Progress report European Decathlon project ([www.decathlon-project.eu](http://www.decathlon-project.eu))

The European Decathlon (Development of Cost efficient Advanced DNA-based methods for specific Traceability issues and High Level On-site applicatioNsproject) was started on December 1, 2013. The project has finished on November 30, 2016.

The project has focused on the development of DNA-based methods for the three areas of the project: 1) food pathogens, 2) GMOs (genetically modified organisms, with a focus on unauthorised GMOs), and 3) customs issues (with a focus on tobacco and endangered species identification). The work in the different Work Packages within the Decathlon project will be summarized here.

In the Decathlon project, Work Package 1 (WP1), the Advanced Analytical Knowledge Center, has brought together all expertise of the consortium in the field of isothermal amplification, (droplet) digital PCR, next generation sequencing (NGS) and bioinformatics. In deliverable report D1.1. the results are described of the assessment of existing as well as newly developed isothermal amplification methods for on-site application in the fields of food testing (Escherichia coli with focus on EHEC isolates), GMO testing and areas relevant to customs issues i.e. tobacco identification. This relates to the following isothermal methods: LAMP (Loop-mediated isothermal amplification), RPA (recombinase polymerase amplification), and HDA (helicase-dependent amplification). It was concluded that LAMP assays would initially only be used as a fast and specific means of identification and characterization of E. coli isolates, LAMP assays for markers for pathogenicity will require further optimisation. For GMO detection, it was concluded that LAMP has shown to be the most promising isothermal amplification technology, rather than the other tested isothermal amplification technologies HDA and RPA, as the amplification mechanism in LAMP is triggered solely by the interplay of the related oligo-nucleotides: it does not require additional enzymes besides a strand displacement DNA polymerase to dehybridise DNA duplexes. As a result, the LAMP reaction mix is much simpler than for other isothermal amplification technologies which increases robustness and facilitates the integration of LAMP based assays in on-site applicable formats.

The work in WP1 has more recently focused on isothermal amplification, digital droplet PCR (ddPCR) and Next Generation Sequencing (NGS). Within WP1 experimental work is performed to aid especially WP2, 3,4 and 5 that further develop the respective techniques in the different areas of application. In WP1 work has been performed on isothermal amplification to underpin the work in WP2 (on-site detection) and WP3 (food pathogens). In WP2 the initial focus is on on-site methods for food pathogen detection and it is important to have isothermal amplification methods available. In WP1 a review has been performed of available isothermal methods for the detection and identification of food pathogens with the emphasis on detection of EHEC and the different target genes were identified. Published assays were assessed in terms of target genes, sample matrices tested, type of amplification product detection, validation data available, etc. In addition to this also ddPCR has been explored as an alternative method for PCR-based quantitation of DNA samples. For the development of multiplex ddPCR systems to quantify GMOs two different analysis strategies have been followed: (i) quantification of twelve EU authorized GM maize lines in a single analysis and (ii) screening for NOS terminator positive and 35S promotor positive samples with conventional qPCR (first step) and ddPCR based quantification of all positive samples for the remaining four EU authorized GM maize lines (second step). As this qPCR modules were designed for use in simplex qPCR reactions, the primers and probes had to be checked carefully for interactions and the ddPCR methods have been optimised accordingly. For NGS a comprehensive report has been written to give an overview of the state-of-the-art of NGS (Next Generation Sequencing) strategies as reported in the scientific literature in the areas of detection and identification of food pathogens, of GMOs and in the field of species identification. Under WP1 all bioinformatics specialists work together in the Task Force Bioinformatics, that has progressed on the development of user friendly modules and pipelines for the routine analysis of whole genome and DNA target enriched datasets for the different Decathlon end-points.

WP2, focusing on on-site applications of DNA-based methods, is divided into two main functions. The first task encompasses development of a microfluidic cartridge for spatial multiplexing of an existing DNA molecular method. This involves discretising a purified DNA sample into a series of equal volumes. Currently the cartridge is performing as planned and can split a DNA sample into the required aliquots. It is scalable; a number of variants have been tested with between 4 and 18 separate wells. Relative to the manufacturing technique used it shows a high degree of reliability. Development of the ‘spin-stand’ test bed is currently being completed. The microfluidic disc has been tested using LAMP primers (provided by WP3) using purified DNA samples (provided by WP3). Furthermore, the issue of the purification of DNA from lysed samples to a quality required for LAMP amplification has been addressed. Based on input from WP3 it was determined to use a silica bead based purification method. A liquid handling protocol was determined. Currently the disc has been designed and initial versions have been tested. Flow visualisation (using stroboscopic imaging) shows the disc is performing as designed. However, minor iterative improvements are required before testing with samples (DNA isolates and later different food matrices) is initialised. As the spin-stand is not yet complete, LAMP amplification has been attempted using a custom heating jig and off-disc fluorescent measurements. However, this phase of testing is at an early stage. With respect to the graphene-based sensor approach. Here, it was found that nanogaps, manufactured under vacuum on free standing graphene membranes, are folded or ruptured when exposed to a water environment. Novel manufacturing methods have been developed to circumvent this issue and thus permit the nanogaps to be deployed in an aqueous solution. Work has focussed here on determining the biomolecular detection limit. This initially involves electrical characterisation of the DNA sensors. An alternative fabrication protocol was designed for, in the end, supported graphene nanogaps. Also, two different protocols to transfer the graphene monolayer on top of the pre patterned structures were tested. Further functionalization of the edge of the etched structure will be studied.

In WP3, focusing on food pathogens, the focus is on the improvement of EHEC diagnostics, using NGS and DNA sequence-based approaches for diagnostic purposes and comprehensive risk assessment of food pathogens. Also, experiments were performed to develop and optimise modules for the extraction of DNA directly from larger volumes of food matrices (minced meat or vegetables) or for the isolation of intact bacteria from irrigation water, with the aim to provide DNA of sufficient quality and quantity to be successively analysed with molecular analytical modules for the detection of the pathogen / virulence factors. The meat/vegetable protocol was tested on samples spiked at concentrations challenging the theoretical limit of detection, and demonstrated significantly improved detectability with recovery rates ranging around 15-20% and 40-50% for meat and vegetables, respectively. Furthermore, two bioinformatics workflows for genome marker identification from larger sets of bacterial genome sequences were developed and evaluated. Based on the comparative genomic analyses, 47 new specific biomarkers for the clinically most relevant STEC subtypes (O157 and non-O157 ‘big six’ serotypes as well as for O104). A multiplex PCR was developed based on the newly discriminative biomarkers, which allow STEC detection and discrimination of the clinically most relevant STEC serotypes as mentioned before. The specificity of the multiplex PCR was shown based on well defined clinical STEC isolates as well as several established E. coli reference strain collections, including the ECOR, DEC and HUSEC collections. In summary, an improved DNA-based typing tool for clinically relevant STEC has been developed, that can not only be used for improved typing and risk assessment in clinical samples, but also in food-related samples. This research will thus contribute to the further improvement of detection and typing approaches of pathogens, as well as the integration of foodborne pathogen detection and subtyping into a single test.

WP4 focused on improved methods for GMO detection and identification, including unauthorised GMOs. A ddPCR protocol for the identification and quantification of all EU-authorised GM maize varieties, which was prepared in WP1, was tested in three different labs of partners. The results of the transferability study have shown that the protocol was successfully transferred to other labs. The in-house validation of the method with two multiplex reactions met all the minimum performance parameters as established in WP6. Therefore the method was selected by WP1 as the method suitable to enter into a full collaborative trial validation lead by WP6. Additional to the performance parameters, the cost-effectiveness of the multiplex ddPCR method was assessed in line with current GMO quantification procedure with qPCR. It was shown that the multiplex ddPCR for quantification of twelve maize lines was more cost effective than qPCR in all cases but one, which is the negative sample, where only screening is done with qPCR. Furthermore, NGS approaches have been initiated. A method for NGS analysis of amplicons has been developed, that may provide in specific cases advantages in cost efficiency and flexibility over current qPCR approaches. It was shown that the NGS PCR amplicon sequencing provides generally the same screening information as qPCR screening. In general it was concluded that the flexibility of NGS amplicon targeting requiring less specificity testing seems currently the biggest benefit of this strategy. Then, for the improvement detection and identification an optimised strategy was developed to identify unknown GMOs starting from unexplained GMO elements in a given sample. The new approach yielded sequence information that is likely to be related to this model-UGMO, indicating the proof-of-principle of this method. The data were analysed based on a newly developed (WP1) NGS data analysis pipeline.

Under WP5, the customs issues are: i) identification of tobacco in reconstituted tobacco, ii) tobacco profiling (comparison of blends, batches, manufacturers) and iii) identification of (endangered) species in complex samples (multiple ingredients). For the identification of tobacco in reconstituted tobacco samples, LAMP- as well as real-time PCR methods have been developed. Out of the assays tested, the PMT-1 target was selected for prevalidation experiments. Based on these experiments, a validation trial was planned and performed for the tobacco specific detection using the PMT-1 real-time PCR assay. Together with this real-time method validation trial, it was planned to collect data for a LAMP assay, as LAMP is a potential on-site method that could be used on-site by, for instance, customs inspection services. The LAMP Nia-Green assay turned out to be the most promising method to test in parallel with the real-time PCR method. Based on these results it was concluded that RAD sequencing could be used for tobacco species identification, possibly including mixtures, but that the the depth of tag used in these experiments is not good enough to yield reliable SNP markers for effective species identification in mixtures. The second issue has to be taken into consideration is that the tag size (33-36 bp) is relatively small. Therefore, the approach is generally prone to yield false positive SNPs when mapping to the scaffold. It was therefore concluded that additional strategies will need to be explored to make it feasible to identify subsamples of individual tobacco batches. For the identification of species, including CITES, species in complex samples, that may be degraded, an informative protocol was developed that can basically identify all species present in a sample, provided that the related sequence information is present in the related databases. In addition to this a bioinformatics pipeline has been developed and made available for (customs) laboratories via a user-friendly webtool. After elaborate testing it was concluded that almost all species in the tested mixtures could accurately be identified at the species or family-level using the proposed DNA metabarcoding approach, even in cases where the larger part of the ingredients (species) were present at 1% dry weight concentration.

WP6 has compiled a comprehensive set of minimum performance parameters (MPPs) for molecular analytical methods together with their associated acceptance values (AAVs), covering techniques for all steps from nucleic acid extraction to next generation (high throughput) nucleic acid sequencing. These MPPs and AAVs can be used to assess and improve the quality of analytical methods and services from developmental stages, via validation to routine application. Much emphasis has been put on making the MPPs and AAVs applicable across sectors and balancing pragmatism with focus on maintaining sufficient quality. Application of these MPPs and AAVs are also expected to reduce costs by efficient resource investments, e.g. in the development phase and when laboratories select among alternative analytical methods. The first method developed in the frame of the DECATHLON project and selected for validation was a digital droplet polymerase chain reaction (ddPCR) method for the quantification of 12 EU- authorized genetically modified (GM) event. However, after the prevalidation, the intended validation could not be completed as a consequence of a long series of unfortunate and largely unpredicted factors. Then, two methods have been developed for the distinction of cigarettes (outer wrapper of paper) and cigarillos/cigars (outer wrapper of reconstituted tobacco/ tobacco): a quantitative real- time PCR (qPCR) assay and a Loop-mediated isothermal AMPlification (LAMP) assay. Both methods have been put to the test in an international ring trial to determine their performance. Overall, it was concluded that the developed DNA analysis methods will provide useful tools for customs: the ability to detect the presence of tobacco in the outer wrapper of these products is important for the tariff-classification of cigarettes and cigarillos/cigars, with regard to import duties and excise duty rates. The development of the combined method and pipeline for CITES species detection was selected for a large inter-laboratory trial validation. The CITESspeciesDetect pipeline is developed to perform reference based parallel identification of multiple amplified barcodes and mini-barcodes from plant and animal species. The input data are paired-end (PE) sequence reads from Illumina MiSeq or HiSeq sequencers. The output is a table of hits of operational taxonomic units (OTUs) from reads mapping to reference databases. It was concluded that the pipeline is a reliable tool, that is fit for purpose. With the recent availability of the online web-interface it became possible to re-analyze the data by each of the individual participants of the inter-laboratory trial.

WP7 is the central WP in the Decathlon project that focuses on communication in every aspect. In WP7 the Communication, Dissemination and Exploitation Plan has been developed to provide the basic structure for the communication within the project as well as the communication with the stakeholders. Communication activities included the compilation of a database of all relevant stakeholders, the development of the project’s brand identity (project logo), the construction of the website (www.decathlon-project.eu), the use of social media, including LinkedIn and Twitter. Regarding internal communication, next to the activities as performed under WP8, the Basecamp platform was effectively used for disseminating information and internal documents. A meeting calendar was available on Basecamp to inform partners of the scheduled meetings. A communication tool (GoToMeeting) was used by partners for the online meetings. An overview of all dissemination activities is updated each half year. Project newsletters, the project brochure, general flyer and roll-up for Decathlon were created as additional promotional materials to provide interested stakeholders and key actors with brief and concise information of the project. The webinars and stakeholder involvement was a very crucial aspect to the successful implementation of the Decathlon project. WP3, WP4 and WP5 have collected information from the respective stakeholders through the organisation of webinars or by contacting them directly by email. Finally, international stakeholder workshops have been organised to enable the Decathlon consortium to interact with stakeholders from industry, academia, as well as policy-makers, to jointly work on the development of new or improved methods that meet the requirements for each field. Also, WP7 has facilitated many training activities, varying from workshops, the development of decision support systems and e-learning tools, training sessions in webinars for stakeholders, to the organisation of exchange visits between partners.