

1 Identification and quantification of meat product 2 ingredients by whole-genome metagenomics (All-Food- 3 Seq) 4

5 Sören Lukas Hellmann^{1#}, Fabian Ripp^{1,2#}, Sven-Ernö Bikar³, Bertil Schmidt⁴, René
6 Köppel⁵, Thomas Hankeln^{1*}

7
8 ¹ Institute of Organismic and Molecular Evolution, Molecular Genetics & Genome
9 Analysis, Johannes Gutenberg University Mainz, Mainz, Germany

10 ² current address: MVZ Labor Volkmann, Karlsruhe, Germany

11 ³ StarSEQ GmbH, Mainz, Germany

12 ⁴ Institute of Informatics, Johannes Gutenberg University Mainz, Mainz, Germany

13 ⁵ Official Food Control Authority of the Canton Zurich, Zurich, Switzerland
14

15 *Correspondence: Prof. Dr. Thomas Hankeln, Institute of Organismic and Molecular
16 Evolution, Molecular Genetics & Genome Analysis, Johannes Gutenberg University
17 Mainz, J. J. Becher-Weg 30A, D-55099 Mainz, Germany, phone +49 6131 393 3277,
18 email hankeln@uni-mainz.de

19 # equal contribution
20

21 Keywords: Food Metagenomics, Species Identification, Doner Kebap, Read Mapping,
22 Next-Generation-Sequencing
23

24 Acknowledgements

25 TH and SLH gratefully acknowledge funding by the Federal Office for Agriculture and Food
26 (project ID: 2816503814).

27 **Abstract**

28 Complex food matrices bear the risk of intentional or accidental admixture of non-
29 declared species. Moreover, declared components can be present in false
30 proportions, since expensive taxa might be exchanged for cheaper ones. We have
31 previously reported that PCR-free metagenomic sequencing of total DNA extracted
32 from sausage samples combined with bioinformatic analysis (termed All-Food-Seq,
33 AFS), can be a valuable screening tool to identify the taxon composition of food
34 ingredients. Here we illustrate this principle by analysing regional Doner kebab
35 samples, which revealed unexpected and unlabelled poultry and plant components in
36 three of five cases. In addition, we systematically apply AFS to a broad set of
37 reference meat material of known composition (i.e. reference sausages) to evaluate
38 quantification accuracy and potential limitations. We include a detailed analysis of the
39 effect of different food matrices and the possibility of false-positive sequence read
40 assignment to closely related species, and we compare AFS quantification results to
41 quantitative real-time PCR (qPCR) and droplet digital PCR (ddPCR). AFS emerges
42 as a potent PCR-free screening tool, which can detect multiple target species of
43 different kingdoms of life within a single assay. Mathematical calibration accounting
44 for pronounced matrix effects can significantly improves AFS quantification accuracy.
45 In comparison, AFS performs better than classical qPCR, and is on par with ddPCR.

46 Introduction

47
48 The determination and quantification of food ingredients is an important issue in
49 official food control. The complexity of foodstuff, difficulties in the traceability of
50 trading channels and the globalisation of food markets opens doors for fraud and
51 failures in correct labelling, stocking and processing procedures [1]. Possible
52 consequences for consumers are manifold beginning with compliance of ethical
53 aspects like halal, kosher or vegan over health risks caused by pathogenic organisms
54 to simple deception because of economic reasons. In fact, biological contaminants
55 made up the vast majority of warning notices released by the German authorities [2]
56 between 2011 and 2015. The majority of these cases were provoked by
57 microbiological contaminations or the presence of non-declared allergenic food
58 components. Therefore, food and drug legislation demands proper declaration of
59 ingredients and compliance to storage and transport conditions [3, 4]. To ensure
60 adherence to law and to maintain consumer's safety, there is a growing need for
61 methods that allow for precise determination of food ingredients, ideally spanning all
62 kingdoms of life including plants, animals, bacteria, fungi and perhaps also extending
63 to viruses. A broad palette of analytical methods for analysing foodstuffs has been
64 developed and is routinely applied at official food control laboratories, but also private
65 and industrial control labs. Among these, DNA-based methods like PCR are probably
66 the most widely used technologies, because of their high sensitivity and the
67 possibility to perform quantitative measurements [5–13]. However, even when
68 multiplexed or performed in the meta-barcoding format, PCR-based approaches have
69 the drawback to detect only a limited range of target species and produce assay-
70 dependent amplification biases [14–17].

71
72 We have previously shown that deep metagenomic DNA sequencing of whole-
73 genome DNA from foodstuffs, followed by dedicated bioinformatic analysis, is in
74 principle able to overcome these issues. DNA sequence reads obtained from food
75 can be bioinformatically assigned to existing reference genomes for species
76 identification, and the number of reads successfully assigned to a respective genome
77 can be counted to give a quantitative measure of the species proportions.
78 Importantly, such whole-genome sequencing of foodstuff DNA (termed All-Food-Seq:
79 AFS; [18, 19]) does not require any *a priori* definition of possible target species. AFS
80 can therefore be viewed as a screening method, which theoretically can detect an
81 infinite spectrum of diverse species, being only limited by our current knowledge of
82 genomes, as represented in the fast-growing public sequence databases. The
83 “identification plus quantification” principle based on read-assignment and read-
84 counting has been successfully demonstrated so far as a proof-of-principle in a
85 limited number of foodstuff samples, i.e. sausages of pre-defined composition
86 prepared as reference material [8, 20]. We therefore decided to further investigate
87 the potential of AFS in a real-life test case, analysing different doner kebab samples
88 obtained from snack bars. We also saw the necessity to evaluate the quantification
89 potential of AFS in more detail. Inferring species proportions from DNA read

90 proportions can be difficult because it may substantially depend on the food
91 composition and processing. As an example, high quality meat may be substituted for
92 in a product by the addition of rind, lard or skin, which could affect DNA amounts per
93 gram tissue, and thus the inference of species proportions within the foodstuff. To
94 study this so-called matrix effect, we have applied the AFS method to an extended
95 set of reference sausage samples, each containing known admixtures of different
96 meat sources, but prepared according to different recipes [8, 20]. We compared the
97 AFS quantification results to those obtained by qPCR and ddPCR on the same
98 samples and evaluated the effects of matrix composition.

99 **Materials and Methods**

100 **Food samples and DNA extraction**

101 Doner kebab samples were purchased at five snack bars distributed in the Rhine-
102 Main area. All meat pieces were identified by eye and selected by sterile forceps for
103 subsequent homogenization in large volume using a standard kitchen device. About
104 1 g of the homogenized matrix, which looked surprisingly different (ranging between
105 an oily and granular texture), was taken for subsequent DNA isolation using the
106 Wizard Plus Miniprep DNA purification system (Promega, Madison, USA) according
107 to the manufacturer's protocol. DNA was quantified by Qubit fluorometry
108 (ThermoFisher Scientific, Schwerte, Germany).

109 Calibration sausage samples containing admixtures of cattle, chicken, pig, sheep and
110 turkey at defined amounts were produced by a professional butchery and provided by
111 the Official Food Control Authority of the Canton Zürich, Switzerland [8, 20]. The
112 samples were prepared for calibration of foodstuff detection methods and reflect
113 three different recipes of sausage production (Online Tab. S1): AllMeat sausage (Kal
114 A-E: meat), Lyoner-style sausage (KLyo A-D: matrix of meat, rind and lard) and
115 Poultry-Lyoner (KGeflLyo A-D: matrix of meat and skin). Total DNA was extracted out
116 of 200 mg homogenized sausage sample using the Wizard Plus system (Promega,
117 Madison, USA) according to the manufacturer's protocol.

118 **Illumina library preparation and sequencing**

119 Sequencing library preparations and sequencing were performed by a commercial
120 provider (StarSEQ, Mainz, Germany). The Nextera DNA Library Preparation Kit
121 (Illumina, San Diego, USA) was applied following the manufacturer's instructions.
122 Typically, 1 ng of total DNA was used. Sequencing was carried out on an Illumina
123 MiSeq instrument using reagent kit v.2 in 150 bp paired-end (reference sausages)
124 and 50 bp single-end (doner kebab samples) mode, respectively. In principle, both
125 sequencing modes deliver comparably valid results [19]. Between 200 k and 2600 k
126 reads were generated per sample (Online Tab. S1). Adjustments by downsampling
127 were omitted, because our previous analysis showed that read numbers > 100 k
128 produced consistent quantification results independent of dataset size [19]. All
129 datasets were quality checked, trimmed and filtered by using FASTQC data
130 evaluation software [21] and trimmomatic v0.33 trimming tool [22]. Datasets have
131 been submitted to the SRA database under the project names PRJNA271645 and
132 PRJEB34001.

133 **Bioinformatic analysis of main ingredients using AFS**

134 The AFS read-mapping pipeline was executed with 3 rounds of iterative mapping and
135 step-wise decreased mapping stringency, as described [18, 19]. This strategy allows
136 for a final number of 2 mismatches after mapping step 3. At each round, reads that
137 mapped against one of the provided reference genomes were cumulatively counted
138 and reported on a 1-100% scale to reflect relative species proportions. In the doner

139 kebab screening analysis, sequence reads were mapped against a selection of
140 reference genomes (accession numbers: cattle: NC_037328.1, sheep:
141 NC_040252.1, goat: NC_030808.1, pork: NC_010443.5, horse: NC_009144.3,
142 chicken: NC_006088.5, turkey: NC_015011.2, maize: NC_024459.2 and soy:
143 NC_016088.3. In the quantification analysis of the calibrator sausages, reference
144 genome choice was limited to the animal species cattle, chicken, pig, horse, sheep,
145 goat, water buffalo (accession number: NC_037545.1) and turkey. Goat and water
146 buffalo genomes were added to test the robustness of AFS towards false positive
147 signals to be expected between closely related species. All evaluations were
148 performed on a standard desktop PC (Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz,
149 16GB DDR4 2667 MHz RAM, 256GB SATA SSD, CentOS Linux release 7.6.1810).

150 Reads that did not match, very likely originate from species not provided as a
151 reference during the AFS mapping step. These unmapped reads (around 3 % per
152 sample), often representing spice plants and microbiota [18], did not undergo further
153 metagenomic analysis in the present study, since the prime goal was to evaluate the
154 quantification properties of AFS for the main meat components.

155 **Calculation of false-positive read assignments**

156 In order to determine false-positive read assignment rates for the tested species, in
157 particular the closely related cattle-buffalo, chicken-turkey and goat-sheep, we
158 created *in silico* datasets of different proportions of reads for each species with the
159 corresponding related species being absent. To this end, we used whole-genome
160 shotgun datasets from the SRA (SRR8588004, SRR9663406, SRR8442931,
161 SRR8560982, SRR6470934) and performed data pre-processing as described for
162 the reference sausages. For each species, proportions of 1, 5, 10, 25, 50, 75 and
163 100 % were extracted using the reformat tool from the BBMap suite [23] and
164 complemented to 1 mio reads with the non-related plant species rice (accession
165 number: NC_008394.4). For the cattle-buffalo species pair, we only inspected the
166 false-positive rate of buffalo assignments given a cattle ingredient, as the opposite
167 direction is irrelevant to food safety inspections in our opinion. To investigate the
168 effect of sequence read length on false positive mapping, all generated datasets
169 were trimmed using the reformat tool to a length of 50, 100 and 150 bp, respectively.
170 Subsequent AFS analyses were performed as described above with 3 mapping
171 rounds (accepting max. 2 mismatches) against buffalo, cattle, chicken, goat, horse,
172 pork, sheep and turkey genomes.

173 **Results**

174 **AFS screening of species composition in doner kebab samples**

175 Doner kebab samples were obtained from five snack bars in the Rhine-Main area.
176 Their meat components were sampled, homogenized and the extracted DNA
177 sequenced. AFS analysis revealed that samples 2 and 3 were prepared from pure
178 beef, while samples 1, 4 and 5 consisted of beef and turkey, with the latter as the
179 dominant component (Tab. 1a). Samples 1, 3 and 5 revealed measurable amounts of
180 soybean DNA (0.5-0.8 %), and sample 1 additionally contained maize DNA (1.8 %).
181 In samples 2, 3 and 5, we observed that 0.1-0.4 % of sequence reads were assigned
182 to goat and sheep. Since the latter also belong to the family of Bovidae, one could
183 interpret the goat/sheep read assignments as candidate false-positives, produced as
184 a consequence of the phylogenetic relatedness and the presence of conserved
185 genomic elements. However, our detailed evaluation of possible false-positive values
186 (see below; Online Fig. S1b, Online Tab. S2) shows that at least for samples 2 and 3
187 the measured values of goat and sheep are slightly higher than expected for a matrix
188 consisting almost only of cattle. We therefore cannot rule out that small amounts of
189 sheep and goat material were indeed present in these doner samples, allegedly
190 caused by the presence of cheese matrices or due to unknown circumstances during
191 doner production. In contrast, the 0.2 and 0.3 % of chicken reads in samples 1 and 4,
192 which are clearly dominated by turkey, may accordingly be considered false-
193 positives.

194 **AFS quantification of meat ingredients in reference sausages**

195 To specifically study the quantification properties of AFS in a broad set of samples, a
196 total of 13 reference sausage samples (Online Tab. S1), prepared according to three
197 different standard recipes, were sequenced and analysed. Datasets were then
198 studied to evaluate quantification accuracy, the impact of different matrices (i.e. meat,
199 rind, lard and skin), and the probability of false-positive read assignments. AFS
200 results were then compared to quantification data previously obtained by qPCR [8,
201 20] and droplet digital (dd) PCR [13] on the very same sausage samples (Tab. 1b-D).

202 • *AFS quantification accuracy*

203 Our sample set covered expected species proportions from 0.5-80 %. Minimal and
204 maximal expected components varied between the three different sample types
205 (meat-only samples Kal A-E 1-55%, mixed-matrix samples KLYo A-D and KGeflLYo
206 A-D 0.5-80%). It turned out that even the low concentrations of ingredients could be
207 detected by AFS with high accuracy (0.5 ± 0.1 %; 1 ± 0.1 %; 2 ± 0.4 %; 4 ± 0.2 %;
208 5.5 ± 0.5 %). As species concentrations increased, absolute deviations of measured
209 values also increased to a maximum of 20.9 % for the 9-36 % interval and 20.4 % for
210 the 55-80 % interval, respectively (Tab. 1b-D). To compare the performance of AFS
211 for the different sausage types, we summed up the individual species deviations for
212 each sausage individually. Results showed that, omitting any calibration calculations

213 (see below), Kal A-E samples were quantified with overall best results (ranging
214 between 6.7 and 11.1 % deviation), followed by KLy0 A-D (6.8 to 17.0 %) and
215 KGeflLy0 A-D (18.4 to 41.9 %).

216 • *Evaluation of false-positive read assignments between related species*

217 The species assignment of sequence reads in AFS is based on classical read-
218 mapping algorithms involving sequence alignment [18]. This implies the potential
219 danger of a mis-classification if a read contains highly conserved DNA sequences,
220 often present in the genomes of phylogenetically closely related taxa. Of course such
221 false positive assignments could have, if present, an eminent effect on detection
222 accuracy. To evaluate the potential of such false-positive read assignment within
223 AFS, we intentionally included in the read-mapping step the reference genomes of
224 species, which are not present in the sausages, but which are evolutionarily close to
225 the real food components. Specifically, we added the genome of the water buffalo
226 (*Bubalus bubalis*), which shared a common ancestor with cattle 13 mio years ago,
227 and the goat (*Capra hircus*), which diverged from sheep 10 mio years ago ([24];
228 Online Fig. S1a). False-positive signals of buffalo and goat from sausages containing
229 cattle or sheep as real ingredients ranged between 0.0-1.7 % and depended on the
230 amount of the corresponding real ingredient species (Tab. 1b-D). For example, the
231 maximal value of 1.7 % false goat reads was obtained for the Kal D sausage
232 containing 55% sheep.

233 To systematically specify the chance of false-positive read assignments between
234 species pairs in AFS, we simulated read datasets with varying, known amounts of
235 reads from the species in our study and mapped them to the respective reference
236 genomes. The amount of false-positive reads in fact scaled linearly with the real
237 ingredient proportions (Online Fig. S1b, Online Tab S2), allowing us to define
238 threshold values for the respective species pairs. Interestingly, but not unexpectedly,
239 the short 50 bp reads produced markedly higher false-positive values than 100 bp
240 and 150 bp reads. For example, a 100% sheep dataset produced 5.1% false-positive
241 goat assignments with 50 bp read length, but only 2.7 % with 150 bp reads (Online
242 Fig. S1b). Some minor 'asymmetric' quantification results (i.e. chicken against turkey
243 genome versus turkey against chicken genome) could be noted and are probably
244 caused by different qualities of the respective reference genomes. Notwithstanding,
245 these calculated values can now be applied by the AFS user to objectively assign
246 quantification values as potential false-positives, as done above in the case of the
247 doner kebab samples.

248 • *Matrix effects and their possible correction by linear regression*

249 Different types of food matrices can bias quantification analyses because different
250 tissues often contain varying concentrations of DNA, and cellular DNA may also be
251 extracted from them at different efficiencies. To study this effect, we included three
252 types of sausage matrices: the Kal samples, consisting only of pure meat, the KLy0
253 samples, in which pork material was represented by three tissues (meat, rind and

254 lard at a ratio of 1:4:15) and the KGeflLyo sausages, containing chicken material as a
255 1:1 mixture of meat and skin (Online Tab. S1).

256 Specifically for the KGeflLyo sausages with their partial replacement of chicken meat
257 by skin (Online Fig. S2), the chicken component showed a substantial
258 overrepresentation on the DNA level, thus severely compromising the quantification
259 results for this matrix type (independent of whether AFS or PCR methods were
260 applied; comp. Tab. 1). While samples containing meat-only chicken showed minimal
261 deviations of 0.1-0.5 % from expected values, the meat/skin matrix led to an almost
262 proportional overestimation of chicken by 9-20 % (Tab. 1d; Online Fig. S2). A
263 second, but milder effect was noticed for pork as an ingredient, which was
264 systematically underestimated by 2.7-7.8 % in the KLyo A-D, 0.1-0.6 % in the
265 KGeflLyo A-D and 0-5.5 % in the Kal A-E samples, respectively.

266 Assuming that the observed effects represent systematic errors, we decided to
267 normalize our measurements by applying linear regression. We did this for every
268 sample type and species separately to consider both matrix-specific and species-
269 dependent effects (see calibrated AFS values in Tab. 1). In fact, the improvement of
270 the quantification values turned out to be massive, showing that AFS (very much like
271 the PCR methods; comp. [8, 13]) will benefit from the establishment of such matrix
272 calibration factors. Indeed, we were able to correct efficiently for most of the
273 systematic error over a broad range of expected values. Note however, that in some
274 cases (e.g. Kal A and E) deviation slightly increased after the normalisation
275 procedure for the very low expected values of 0.5 and 1 %, respectively (Tab. 1b).

276 • *Limits of detection and precision*

277 Using normalised values gained by linear regression, we calculated the limit of
278 detection (LoD) of AFS at a confidence level of 95%, applying the procedure
279 described by [25]. The LoD describes the lowest quantity of an analyte that can be
280 reliably detected above the observed background noise. In the case of read-mapping
281 approaches, LoD will depend on genome relatedness and resulting chance for false-
282 positive read assignment, which in turn partly depends on read length (see above). If
283 closely related genomes (e.g. sheep vs. goat and cattle vs. buffalo) are included in
284 an AFS mapping procedure using 150 bp reads, the method produces a LoD of 1.6
285 %. If only distant species are tested for, the LoD decreases to 1.0 %.

286 To also infer the random error produced by AFS, and thus the precision of the
287 method, we calculated 95% confidence intervals for every instance of the expected
288 species proportions between 0.5 and 80% (Online Tab. S3). Proportion components
289 below 2% are measured with about 50 % uncertainty. Measurement error decreased
290 to about 10 % for proportions between 2 and 36 % and 4 % for proportions above
291 36%. Overall, CIs turned out to be close to the expected values and therefore are an
292 excellent indication of high AFS precision over the entire range of expected values
293 from 0.5 to 80%.

294 Discussion

295

296 Classical DNA-based species identification in food is routinely performed as a
297 targeted approach using PCR-based methods, which can detect only a certain range
298 of taxa, for which the PCR primers ideally fit [5–8, 10–13]. AFS in contrast analyses
299 the complete DNA of a foodstuff without amplification and is therefore a non-targeted,
300 whole-genome screening approach [18]. To investigate the potential of AFS to detect
301 unforeseen species components, we chose to study a real-case food control scenario
302 and sequenced the meat from five doner kebab samples from the Rhine-Main area.
303 According to German food legislation, snacks sold under the label “doner kebab” are
304 expected to consist only of sheep and/or beef [26]. However, occasional surveys
305 conducted by food authorities [27] or even occasioned by broadcasting stations [28]
306 have already pointed at a considerable heterogeneity of animal species components
307 in doner kebab samples from Germany, which very often contained unlabelled poultry
308 (chicken, turkey) and in rare cases even pork. Using AFS, we found that 3 of our 5
309 samples indeed contained turkey meat, two samples even at a major extent (90 % or
310 more). None of the samples, however, was openly advertised to the consumer as
311 “poultry doner”. In addition, AFS detected in four cases soy as an unexpected and
312 unlabelled ingredient, which may be critical for consumers suffering from allergy
313 towards soybeans. Soybean DNA may originate from the usage of spice coating
314 (panada). One sample additionally contained maize DNA, the origin of which is
315 unclear. AFS thus confirmed the results previously obtained by other labs in doner
316 kebab species screens and thus should function well as a method in routine food
317 screening.

318 The performance of AFS for the quantification of species in different types of food
319 matrices has not yet been investigated systematically. The main focus of the current
320 study was therefore to explore the quantification potential of AFS to infer species
321 proportions of reference sausages, which have previously been used in the field to
322 evaluate PCR-based quantification methods. To directly compare AFS to
323 quantification results obtained by qPCR [8] and ddPCR [13], we calculated for
324 simplicity the sum of the % deviation (measured vs. expected) for each sausage
325 sample (Tab. 1b-d). Results showed that AFS data -very much like the qPCR and
326 ddPCR data- need to be calibrated for matrix-dependent biases to generate the most
327 accurate results. Indeed, in 8 of 12 cases “AFS-cal” produced the best results, while
328 ddPCR turned out to be clearly superior in one case (Kal C) and slightly better in
329 three cases (KLyo A, C, D). AFS readily identifies and quantifies proportions of
330 species over a broad % range. Most importantly it works at the 1% level, a value
331 often approximatively taken by food authorities to distinguish problematic species
332 amounts from trace amounts, e.g. originating by unavoidable contamination.

333 Very much like for other DNA-based methods, the limitations of AFS are set by
334 sequence similarities between closely related genomes and by the so-called matrix
335 effect, which ultimately determines the extent to which species proportions in food

336 can be indirectly inferred from DNA proportions. Our theoretical evaluation of
337 possible wrong read assignments between closely related taxa provides the applicant
338 of AFS with a means to readily distinguish between true and false quantification
339 results. Food consisting of species, which have diverged at minimum 10 mio years
340 ago (e.g. sheep-goat or cattle-buffalo), may thus be analysed without much
341 problems. If AFS is performed for other, possibly closer taxa, the limits of false-
342 positives can easily be determined by the procedure, which we have outlined in the
343 methods section.

344 As previously noticed for PCR-based quantification methods [8, 13], the AFS requires
345 mathematical calibration for matrix effects to achieve best results (see above).
346 Theoretically, for instance, it should be necessary for AFS to take into account that
347 birds have only 1/3 the genome size of mammals. In practice, this consideration
348 proved to be not useful at all for quantifying food containing a mixture of bird and
349 mammalian material by AFS (data not shown). The possible reason is that chicken
350 meat may contain more DNA per gram tissue than, e.g., pork [29], thus
351 compensating for the smaller genome size. It will be almost impossible to define the
352 DNA amounts for all conceivable tissues from food-relevant species. However, the
353 application of food matrix reference material, as done in the present study, facilitates
354 a guided calibration of matrix effects and thus efficiently circumvents this problem.

355 In conclusion, we confirm here that AFS is a potent additional screening and
356 quantification tool in the repertoire of foodstuff analysis. We have calculated that AFS
357 sequencing reagent costs (50 libraries prepared in parallel, 500 k reads each, all
358 loaded on 1 Illumina MiSeq flowcell) currently would amount to appr. 90 EUR per
359 sample (see [19] for high-multiplex estimations). The computer skills required match
360 those of a typical bioinformatics master student, and routine screening of 1 mio reads
361 against up to 10 eukaryotic genomes can be performed on a laptop PC requiring a
362 computation time of appr. 20 min (see Materials and Methods for hardware used).
363 We like to point out that, in contrast to standard PCR analytics and depending on the
364 desired depth of analysis, the AFS can go well beyond the mere identification of
365 animal and plant species into the world of food microbiota, even including viruses
366 [18]. Ideally, AFS would screen for the ever-growing number of sequenced animal,
367 plant, fungal and bacterial genomes in one single analysis on standard computers.
368 However, due to the usage of algorithms involving read mapping and sequence
369 alignments, the screening power of AFS is currently limited to 20-30 species with
370 large eukaryotic genomes in one analysis. We therefore investigate the applicability
371 of novel non-alignment-based, memory-efficient algorithms for AFS. At the same
372 time, the identification and quantification of microbiota from foodstuff by AFS is a goal
373 worth of pursuing in future.

374 **Conflict of Interest**

375 The authors declare that they have no conflict of interest.

376

377 **Compliance with ethics requirements**

378 This article does not contain any studies with human or living animal subjects.

379

380 References

- 381 1. German Federal Office for Risk Assessment: Food safety and globalisation -
382 challenges and opportunities (Bundesamt für Risikobewertung).
383 [https://www.bfr.bund.de/en/press_information/2014/13/food_safety_and_global](https://www.bfr.bund.de/en/press_information/2014/13/food_safety_and_globalisation___challenges_and_opportunities-190341.html)
384 [isation___challenges_and_opportunities-190341.html](https://www.bfr.bund.de/en/press_information/2014/13/food_safety_and_globalisation___challenges_and_opportunities-190341.html). Accessed 28 Aug 2019
- 385 2. German Federal Office of Consumer Protection and Food Safety (Bundesamt
386 für Verbraucherschutz und Lebensmittelsicherheit).
387 https://www.bvl.bund.de/DE/Home/homepage_node.html. Accessed 28 Aug
388 2019
- 389 3. German Drug Law (Bundesministerium der Justiz und für Verbraucherschutz:
390 Gesetz über den Verkehr mit Arzneimitteln). [https://www.gesetze-im-](https://www.gesetze-im-internet.de/amg_1976/AMG.pdf)
391 [internet.de/amg_1976/AMG.pdf](https://www.gesetze-im-internet.de/amg_1976/AMG.pdf). Accessed 28 Aug 2019
- 392 4. Swiss Food Legislation (Schweizerisches Bundesgesetz über Lebensmittel und
393 Gebrauchsgegenstände (Lebensmittelgesetz, LMG) vom 20. Juni 2014).
394 <https://www.admin.ch/opc/de/official-compilation/2017/249.pdf>. Accessed 28
395 Aug 2019
- 396 5. Brodmann PD, Moor D (2003) Sensitive and semi-quantitative TaqMan™ real-
397 time polymerase chain reaction systems for the detection of beef (*Bos taurus*)
398 and the detection of the family Mammalia in food and feed. *Meat Sci* 65:599–
399 607. [https://doi.org/10.1016/S0309-1740\(02\)00253-X](https://doi.org/10.1016/S0309-1740(02)00253-X)
- 400 6. Zhang C-L, Fowler MR, Scott NW, et al (2007) A TaqMan real-time PCR
401 system for the identification and quantification of bovine DNA in meats, milks
402 and cheeses. *Food Control* 18:1149–1158.
403 <https://doi.org/10.1016/J.FOODCONT.2006.07.018>
- 404 7. Köppel R, Ruf J, Zimmerli F, Breitenmoser A (2008) Multiplex real-time PCR for
405 the detection and quantification of DNA from beef, pork, chicken and turkey.
406 *Eur Food Res Technol* 227:1199–1203. [https://doi.org/10.1007/s00217-008-](https://doi.org/10.1007/s00217-008-0837-7)
407 [0837-7](https://doi.org/10.1007/s00217-008-0837-7)
- 408 8. Köppel R, Ruf J, Rentsch J (2011) Multiplex real-time PCR for the detection
409 and quantification of DNA from beef, pork, horse and sheep. *Eur Food Res*
410 *Technol* 232:151–155. <https://doi.org/10.1007/s00217-010-1371-y>
- 411 9. Köppel R, Eugster A, Ruf J, Rentsch J (2012) Quantification of Meat
412 Proportions by Measuring DNA Contents in Raw and Boiled Sausages Using
413 Matrix-Adapted Calibrators and Multiplex Real-Time PCR. *J AOAC Int* 95:494–
414 499. <https://doi.org/10.5740/jaoacint.11-115>
- 415 10. Ulca P, Balta H, Çağın İ, Senyuva HZ (2013) Meat species identification and
416 Halal authentication using PCR analysis of raw and cooked traditional Turkish
417 foods. *Meat Sci* 94:280–284. <https://doi.org/10.1016/j.meatsci.2013.03.008>
- 418 11. Floren C, Wiedemann I, Brenig B, et al (2015) Species identification and
419 quantification in meat and meat products using droplet digital PCR (ddPCR).
420 *Food Chem* 173:1054–1058. <https://doi.org/10.1016/j.foodchem.2014.10.138>

421

- 422 12. Song K-Y, Hwang HJ, Kim JH (2017) Ultra-fast DNA-based multiplex
423 convection PCR method for meat species identification with possible on-site
424 applications. *Food Chem* 229:341–346.
425 <https://doi.org/10.1016/j.foodchem.2017.02.085>
- 426 13. Köppel R, Ganeshan A, Weber S, et al (2019) Duplex digital PCR for the
427 determination of meat proportions of sausages containing meat from chicken,
428 turkey, horse, cow, pig and sheep. *Eur Food Res Technol* 245:853–862.
429 <https://doi.org/10.1007/s00217-018-3220-3>
- 430 14. Markoulatos P, Siafakas N, Moncany M (2002) Multiplex polymerase chain
431 reaction: A practical approach. *J Clin Lab Anal* 16:47–51.
432 <https://doi.org/10.1002/jcla.2058>
- 433 15. Berry D, Mahfoudh K Ben, Wagner M, Loy A (2011) Barcoded Primers Used in
434 Multiplex Amplicon Pyrosequencing Bias Amplification. *Appl Environ Microbiol*
435 77:7846–7849. <https://doi.org/10.1128/AEM.05220-11>
- 436 16. Tedersoo L, Anslan S, Bahram M, et al (2015) Shotgun metagenomes and
437 multiple primer pair-barcode combinations of amplicons reveal biases in
438 metabarcoding analyses of fungi. *MycKeys* 10:1–43.
439 <https://doi.org/10.3897/mycokeys.10.4852>
- 440 17. Sze MA, Schloss PD (2019) The Impact of DNA Polymerase and Number of
441 Rounds of Amplification in PCR on 16S rRNA Gene Sequence Data. *mSphere*
442 4:e00163-19. <https://doi.org/10.1128/msphere.00163-19>
- 443 18. Ripp F, Krombholz C, Liu Y, et al (2014) All-Food-Seq (AFS): a quantifiable
444 screen for species in biological samples by deep DNA sequencing. *BMC*
445 *Genomics* 15:639. <https://doi.org/10.1186/1471-2164-15-639>
- 446 19. Liu Y, Ripp F, Koeppel R, et al (2017) AFS: identification and quantification of
447 species composition by metagenomic sequencing. *Bioinformatics* 33:btw822.
448 <https://doi.org/10.1093/bioinformatics/btw822>
- 449 20. Eugster A, Ruf J, Rentsch J, Köppel R (2009) Quantification of beef, pork,
450 chicken and turkey proportions in sausages: use of matrix-adapted standards
451 and comparison of single versus multiplex PCR in an interlaboratory trial. *Eur*
452 *Food Res Technol* 230:55–61. <https://doi.org/10.1007/s00217-009-1138-5>
- 453 21. FASTQC. A quality control tool for high throughput sequence data.
454 <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed 28 Aug
455 2019
- 456 22. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for
457 Illumina sequence data. *Bioinformatics* 30:2114–2120.
458 <https://doi.org/10.1093/bioinformatics/btu170>
- 459 23. BBTools. A suite of fast, multithreaded bioinformatics tools designed for
460 analysis of DNA and RNA sequence data.
461 <https://sourceforge.net/projects/bbmap/>. Accessed 28 Aug 2019

462

- 463 24. Kumar S, Stecher G, Suleski M, Hedges SB (2017) TimeTree: A Resource for
464 Timelines, Timetrees, and Divergence Times. *Mol Biol Evol* 34:1812–1819.
465 <https://doi.org/10.1093/molbev/msx116>
- 466 25. Armbruster DA, Pry T (2008) Limit of blank, limit of detection and limit of
467 quantitation. *Clin Biochem Rev* 29 Suppl 1:S49-52
- 468 26. Marking of “doner kebab” and “similar” products by bulk delivery.
469 (Kenntlichmachung von „Döner Kebab“ und „ähnlichen“ Erzeugnissen bei loser
470 Abgabe. Bayerisches Landesamt für Gesundheit und Lebensmittelsicherheit).
471 https://www.lgl.bayern.de/downloads/lebensmittel/doc/merkblatt_doener_kebab
472 .pdf. Accessed 28 Aug 2019
- 473 27. The composition and labelling of doner kebabs. LACORS.
474 <https://www.ihsti.com/lacors/ContentDetails.aspx?id=21001>. Accessed 28 Aug
475 2019
- 476 28. Frequently minced meat and additives in doner kebab. (Häufig Fleischbrät und
477 Zusatzstoffe im Döner. Norddeutscher Rundfunk).
478 [https://www.ndr.de/ratgeber/verbraucher/Haeufig-Fleischbraet-und-](https://www.ndr.de/ratgeber/verbraucher/Haeufig-Fleischbraet-und-Zusatzstoffe-im-Doener,doener164.html)
479 [Zusatzstoffe-im-Doener,doener164.html](https://www.ndr.de/ratgeber/verbraucher/Haeufig-Fleischbraet-und-Zusatzstoffe-im-Doener,doener164.html). Accessed 28 Aug 2019
- 480 29. Cai Y, Li X, Lv R, et al (2014) Quantitative analysis of pork and chicken
481 products by droplet digital PCR. *Biomed Res Int* 2014:810209.
482 <https://doi.org/10.1155/2014/810209>

483 **Tab. 1 Quantification results of AFS pipeline.** Raw results obtained by AFS
 484 analysis as well as calibrated AFS values obtained by linear regression are
 485 compared to PCR-based quantification for **(a)** Doner Kebab samples, **(b)** Kal A-E, **(c)**
 486 Klyo A-D and **(d)** KGeflLyo A-D. qPCR data were obtained from [8, 20], ddPCR data
 487 from [13]. “Sum dev” represents the sum of % deviation from expected proportions.
 488 Best results for each sausage are shaded grey. (n.a. = not analysed)

489 **(a)**

	beef	turkey	soy	maize	horse	pork	chicken	sheep	goat
Doner Kebab 1	8.6	88.6	0.7	1.8	0.0	0.0	0.2	0.0	0.0
Doner Kebab 2	99.5	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.3
Doner Kebab 3	98.7	0.0	0.4	0.0	0.0	0.0	0.0	0.4	0.4
Doner Kebab 4	5.1	94.5	0.0	0.0	0.0	0.0	0.3	0.0	0.0
Doner Kebab 5	27.7	71.3	0.4	0.0	0.0	0.0	0.3	0.1	0.1

490

491 **(b)**

		beef	pork	sheep	horse	chicken	turkey	goat	buffalo	sum dev
Kal A	expected	1.0	35.0	9.0	55.0	0.0	0.0	0.0	0.0	
	AFS	1.4	30.9	10.1	57.3	0.0	0.0	0.3	0.0	8.2
	AFS cal.	0.3	34.9	10.3	54.3	0.0	0.0	0.3	0.0	3.1
	qPCR	0.4	39.3	8.9	51.5	n.a.	n.a.	n.a.	n.a.	8.6
Kal B	expected	9.0	55.0	1.0	35.0	0.0	0.0	0.0	0.0	
	AFS	11.2	49.5	1.3	37.8	0.0	0.0	0.1	0.1	11.0
	AFS cal.	9.6	54.2	1.0	35.1	0.0	0.0	0.1	0.1	1.7
	qPCR	23.0	51.0	1.5	24.4	n.a.	n.a.	n.a.	n.a.	29.1
Kal C	expected	25.0	25.0	25.0	25.0	0.0	0.0	0.0	0.0	
	AFS	25.8	19.8	28.2	25.1	0.0	0.0	1.0	0.2	10.5
	AFS cal.	23.7	22.4	29.0	23.7	0.0	0.0	0.9	0.2	10.3
	qPCR	34.0	18.8	25.8	21.4	n.a.	n.a.	n.a.	n.a.	19.6
	ddPCR	25.6	25.3	24.6	24.5	n.a.	n.a.	n.a.	n.a.	1.8
Kal D	expected	35.0	9.0	55.0	1.0	0.0	0.0	0.0	0.0	
	AFS	38.2	7.7	50.8	1.3	0.0	0.0	1.7	0.3	11.1
	AFS cal.	35.3	9.1	52.1	1.5	0.0	0.0	1.7	0.3	5.8
	qPCR	29.9	7.6	61.7	0.9	n.a.	n.a.	n.a.	n.a.	13.3
Kal E	expected	55.0	1.0	35.0	9.0	0.0	0.0	0.0	0.0	
	AFS	56.9	1.0	32.2	8.5	0.0	0.0	1.1	0.4	6.7
	AFS cal.	54.5	1.9	33.8	8.4	0.0	0.0	1.1	0.4	4.7
	qPCR	51.2	1.0	37.2	10.5	n.a.	n.a.	n.a.	n.a.	7.5

492

493

494

(c)

		beef	pork	sheep	horse	chicken	turkey	goat	buffalo	sum dev
KLyo A	expected	14.0	80.0	0.0	0.0	0.5	5.5	0.0	0.0	
	AFS	19.3	74.6	0.0	0.0	0.7	5.3	0.0	0.1	11.2
	AFS cal.	13.0	80.8	0.0	0.0	0.3	5.7	0.0	0.1	2.3
	ddPCR	14.2	80.2	n.a.	n.a.	0.4	5.3	n.a.	n.a.	0.7
KLyo B	expected	36.0	58.0	0.0	0.0	2.0	4.0	0.0	0.0	
	AFS	41.8	52.4	0.0	0.0	2.2	3.4	0.0	0.3	12.5
	AFS cal.	35.9	57.9	0.0	0.0	2.1	3.8	0.0	0.3	0.8
	ddPCR	34.9	60.1	n.a.	n.a.	1.4	3.7	n.a.	n.a.	4.1
KLyo C	expected	58.0	36.0	0.0	0.0	4.0	2.0	0.0	0.0	
	AFS	66.0	28.3	0.0	0.0	3.9	1.4	0.0	0.4	17.0
	AFS cal.	60.6	33.1	0.0	0.0	4.1	1.7	0.0	0.4	6.3
	ddPCR	58.4	36.8	n.a.	n.a.	2.9	1.9	n.a.	n.a.	2.4
KLyo D	expected	80.0	14.0	0.0	0.0	5.5	0.5	0.0	0.0	
	AFS	82.8	11.3	0.0	0.0	5.0	0.4	0.0	0.5	6.8
	AFS cal.	77.7	15.7	0.0	0.0	5.3	0.7	0.0	0.5	4.9
	ddPCR	79.2	15.9	n.a.	n.a.	4.3	0.6	n.a.	n.a.	4.0

495

496

(d)

		beef	pork	sheep	horse	chicken	turkey	goat	buffalo	sum dev
KGeflLyo A	expected	0.5	5.5	0.0	0.0	14.0	80.0	0.0	0.0	
	AFS	0.6	5.1	0.0	0.0	29.8	64.5	0.0	0.0	31.9
	AFS cal.	0.2	5.6	0.0	0.0	10.0	84.1	0.0	0.0	8.5
	ddPCR	0.8	9.4	n.a.	n.a.	24.6	65.1	n.a.	n.a.	29.7
KGeflLyo B	expected	2.0	4.0	0.0	0.0	36.0	58.0	0.0	0.0	
	AFS	2.0	3.5	0.0	0.0	56.9	37.6	0.0	0.0	41.9
	AFS cal.	2.2	3.9	0.0	0.0	40.3	53.5	0.0	0.0	9.1
	ddPCR	2.1	5.0	n.a.	n.a.	41.7	51.2	n.a.	n.a.	13.6
KGeflLyo C	expected	4.0	2.0	0.0	0.0	58.0	36.0	0.0	0.0	
	AFS	3.4	1.4	0.0	0.0	75.7	19.5	0.0	0.0	35.4
	AFS cal.	4.4	1.8	0.0	0.0	61.2	32.6	0.0	0.0	7.2
	ddPCR	4.2	2.4	n.a.	n.a.	62.6	30.8	n.a.	n.a.	10.4
KGeflLyo D	expected	5.5	0.5	0.0	0.0	80.0	14.0	0.0	0.0	
	AFS	3.9	0.4	0.0	0.0	89.2	6.5	0.0	0.0	18.4
	AFS cal.	5.2	0.7	0.0	0.0	76.4	17.8	0.0	0.0	7.9

497