The authors regret to inform that there was an unintentional train-test leakage in our machine learning experiment. The reason for such leakage was not taking into account highly correlated LC-MS repetions of individual physical samples. In order to correct train-test leakage, splitting into train and test was done such that either all repetitions from one physical sample were put in train set or one (random) of them was put in test set while the rest repetitions were ignored. Overall, accuracy and F1-score of our models was reduced by around 2% to 15% for different models. All the results that were affected by this leakage were recalculated the following corrections are suggested (in addition to tables and figures) We used blue font color for our corrections to the relevant parts of the original text.

Corrigenda

1. (abstract) Even with elimination of all retention time values accuracies of up to 96% and 92% were achieved on validation set for plant species and plant organ identification respectively.
   Even with elimination of all retention time values accuracies of up to around 85% were achieved on validation set for plant species and plant organ identification.
2. (results) Encoded data vectors with 25 variables were used to train logistic regression and continuous Bayes classifiers (both Naive Bayes and hybrid Bayesian Network) with resulting identification accuracy of 96% and 84–87% on Test 1 respectively. All above-mentioned models showed accuracy of 68–77% on Test 2.
   Encoded data vectors with 25 variables were used to train logistic regression and continuous Bayes classifiers (both Naive Bayes and hybrid Bayesian Network) with resulting identification accuracy of 85% and 68-69% on Test 1 respectively. All of the above-mentioned models showed accuracy of 68-75% on Test 2.
3. (results) According to the Table 1 Part 1, classifier based on Tucker decomposition with principal angle distance measure performs well (93% and 86% respectively for Test 1 and Test 2).
   According to the Table Table 1 Part 1 Part 1, classifier based on Tucker decomposition with principal angle distance measure performs well (78% and 84% respectively for Test 1 and Test 2).
4. (discussion) Results show that with careful selection of feature space and model tuning it is possible to achieve up to 96% classification accuracy even with large and heterogeneous negative class.
   Results show that with careful selection of feature space and model tuning it is possible to achieve up to 85% classification accuracy even with large and heterogeneous negative class.
5. (discussion) The most obvious increase was shown by BN on Test 2, where emergence of correct labels in Top5 jumped by more than 20% compared to "winner takes all" approach. Although exact accuracy values may drop when using larger and more diverse datasets, this shows great potential of discrete BNs in such applications. All in all, TopN representation can be considered a more preferable way of output – narrowing possible candidates to 3–5 with 95% or more accuracy can be more beneficial than 80% accurate single candidate species.
   The most obvious increase was shown by bayesian networks on Test 2, where emergence of correct labels in Top5 jumped by around 20% compared to "winner takes all" approach. Although exact accuracy values may drop when using larger and more diverse datasets, this shows great potential of BNs in such applications. All in all, TopN representation can be considered a more preferable way of output – narrowing possible candidates to 3-5 with ∼90% accuracy can be more beneficial than 75% accurate single candidate species.
6. (discussion) Algorithms showed high distinguishing ability between most classes (up

to 92% accuracy), excluding very similar pair of classes (roots, roots and rhizomes).
Algorithms showed high distinguishing ability between most classes (up to 86% accuracy), excluding very similar pair of classes (roots, roots and rhizomes).

7. Table 1 , Figure 2, Figure 5, Figure 6, SupplementaryS1
Were remade and complied accoring to recalculation results and are included in this corrigendum.

Respectfully,

Pavel Kharyuk
E-mail: kharyuk.pavel@gmail.com
Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Building 3, Moscow 143026, Russia; Institute of Numerical Mathematics of Russian Academy of Sciences

Dmitry Nazarenko
E-mail: dmitro.nazarenko@gmail.com
Lomonosov Moscow State University, Faculty of Chemistry, Moscow, 119991, Russia

Ivan Oseledets
E-mail: i.oseledets@skoltech.ru
Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Building 3, Moscow 143026, Russia. Institute of Numerical Mathematics, Russian Academy of Sciences. Gubkina St. 8, Moscow 119333, Russia.

Igor Rodin
E-mail: rodin@analyt.chem.msu.ru
Lomonosov Moscow State University, Faculty of Chemistry, Moscow, 119991, Russia

Oleg Shpigun
E-mail: shpigun@analyt.chem.msu.ru
Lomonosov Moscow State University, Faculty of Chemistry, Moscow, 119991, Russia

Andrey Tsitsilin
E-mail: vilarnii@mail.ru
All Russian research institute medicinal and aromatic plants (VILAR), Moscow, 117216, Russia

Mikhail Lavrentyev
E-mail: mihaillavrentev@yandex.ru
Saratov State University, Department of Botanics and Ecology, Saratov, 410012, Russia