytical chemistry, 78(3), pp.779-787.

Tautenhahn, R., Patti, G.J., Rinehart, D. and Siuzdak, G., 2012. XCMS Online: a web-based platform to process untargeted metabolomic data. Analytical chemistry, 84(11), pp.5035-5039.

Wishart, D.S., 2020. Metabolomic data exploration and analysis with the human metabolome database. In Computational Methods and Data Analysis for Metabolomics (pp. 165-184). Humana, New York, NY.

Yu, T., Park, Y., Li, S. and Jones, D.P., 2013. Hybrid feature detection and information accumulation using high-resolution LC–MS metabolomics data. Journal of proteome research, 12(3), pp.1419-1427.

Data availability: The datasets MT02 and SZ22 are available at https://github.com/shuzhao-li/data. The BM21 and HZV029 datasets are in the submission process to Metabolomics Workbench (https://www.metabolomicsworkbench.org/), and will be made publicly available at the time of publication.

The public datasets used in this work are under Study IDs ST001667 and ST001237 on Metabolomics Workbench. The large SLAW dataset was retrieved from MassIVE by study ID MSV000086486 (Delabriere et al, 2021).

Code availability: The asari source code is available at GitHub, https://github.com/shuzhao-li/asari, and as a Python package via https://pypi.org/project/asari-metabolomics/.

Acknowledgements: This work was in parted funded by NIH grants (to SL) U01 CA235493 (NCI) and R01 AI149746 (NIAID).

Author contributions: S.L. designed the study, wrote the asari software and the manuscript. A.S. and S.L. performed data analysis and software testing. M.T. performed the experiments of HZV029, MT02 and BM21. S.Z. performed the experiment of SZ22.

Methods

Software design of asari. Asari is written in Python 3, and can be used as a standalone command line tool or imported as a package. Its library dependency includes numerical computing via numpy and scipy, data wrangling via pandas, and visualization via panel and hvplot. Pymzml is used to parse mzML format. Data structures, annotation, search and chemical calculation make use of our supporting packages metDatamodel, mass2chem and jms-metabolite-services. Implementation of new and previous algorithms was coded from ground up, where numerous details contributed to the computing speed, e.g., discrete mathematics is preferred over continuous curves, and intermediary indexing and caches are employed. Processed mass tracks are cached on disk to reduce memory footprint. Mass tracks are explicitly linked with features and peaks, and the information is exported as JSON in asari output. The quality metric mSelectivity is used internally and not exported by default. The annotation and search functions are generic to accommodate reference databases, and the default is HMDB (Wishart 2020).

Evaluation of feature detection and computational performance. The features from different software tools in the same data are considered matched when their m/z values are within 5 ppm and retention times are within 10 seconds. The results from XCMS did not use peak filling, which often creates artifacts. The merging of adjacent peaks in XCMS is dependent on input parameters, and often resulted more split peaks than the results in asari. The 9 XCMS features that were not accepted by asari (Figure 2A) are (m/z @ retention time in seconds): 129.1022@25**,** 28.0197@17, 210.9937@16, 174.1854@23, 256.2999@23, 156.1133@14, 129.1023@13, 120.0808@15, 100.1121@15.

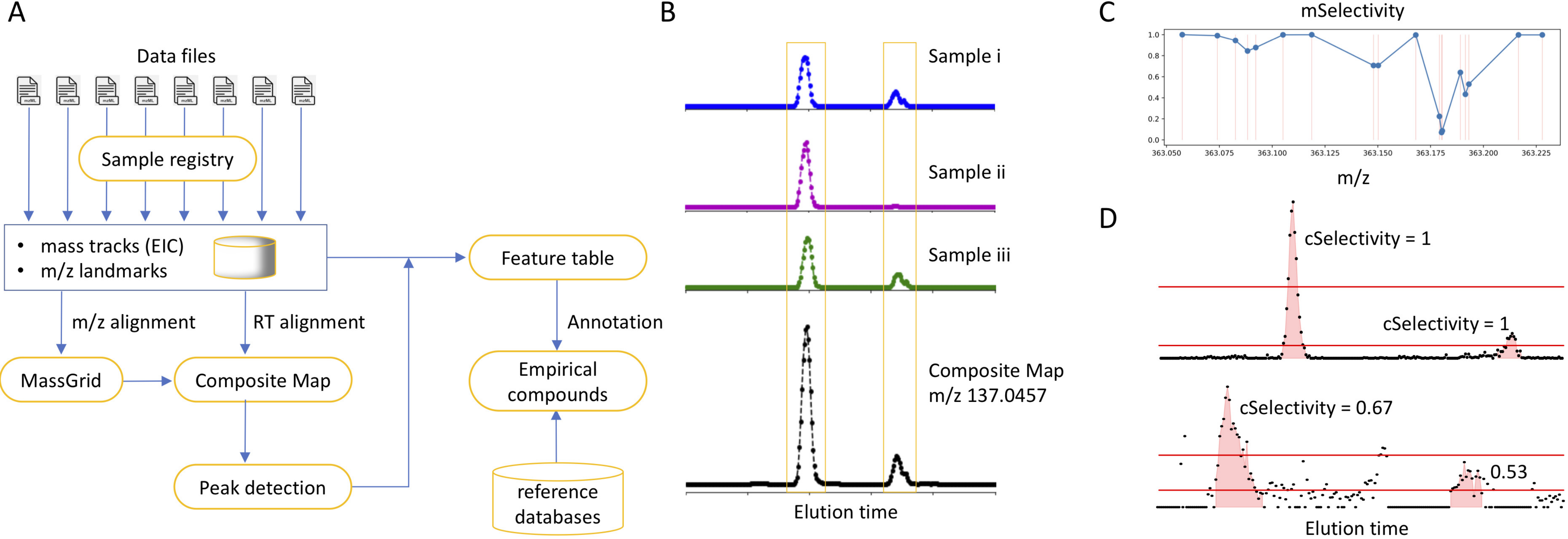
The “ground truth” features in the NIST SRM 1950 sample were manually verified and counted as 39 true positive m/z features. The reported isomers are not distinguished here since our retention time is not comparable to the previous publication. A positive match to either asari or XCMS results requires a feature to be within 5 ppm. For the credentialed E. coli samples, a feature is considered to be true positive when a) it is present in all six samples, b) presence of the isotopic peak by 1.003355 m/z difference at the same retention time, c) the $^{12}C/^{13}C$ ratio > 1 in the unlabeled samples, and d) the $^{12}C/^{13}C$ ratio is > 2-fold higher in the labeled samples than unlabeled samples. The difference from the Mahieu et al (2014) paper was due to that we analyzed the labeled and unlabeled samples separately, while the previous work mixed them at specific ratios.

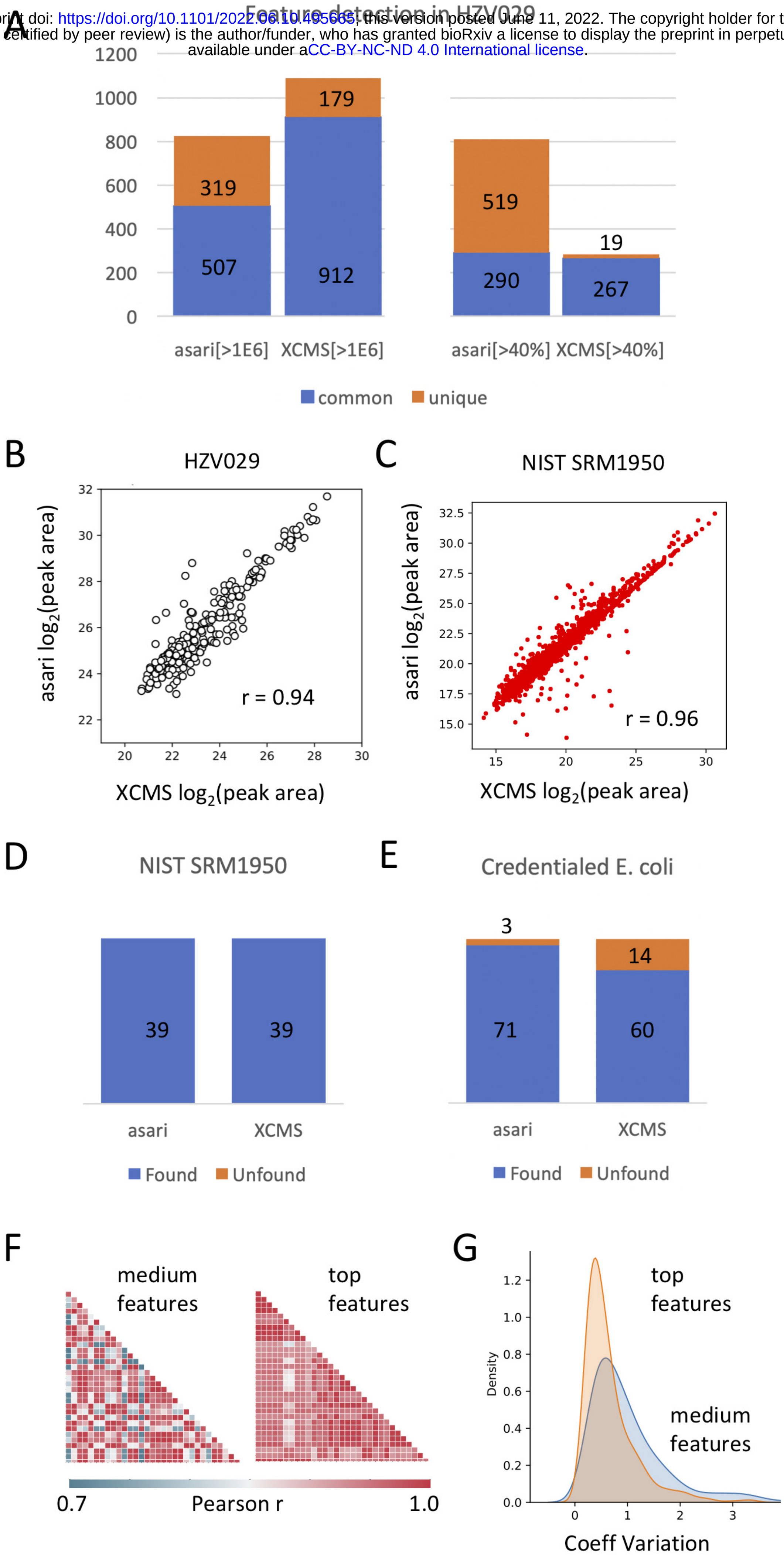
The evaluation of computational performance was performed on a desktop computer with Intel i7-8809G CPU and 32 GB of memory, running Mint Linux 20.2. The asari version was 1.9.2. The XCMS version was 3.18.0. The R script for XCMS is provided in asari repository (https://github.com/shuzhao-li/asari) under doc/ directory. The time and memory use was measured by `/usr/bin/time –p`, and "User time" was used as CPU time (equivalent to CPU time used on a single core).

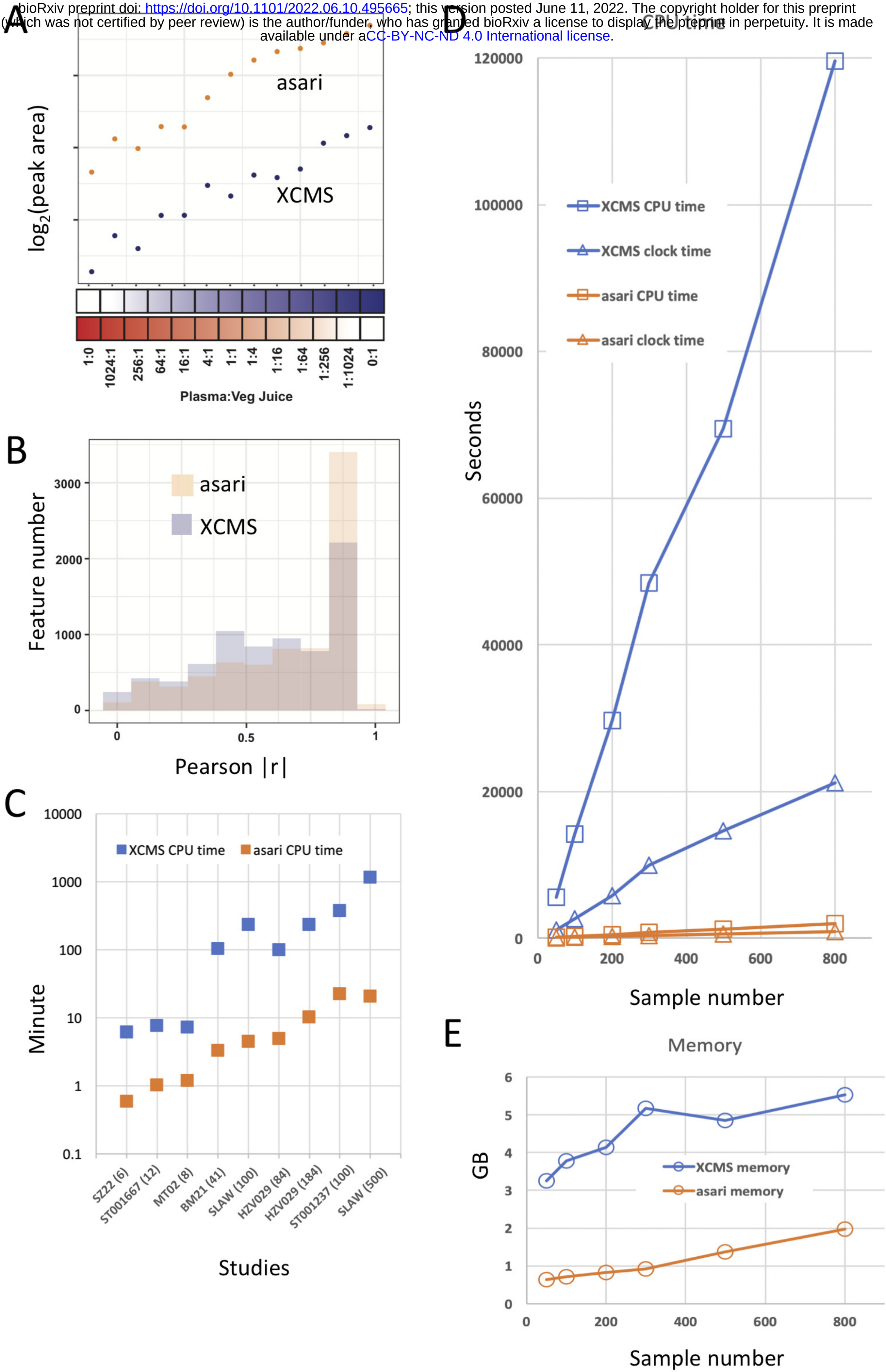
LC-MS metabolomics experiments. The human plasma samples used in this study were a pooled deidentified QC sample in a vaccination cohort, NIST SRM 1950 (https://www-s.nist.gov/srmors/view_detail.cfm?srm=1950), and a commercial reference sample Qstd (Sterile Filtered Human Plasma (K2) EDTA, Equitech Bio, Inc. KERRVILLE, TEXAS). The BM21 experiment included a serial mixture of human plasma (Qstd) and vegetable juice, at the ratio of 1024:1, 256:1, 64:1, 16:1, 4:1, 1:1, 1:4, 1:16, 1:64, 1:256 and 1:1024. Along with the 11 serial mixture samples, 100% vegetable juice and 100% plasma were also included. All samples were analyzed in triplicates, while one replicate was used for data analysis in this study for simplicity. The dry extracts of unlabeled and ^{13}C labeled *E. coli* (Cambridge Isotope Laboratories, Inc.; Catalog number: MSK-CRED-DD-KIT) were reconstituted in 100 µL of ACN/H_2O (1:1, v/v) then sonicated (10 mins) and centrifuged (10 mins at 13,000 rpm and 4°C) before overnight incubation at 4°C. The supernatant for each ^{12}C/^{13}C *E. coli* extract was collected and then prepared for LC-MS analysis. These samples were run in triplicates.
Metabolites extraction was carried out by protein precipitation technique using extraction solvent, acetonitrile:methanol (8:1, v/v) containing 0.1% formic acid and isotope labelled Trimethyl-13C3]-caffeine, [13C5]-L-glutamic acid, [15N2]-Uracil, [15N,13C5]-L-methionine, [13C6]-D-glucose and [15N]-L-tyrosine as spike-in controls. 30 µl of plasma samplewas taken and 60 µl of extraction solvent was added. Extraction blanks were also prepared to remove features of non-biological origins. All samples were vortexed and incubated with shaking at 1000 rpm for 10 min at 4°C followed by centrifugation at 4°C for 15 min at 15,000 rpm. The supernatant was transferred into mass spec vials and 2 µl injected into UHPLC-MS.
All samples were maintained at 4 °C in the autosampler, and analyzed using a Thermo Scientific Orbitrap ID-X Tribid Mass Spectrometer coupled to a Thermo Scientific Transcen LX-2 Duo UHPLC system, with a HESI ionization source, using positive and negative ionizations. The MS settings are: spray voltage, 3500 V; sheath gas, 45 Arb; auxiliary gas, 20 Arb; sweep gas, 1 Arb; ion transfer tube temperature, 325 °C; vaporizer temperature, 325 °C; mass range, 80-

1000 Da; maximum injection time, 100 ms. The resolution was set at 120,000 in the HZV029 experiment, 60,000 in the BM21 and SZ22 experiments.

Data were acquired using hydrophilic interaction liquid chromatography (HILIC) positive and reversed phase (RP) negative polarities in full scan mode with mass resolution of 120,000 simultaneously. An AccucoreTM-150-Amide HILIC column (2.6 μm, 2.1 mm x 50 mm) and a Hypersil GOLDTM RP column (3 μm, 2.1 mm x 50 mm) maintained at 45 ºC were used for chromatographic separation. 0.1% formic acid in water and 0.1% formic acid in acetonitrile were used as mobile phase A and B respectively for RP acquisition. 10 mM ammonium acetate in acetonitrile:water (95:5, v/v) with 0.1% acetic acid as mobile phase A and 10 mM ammonium acetate in acetonitrile:water (50:50, v/v) with 0.1% acetic acid as mobile phase B were used for HILIC method. For HILIC acquisition, following gradient was applied at a flow rate of 0.55 ml/min: 0-0.1 min: 0% B, 0.10-5.0 min: 98% B, 5.00-5.50 min: 0% B and 4.5 min for cleaning and equilibration of column. For RP column, following gradient was applied at a flow rate of 0.4 ml/min: 0-0.1 min: 0% B, 0.10-1.9 min: 60% B, 1.9-5.0 min: 98% B, 5.00-5.10 min: 0% B and 4.9 min cleaning and column equilibration. The chromatographic run time was 5 min followed by 5 min washing step after each sample.


A
Data files
mzML
Sample registry
mass tracks (EIC)
m/z landmarks
m/z alignment
RT alignment
MassGrid
Composite Map
Peak detection
Feature table
Annotation
Empirical compounds
reference databases
B
Sample i
Sample ii
Sample iii
Composite Map m/z 137.0457
Elution time
C
mSelectivity
1.0
0.8
0.6
0.4
0.2
0.0
363.050
363.075
363.100
363.125
363.150
363.175
363.200
363.225
m/z
D
cSelectivity = 1
cSelectivity = 1
cSelectivity = 0.67
0.53
Elution time


A
Feature detection in HZV029
1200
1000
800
600
400
200
0
319
507
179
912
519
290
19
267
asari[>1E6]
XCMS[>1E6]
asari[>40%]
XCMS[>40%]
common
unique
B
HZV029
asari $\log_2$(peak area)
XCMS $\log_2$(peak area)
r = 0.94
C
NIST SRM1950
asari $\log_2$(peak area)
XCMS $\log_2$(peak area)
r = 0.96
D
NIST SRM1950
39
39
asari
XCMS
Found
Unfound
E
Credentialed E. coli
3
71
14
60
asari
XCMS
Found
Unfound
F
medium features
top features
0.7
Pearson r
1.0
G
top features
medium features
Density
Coeff Variation


A
log$_2$(peak area)
asari
XCMS
1:0
1024:1
256:1
64:1
16:1
4:1
1:1
1:4
1:16
1:64
1:256
1:1024
0:1
Plasma:Veg Juice
B
Feature number
3000
2000
1000
0
asari
XCMS
0
0.5
1
Pearson |r|
C
Minute
10000
1000
100
10
1
0.1
XCMS CPU time
asari CPU time
SZ22 (6)
ST001667 (12)
MT02 (8)
BM21 (41)
SLAW (100)
HZV029 (84)
HZV029 (184)
ST001237 (100)
SLAW (500)
Studies
D
CPU time
Seconds
120000
100000
80000
60000
40000
20000
0
XCMS CPU time
XCMS clock time
asari CPU time
asari clock time
0
200
400
600
800
Sample number
E
Memory
GB
6
5
4
3
2
1
0
XCMS memory
asari memory
0
200
400
600
800
Sample number