EUROPEAN COMMISSION

Community Research

SEVENTH FRAMEWORK PROGRAMME

GRACE Stakeholder Consultation on animal feeding studies and *in vitro* studies in GMO risk assessment

Addendum

FP7 Collaborative Project
GRACE 311957

Responses of GRACE team members to questions and comments raised by stakeholders

Addendum to the Stakeholder Consultation Report

Draft Version
December 2013

GRACE

GMO Risk Assessment and Communication of Evidence

This addendum to the report “GRACE Stakeholder Consultation on animal feeding studies and in vitro studies in GMO risk assessment” dated March 2013 holds the responses of GRACE team members to questions and comments raised by stakeholders participating in the consultation in December 2012. Please, note that this compilation is not yet complete. Questions and comments still lacking a response are provided in the second section of this addendum. A complete version of responses will replace this document.

“GMO Risk Assessment and Communication of Evidence”, GRACE is a Collaborative Project of the Seventh Framework Programme of the European Community for Research, Technological Development and Demonstration Activities.
Grant agreement no: 311957
Project duration: 1st June 2012 – 30th November 2015
GRACE website: www.grace-fp7.eu

Table of contents

Part I: Responses to questions and comments

Q/A	Comment/suggestion	Adopted in study plan	Explained in study plan	Not in study plan	Brief explanation/argumentation addendum
	General				
1	Scope of the 90-day feeding study (hypothesis to be tested):		x		Following from the tasks defined in the Call GRACE is going to evaluate and compare the approach "90-day feeding study" with other potential tools or sources of information for the risk assessment of whole food/feed derived from GM crops. Outcomes from 90-day feeding studies will be compared with outcomes of the compositional analysis, ~omics of plant material and animal tissues, in vitro studies with animal tissues, longitudinal metabolomics and extended feeding studies, and literature data. Hence, a comparison of available and suggested methods is achieved. Aside the consideration of alternative or improved methodologies the investigations intend to consider two critical preconditions for feeding studies - robustness and sensitivity. "Robustness" issues addressed are (i) how do the variety or the kind of crop affect testing and outcome, or (ii) do different trials (and labs) achieve equivalent results. "Sensitivity" is addressing that the test indicates a risk when there is a risk e.g. for human consumption. In contrast to initial planning the feeding trials and analytical efforts have to be focussed due to cost constraints (extended analysis following the stakeholder consultations ; increased costs for services) and the retreat of providers (USDA). Therefore, histopathological examinations are only conducted for high dosage and control groups as well as an extended analyses will tentatively be conducted in the two initial feeding studies only. Yet the full set of samples will be available for later examination, if required. Key issues can still be considered by combining the different approaches. Feeding trials conducted with MON810 maize will address the comparison of different approaches. All feeding studies with maize varieties and potato lines form the base to explore robustness of testing. Intended feeding trials with potatoes from USDA expressing high and low glycoalcaloid content have been suspended because USDA will not support GRACE any further. (Glycoalkaloids are known toxins for humans). "Sensitivity" will be

Q/A	Comment/suggestion	Adopted in study plan	Explained in study plan	Not in study plan	Brief explanation/argumentation addendum
					considered based on the other trials as welll as by reviewing recent publications. The setting of the feeding trials comprise protocols that essentially follow the guidlines of EFSA (EFSA 2011) comparing a GM crop (line) and its (near-)isogenic counterpart i.e. testing whether the transgenic and non-transgenic plants differ relevantly in the set of endpoints (hypothesis testing: similar vs. different). These tests will be compared to alternative approaches (testing outcome feeding trials vs. alternatives). "Historical data" will be generated by testing further (conventional) maize and (conventional, transgenic) potato as whole food/feed. The "historical data" inform about the quality of the assessment based on the generic variability for whole food/feed test data generated in the same lab (OECD 2000). The pathological review of the endpoint values informs about harmfull effects or critical uncertainties beyond the testing for statistically significant differences (which should be expected to show between varieties if the approach/es is/are sensitive to subtle changes in plant composition, at all).

Q/A	Comment/suggestion	Adopted in study plan	Explained in study plan	Not in study plan	Brief explanation/argumentation addendum
2	Which guidance is followed (OECD TG409, EFSA, other)?		x		The 90-day feeding trials simulate the procedure suggested by EFSA (transgenic vs. non-transgenic conterpart). The corresponding OECD guidline is No. 408: Repeated Dose 90-Day Oral Toxicity Study in Rodents (1998). In addition and in contrast to the EFSA guidance, but following the explanations for toxicity tests published by the OECD a reliable interpretation of outcomes should be based on historic data generated in the same lab (Guidance Notes for Analysis and Evaluation of Repeat-Dose Toxicity Studies - ENV/JM/MONO(2000)18). These historic data need to be generated for SZU because trials with whole food/feed maize have not been conducted there before.
4	Ethical aspects:				
5	Why repeating a trial, where safety is already demonstrated (via trials, consumption)?		x		As outlined in Q&A1 the scope of the 90-day feeding trial with MON810 is not to demonstrate or doubt the safety of this event. MON810 has been selected for several reasons, including the availability of several MON810 and near-isogenic varieties which can be cultivated in Europe and the existence of several data sets from feeding trials with MON810.
6	Has an independent and impartial project review determined that the benefits of this project outweigh harm to the animals?			x	Harm to the animals will be minimal as they will not be subjected to any procedures and the diets are not expected to be seriously toxic. The proposal of the project has been checked for ethical concerns during negotiation by the EC.
7	Cost – benefit analysis: increasing the numbers vs. added statistical power?			x	The number of animals is as recommended by EFSA, but see below (Q&A 41 - 47) for further justification

Q/A	Comment/suggestion	Adopted in study plan	Explained in study plan	Not in study plan	Brief explanation/argumentation addendum
8	How to assure independency of GRACE team members (e.g. link with EFSA)?			x	The proposal underwent a scientific evaluation according to EC rules. All team members of GRACE have been involved in research on GMO and are either members of scientific institutions, authorities or SME. Several senior scientists have or had been members of EFSA scientific committees. The list of partners in the GRACE project is freely available. Moreover the Consortium publicly presents and discusses detailed test protocols and outcomes to maximize transparency as it has been initiated with the stakeholer meeting in Vienna and the written consultations afterwards.
9	A 90-day feeding study cannot be conclusive as this is a sub-chronic test while the maize is conceived to be chronically consumed.			x	The overall purpose of the project is to reconsider the value of 90-day feeding studies for the risk assessment comparing various alternative approaches. Three positions about the necessity have been presented during the stakeholder meeting: always necessary - necessary only if there are distinct indications of effects - obsolete. This study will measure many endpoints (characters) and compare them with the control in well designed experiments. Ninety days is sufficient in toxicity tests with chemicals to detect any important differences. The set of 90-day feeding studies to be conducted in the course of GRACE will elucidate whether this crucial assumption holds true for feeding trials with whole food/feed. The aim of these kinds of feeding trials is not to determine the mode of action of a toxin. It is to detect any sign of toxicity. If distinct toxicity should be detected/verified then the mode of action might be investigated in an appropriate subsequent studies. In addition, an extended study is planned in the frame of GRACE.
10	A 90-day sub-chronic time frame is inappropriate given that the maize could be consumed over a lifetime.			x	
11	A 90-day feeding study is very probably only of limited value because the mode of action cannot be determined as the potential toxicants, if any, are only in the mg-range.			x	

Q/A	Comment/suggestion	Adopted in study plan	Explained in study plan	Not in study plan	Brief explanation/argumentation addendum
12	The findings from this one study should not be given undue weight relative to the findings from other, equally relevant, studies.			x	The purpose of the 90-day feeding studies with MON810 maize shall enable to compare similar studies and reconsider variation between results and approaches. Protocols and results will be made publicly available and should broadly faciliate the comparison of findings from different sources.
13	Information gained from the literature survey on previous GM food/feed animal tests should be used in designing the present study.	x			The design is based on the EFSA guidelines which took such literature into account.
14	Information could be generated including a broad range of in vitro toxicity tests with the plant material before running the feeding study, to be more targeted in the hypothesis to be tested.			x	The 90-day feedings studies will be compared to a set of complementary or alternative approaches in the course of GRACE ranging from compositional analysis to in vitro tests. In case of 90-day feeding studies of whole food/feed as a mandatory part of the risk assessment the underlying hypothesis is per se not specificly targeted. The interplay and value of the different approaches is explored and demonstrated between the different tests conducted throughout GRACE. MON810 maize serves as a well described and controlled model GM crop.
15	It would be interesting to have some animals left and not be killed after 90 days, but to investigate for a longer period of			x	Splitting the 90-day feeding trials will alter the statistical value as well as the analysis of the few extendedly fed animals will lack statistical power. GRACE will perform an additional extended feeding study, reconsidering the outcome of the 90-day feeding trials and ensuring sufficient statistical power of each trial to evaluate the approaches.

Q/A	Comment/suggestion	Adopted in study plan	Explained in study plan	Not in study plan	Brief explanation/argumentation addendum
	feeding.				
	On the protocol				
	Staff, GLP, data recording				
16	All Principal Investigators (PIs) should be listed, thus the nature of all activities conducted outside of Slovak Medical University (SMU) should be clearly defined as GLP or non-GLP and PIs assigned as appropriate. Similarly, Quality Assurance (QA) professionals for GLP-compliant activities conducted outside of SMU and TOPALAB should also be included.	x			see Study Plan A and B (private data will only be disclosed according to law standards).
18	Indicate GLP-compliant and non-compliant parts in protocol.	x			see Study Plan A and B

Q/A	Comment/suggestion	Adopted in study plan	Explained in study plan	Not in study plan	Brief explanation/argumentation addendum
19	Collect data in an electronic data acquisition system, enabling data tracking, offering data protection, and providing an audit-trail.			x	An electronic data acquisition system is currently not avialble. While such a system is of interest, it would require financial investment and careful staff training which could take some months. It would also require investment in automated data collection. Such investment is not justified and has not been considered in the budget. All paper records can be archived, and data integrity will be carefully checked to provide an audit trail. Nevertheless, SZU and JKI will work on improvements that can be realized during the course of the Project .
	Objective				
20	Justify taking MON810 (and not NK603 or MON863).			x	see Q&A 1.
21	Adverse findings will probably not be observed with MON810, then what about the relevance of the metabolomics data?			x	Within GRACE the same plant materials that will be assessed in the animal feeding trials will also be assessed using omics approaches (transcriptomics, proteomics and metabolomics). In this way it can be assessed whether these advanced analytical technologies applied to the plant materials are more informative compared to animal feeding trials and whether, in the case of animal experiments, early biomarkers of an alteration in metabolic pathways could be detected even if no clinical manifestation of potential unintended effect of the genetic modification is observed and interpret their relevance and predictive value. Background: Where animal feeding trials have clearly shown their added value in the case of well-characterised single substances, it is well recognised in the scientific community that there are important limitations to the performance of animal feeding trials in the case of whole foods. In the latter case the sensitivity of animal feeding trials is currently scientifically debated as only a limited amount of whole foods can be incorporated in the animal's diet without causing nutritional disturbances. It has therefore been questioned regularly whether animal feeding trials may be the best approach to identify unintended effects that may be linked to the breeding procedure, including the genetic modification. Also, it has been observed

Q/A	Comment/suggestion	Adopted in study plan	Explained in study plan	Not in study plan	Brief explanation/argumentation addendum
					that GM plants with intended effects that lead to clear physiological changes that could be identified in targeted compositional analyses did not lead to observed effects in animal feeding trials. GRACE thus aims to directly compare the two approaches on the same plant materials with the aim to provide more conclusive data on the added value of both types of studies (animal feeding trials vs omics analytical assessment) to assess both intended as well as unintended effects of the plant breeding procedures, including the genetic modification.
22	Why 2 trials? Clearly indicate in protocol, or make 2 protocols.		x		There is a study plan for each study. Two 90-day studies are conducted for MON810 to consider generic variability of the test and to increase the pool of "historical data". The timing of trials is due to the available ressources and the limited time frame of the project. (see Q&A 1, 25)
23	Why 2 trials in parallel and not one after the other?		x		
24	Replace the word “autopsy” with “necropsy”.	x			See throughout the revised Study Plan A and B

	Test and control crops				
25	Why 2 MON810 varieties?			x	Seed companies have introduced the MON810 event into a large number of commercial varieties by conventional breeding methods while tests may be perfomed with a selection of pre-market lines. It is a precondition for the risk assessment that such an approach demanded for authorisation of a GM crop is robust towards the gentic background and independent from the test run. This is simulated with the two trials and two MON810 varieties. The expected pattern of the data of endpoints should not lead to contradicting results in both trials or, vice versa, if there is a distinct effect caused by the event it should be indicated in both trials.
26	Why 2 commercial non-GM varieties in each of the trials (4 in total) (include purpose in protocol).			x	OECD (2000) stated if "using historical control data as an aid to evaluation, it must be kept in mind that “normal values” in haematological and clinical chemical measurements depend on the specific methods used to generate the data. Therefore, only values produced by the identical methods from the same laboratory are valid in such comparisons." Since the lab at SZU has not conducted any feeding trials with maize these comparative data are essentially lacking. Thus the "normal" variability (for controls) has to be assessed. The further use of historic data i.e. not including additional controls in further maize trials is depending on such a reliable data base. The testing of non-GM varieties will allow to identify the immanent level of uncertainty for the intepretation of differences between varieties in this lab. Pathological effects are not directly linked to the actual differences between the values of endpoints detected with different varieties, but with the absolute value and pathological thresholds. The typical absolute effect size(s) for each endpoint - i.e. the absolute values that trigger the pathological evaluation - are estimated and qualified by control groups. They allow to track for the comparabilty of different, independent trials.
27	Including commercial non-GM varieties creates data ‘noise’, masking any effect of the genetic modification			x	
28	Why not 6 non-GM varieties as in field trials?			x	The field trials required for the authorisation of GM crops intend to qualify effects of environmental variability in the *environmental* risk assessment especially with regard to the performance during cultivation. The number of control varieties and animals employed in the two 90-day feeding studies with MON810 maize varieties are part of the *toxicological* risk assessment and enable the data analysis to a sufficient degree in regards of statistical quality and data interpretation for toxicity and pathology. (See EFSA

					guidance notes)
29	Why not only 1 non-GM variety to reduce the number of animals?			x	(see Q&A 1, 26 - 28)
30	Why not include SmartStax (higher Bt content)			x	Whole-product feeding studies ought to test the safety of the whole product, including possible unintended effects of the genetic modification, and are not intended to test the safety of the newly expressed proteins. Targeted toxicity testing of the transgenic protein can be performed with purified protein.
31	Why not comparing with near-isogenic line only?			x	(see Q&A 1, 26 - 28)
32	Include details on the production plan of all test, control and reference substances.		x		A brief description is added to the study plan. The production is monitored and detailed protocols are kept at the test site responsible institution. See Study Plan A and B.
33	Replicates are lacking (impact of growing location), how to deal with this?			x	The set of varieties can shed some light on that issue as they allow to consider (subtle) compositional variations. The different varieties were cultured by the same farmer, following the same practices and in a very reduced area with the same type of soil, etc. . Additional locations for cropping would exceed the project ressources.
34	Define test substance: grain.	x			The test substances are maize grain or potato tubers; details see Study Plan A and B
35	Characterisation of the grain (components, methods).		x		Analysis will be done on the milled grain used for making the feed as well as the feed itself. This comprises a wide range of substances, including nutrients, anti-nutrients, toxins, secondary metabolites, contaminants, pesticide residues, mycotoxins, and GMOs. See Study Plan A and B

36	Samples of feed material to be retained in compliance with GLP and for potential future testing.	x			Grain samples, taken according to the iso protocol, have been kept before sending the material to the company preparing the feed; also samples of the feed after irradiation and after storage are to be retained. See Study Plan A and B
	Test System				
37	Why Wistar rats?			x	The Wistar/Han stock is widely used in toxicity testing and has no known disadvantages. Familiarity of the lab, toxicologist etc.(histopathologic team is familiar with the animal model); low incidence of spontaneous malformations and carcinomas; to be able to compare; to be used in extended feeding study; lives longer
38	Explain the reason for a surplus of 10 animals.			x	Each animal will be carefully examined clinically before being accepted. These will be available to replace any animals which are found not to be entirely normal.
39	Include precise description of the age at start (the younger the better)	x			4-5 weeks + 1 week of acclimatisation → 5 to 6 weeks old at study start. See Study Plan A and B
40	Include precise description on weight (keep differences small, to reduce variability in results; females vs. males).	x			See Study Plan A and B

41	Indicate in protocol the rationale for 16 animals per dose per gender vs. 12 ('sample size must be justified as necessary to achieve the scientific objective'): statistical power.			x	There is no perfect way of determining sample size in an experiment. In the past sample sizes in toxicity testing (such as the 90-day toxicity study for small chemicals in OECD 408) have been based on previous experience. In clinical trials sample size is based on power analysis. This requires an estimate of the effect size of clinical or scientific interest. When this is not available (e.g. with multiple outcomes as in toxicity testing) power calculations can be based on the standardised effect size (SES: the difference in means between control and treated groups divided by the standard deviation (SD)). In clinical trials SESs of 0.2, 0.5 and 0.8 SDs are regarded as "small", "medium" and "large" effect sizes. EFSA re-designed OECD 408 for evaluating the toxicity of food. Power was increased by reducing the number of doses and increasing the group size in the top dose and control groups. Animals are housed two/cage, with the cage being the unit of statistical analysis. It was judged that an SES of 1.0 SD or less is unlikely to be of toxicological importance. As a result the total number of animals was increased from 80 to 96, being three dose levels of 16 animals (8 cages) per gender and dose. For GRACE is obliged to evaluate the design of feeding studies, the approach suggested by the EFSA guidelines on group sizes will not be rejected a priori.
42	Include power analysis to estimate a sample size capable of detecting a pre-specified biologically relevant effect size, with a significance level. The data used for such analysis need to be clearly presented and evaluated.			x	See above (Q&A 41). We assume that a standardised effect size of 1.0 or less is unlikely to be of toxicological significance. Assuming a power of 0.8, a significance level of 0.05 and a two-sided test this would be achieved by 17 animals per group. We have rounded down to an even number as the rats are housed in pairs. There are two additional factors to take into account. First the experimental unit is the cage. Animals within a cage do not provide independent observations, but having two animals per cage reduces measurement error to an unknown extent. Second, in some circumstances it may be appropriate to average across the two sexes provided there is no sex by treatment interaction and the variances are homogeneous. One of the aims of the study will be to use the data to investigate these factors.
43	Use data from Monsanto trial: the variability found in that trial could be a			x	GRACE is following an extened design of the trials described by the EFSA Guidelines (2011) and OECD recommendations (2000). See also explanations on historic data and statistics (Q&A 26, 27; 41, 42). The number of animals used is indeed minimized while the whole setting of the

	factor determining the group size.				trial(s) should sufficiently satisfy demands on statistical power. The chosen strategy is already a compromise in between the different demands raised during the stakeholder consultation. The design will allow to recalculate the options based on the trial data and following recommendations can sufficiently be substantiated.
44	A sample size of 10-12 animals per group would provide a more reasonable balance between statistical power and animal welfare considerations.			x	
45	Consider one study with 16 animals per group, the second one with 12.			x	
46	Do the study with 16 animals, later compare with results of 12 randomly chosen animals.	x			
47	A group size of 16 gives only 8 experimental units, whereas up till now 10 independent units were used in the toxicological experiments.			x	
48	Animal housing: which environmental enrichment?	x			Throughout the studies plastic tunnels/tubes were added to each cage
49	Diet formulation, sampling and analysis: list the individuals responsible for the	x			See Study Plan A and B (Private data will only be disclosed according to law standards.)

	conduct of each activity.				
51	Description diet formulation.	x			See Study Plan A and B
52	Include specifics regarding sample size, number, identification, and storage conditions; shipping contact information; shipping conditions; and the analyses to be conducted on the samples.		x		
53	Include physical description of each diet		x		
54	Listing GMOs among known toxins and pesticides could be interpreted as prejudicial; suggest using "molecular identity" as a more neutral way of defining this analysis.			x	The concept of the risk assessment of whole food/feed derived from GMO picked up toxicity testing procedures developed for chemicals especially toxins and pesticides. It is the task of GRACE to reconsider this adapted approach.
56	Include analysis on GM content (intended and not-intended).	x			See Study Plan A and B

57	Determine Cry1Ab content in feed and different tissues (e.g. intestines and blood; allergenicity potential), which protocol(s) of ELISA? suggestion: publication by Szecaks et al. (2012) for plant material.		(x)	(x)	See Study Plan A and B (detection of Cry1Ab antibodies); further investigation cannot be covered by the budget
58	Test shape and colour of feed particles	x			Standard shape and colour of feed material is used
59	Determine protease inhibitors in diet, e.g. trypsin inhibitor.	x			See Study Plan A and B
61	Include analysis on vitamin D content (effect on the immune system).	(x)			See Study Plan A and B
62	Store at -24°C, or lower to prevent deterioration			x	See Study Plan A and B
	Experimental design				
63	use the word “group” or “treatment” and use it consistently	(x)			See throughout Study Plan A and B
64	Prior feed exposure: diet should be specified including an analysis for admixtures (to exclude GM	(x)		(x)	Following to the stakeholder consultations the analysis including transgenes in plants and diets has been extended (see Study Plan A and B; Q&A 59, 61, 62.). Yet, it resulted in time delays for testing as well as in substantial additional costs. Nevertheless, it was considered being neccessary in order to improve reliability of the generated data. The diet of femal parents was certified as being GMP-free by Harlan (provider of rats).

	components):				
65	The diet of female parent animals should be GM-free.	x			
66	Acclimatisation period is too long (OECD TG408: ‘dosing as soon after weaning as possible’): impact of the GM material on young, growing animals may be higher.		x		The protocoll is adjusted according to OECD TG 408 recommendations: "Dosing should begin as soon as possible after weaning and, in any case, before the animals are nine weeks old." see Study Plan A and B
67	Explain the choice of the statistical design (completely randomised design vs. randomised block design; males and females taken together or not, etc.).		x		A complete randomised block design was used. The young rats were homogeneous with a narrow weight range < 10% variation. See Study Plan A and B
68	If the range within each sex is as high as mean ±20% then this might be sufficient to justify introducing a blocking structure based on starting weights.		x		Agreed, but it is expected to be less than 20%; see Q 67

69	Include a description of randomisation method and allocation to treatment groups to ensure that starting weight is well-balanced across treatment groups.		x		The first five cages will be assigned at random to the five treatment groups using software that provides uniform random numbers (EXCEL statistical functions). The rest of the cages will similarly be assigned in groups of five. However, cages with the same treatment will be arranged in vertical columns. See Study Plan A and B
70	Cage rotations should be according to a set schedule.	x			Vertical randomisation of whole columns will be done weekly at random, using a formal randomisation process, as above. See Study Plan A and B
71	Diet ad libitum: specify with “Food will be supplied ad libitum (until the day before the animal is scheduled for necropsy)…”,	x			See Study Plan A and B
73	Why doses 0, 11%, 33% and not higher?			x	The provider of the diet (Harlan) suggested to limit maize to maximum 40% . Commonly used are 33%. The actual factor between 33% and 40% is toxicologically negligible. We choose the "standard" amounts. Nevertheless, the maximum amounts of plant material used should be reconsidered for each crop with regards to a well balanced diet and maximum "dosage".
74	Why not 16,5% (instead of 11%) and 33%?			x	The concentrations of plant material have been choosen to support the detection of any differences by dosage. The toxicological information gained when applying a middle dose for MON810 maize is likely of no use, since previous studies didn't indicate effects within this range and the spread of dosages is actual low (factor ~ 3) compared to tests with chemical compounds.
75	Why not 3 doses (11%, 22%, 33%)?			x	
76	Extra dose instead of a commercial non-GM variety (to keep animal numbers equal)			x	
77	Colour coding: avoid			x	See Study Plan A and B; colour coding as realised by the provider of the

	to use of red and green				diet.
78	Colour coding for blinding vs. explanatory table in protocol; how to maintain blinding?		(x)	(x)	See Study Plan A and B
	Periodical Observations				
79	Include detailed time schedule for examinations.	x			See Study Plan A and B
80	What are the cut-off criteria to withdraw an animal from the experiment (OECD guideline)?		x		See Study Plan A and B; OECD TG 408, 20.
81	Specify observations for moribund animals: "once in the morning and once in the afternoon" to ensure that the observations are conducted far enough apart to be of value in prevent loss of tissues to cannibalism or autolysis.		(x)	(x)	
82	Moribund animals should be weighed and physically examined before euthanasia and subsequent		x		

	necropsy.				
83	Include a description of the order in which animals are observed or sampled to avoid bias in time.			x	Each group was subdivided into two subgroups A and B ; all animals in subgroup A were observed/sampled on the first day, all animals of subgroup B were observed/sampled on the second day.
84	Add a heading to cover unscheduled observations, so that any findings noted outside the defined observation periods may be recorded without creating a protocol deviation.	x			In the used protocol there is a column for inclusion of clinical observations for each animal.
85	Cage side observations: clarify whether the animals will be removed from the cage for a close-up examination or not.	x			Daily observation of animals in the cage to avoid additional stress; see also Study Plan A and B
86	During acclimatisation: inspection of individual animals by a qualified veterinary technician upon receipt; twice daily observation of all animals for changes in general		x		All points were covered by the actual SOP.

	appearance or behaviour; conduct at least one detailed physical examination on individual animals to detect pre-existing conditions; and record body weights and food consumption to monitor general health.				
87	Detailed clinical observations (including functional assessment) should be made of all animals at least once prior to the first exposure. functional assessment weekly thereafter.		(x)	(x)	Detailed clinical observation except funtional assessment was performed prior to first exposure. Functional assessment was performed in week 10 according to the actual SOP.
88	Include reactions to sensory stimuli, assessment of grip strength and motor activity assessment; in any case this should not happen earlier than in week 11.			x	The analysis was perfomed in week 10 as ophthalmology was perfomed in week 11.

89	Include ophthalmologic examination also before the start and at the termination of the study, preferably in all animals but at least in the high dose and control groups.	x			See Study Plan A and B
90	In measuring feed consumption feed spillage must be taken into account.			x	Controlled, but spillage did not occur because the feed was completly consumed
91	Having 2 animals in one cage may not distinguish between equal, average consumption, and very unequal consumption between the 2 rats.			x	Two animals per cage is required by 210/63/EU animal welfare legislation. The procedure followed in the study is in line with the EFSA recommendations and the statistical analysis is adjusted considering one cage as experimental unit.
92	Additional analyses: plant DNA, RNA and proteins in blood and organs.			x	The presence of plant DNA, RNA and proteins may be interesting in subsequent targeted analyses. The intention of the 90-day feeding study of whole food/feed as promoted in EU legislation is a non-targeted approach.
93	Immune reactions and impact on the hepatorenal system should be investigated	x			see Study Plan A and B

94	Clinical pathology: for the differential blood cell count all cell types should be counted. If the instrument is not appropriate to measure all cell types, a blood smear should be made and evaluated microscopically.	x			Blood smears have been prepared and will be evaluated microscopically.
95	Clinical pathology: the 3 components of the Differential Leukocyte Count should be indicated.	x			
96	Clinical pathology: also include reticulocytes in haematology evaluations. Some anticoagulants can interfere with APTT and PT assays.			x	It is not included in OECD or EFSA guidelines.
98	The selection of hepatocellular enzymes is limited and it would be recommendable to measure more liver and bile enzymes.			x	The number of enzymes measured complies with OECD TG 408; see Study Plan A an B
99	Clinical chemistry: detail how fasting will be conducted.			x	Animals were fastened for 18 h.

100	Clinical chemistry: inconsistency (plasma samples vs. tests on serum)			x	Serum samples for clinical chemistry and plasma samples for metabolomic and immunological analyses. Compare with Study Plan A and B.
102	Include description of collection procedures for the urine and blood samples	(x)		(x)	Blood samples for hematology taken from tail vein, blood samples for biochemistry were taken from abdominal vein; urine samples were not collected and are not included in the Study Plan A and B.
103	Additional urinalysis parameters: leukocytes, nitrites and specific gravity			x	Urine samples were not collected and are not included in the Study Plan A and B.
104	Typos to be corrected: aspartate, alanine, bilirubin.	x			fully adopted in the revised Study Plan A and B
	Metabolomics				
105	The objective of the metabolomic analyses needs to be clearly stated in the protocol.			x	Immunologic and metabolomic analyses on animals, in connection and in parallel with the 90-day feeding trials, aim to identify possible early indicators of potential unintended (adverse) effects of the genetic modification regarding altered immune response to the newly expressed protein and to endogenous proteins of the GM maize (targeted approach) and more generally disruption or alteration in metabolic pathways using non-targeted metabolomic analysis of body fluids and tissues/organs of GM-fed rats in comparison with rats fed conventional counterpart . The investigations with in vitro systems at FUB aim to investigate the expression pattern of specific cell types treated with GM feed. Transcriptomic, proteomic methods as well as vitality parameters will be used to identify cellular effects in that system. Metabolomic approaches will not be part of these investigations. (See also Q&A 21.)
106	Metabolomics: indicate whether or not the analyses will be validated and/or GLP-compliant; if GLP, identify a PI		x		Analyses will be performed using the quality reference system developed and used for research and experimentations at INRA in order to meet the objectives of INRA's quality policy : traceability of research activities and reliability of measurable results.

	and QA professional.				
107	Metabolomics: indicate the INRA study number			x	NA
109	Metabolomics: the samples should be covered with argon or nitrogen to avoid oxidation of some metabolites.			x	Samples are kept at - 72°C.
110	Metabolomics: preserve some additional cell types, e.g., intestinal cells, for ex vivo studies, as generally outlined in WP 2			x	Not foreseen in the study plan for immuno/metabolomic studies (because of technical difficulties and possible pitfalls); Additional cells might be stored if subsequent analysis appears necessary. However, additional metabolomic studies will significantly increase the workload and exceed the financial ressources. (See also Q&A 105.)
	Pathology				
111	Method of euthanasia should be briefly mentioned next to referring to SOP.	x			See Study Plan A and B
112	Indicate procedure for euthanization and subsequent necropsy to minimise temporal bias.			x	To avoid temporal bias during necropsy 12 people worked in parallel.
113	Representative tissue sampling is to be done in parallel for the different approaches.			x	Question appears unclear.

114	More details on the tissue dissection should be provided (e.g., the dissection scheme, the variation among dissectors, the criteria used to decide possible needs for additional tissue types).			x	Dissection schemes are available in the laboratory where the necropsy takes place. Individual dissectors were responsible for specific tissues throughout all groups. Additional tissue types were selected after weighing as requested for omics studies by partners.
116	Inaccuracies in describing the organs to be collected and examined for pathology; e.g. epididymides, Peyer's patches	x			see corrections Study Plan A and B
119	Indicate the state of the organs (fresh or fixed) at the time of weighing.	x			fresh weight
120	Indicate whether paired organs are weighed together or not, etc.			x	Organs were weighed seperatly.
121	Paired organs: some organs are mentioned in singular (thyroid, parathyroid, adrenal) when compared to others (gonads, kidneys, eyes). It should be homogeneous.	x			The wording has been revised.

123	Measure the carcass weight to reduce variability due to gut content.			x	Total weight according to OECD TG 408.
124	Weigh the gut tissue in addition to performing histology, because the empty intestinal weight is a very good indicator of the effect of 'antinutrients' in the diet.			x	Gut was not weighed according to OECD TG 408.
125	Suitability of fixatives for certain tissues is questioned.			x	Based on the long-lasting experience at SZU the fixation in formalin is considered being appropriate.
126	Testis, epididymis, eye, optic nerve and harderian gland should be preserved in Davidson's fixative instead of Formalin.			x	
130	Allergenicity potential of Bt toxin: focus on intestine histology, digestive and gut wall enzyme activities, stress cell and immune responses or allergenicity markers.			x	The approach follows a non-targeted protocol and specific parameters for allergenicity are not included in OECD TG 408 and the EFSA guideline.
131	The fate of work products and study records should be added to the document.	(x)			Specialised companies take care of the disposal of work products. Records are archived as described in the Study Plan A and B.

132	Quality of the analysis: peer-review of the micropathology by a second pathologist.			x	A second independent pathologist will perform the peer-review of slides.
135	Raw data will be made available; pictures on the website (possibly selection rules, if selection is needed for technical reasons, set beforehand).			x	Raw data will be made available, images from the histopathology cannot be provided because of extremely high costs which are not covered by grant.
	Data evaluation and statistical analysis				
136	Consider the statistical methods used by Séralini.			x	A major criticism of the paper by Seralini et al is that they did not do *any* statistical analysis of their data on qualitative variables (tumour incidence and pathological lesions). Had they done so they would have found no significant differences among groups because their sample (group) size was too small to allow any sensible analysis of such data. With the quantitative variables such as haematology and clinical biochemistry they did not present tables of means and standard deviations as is usual in such studies and they used novel statistical methods which have not been used previously in similar studies, and whose validity can't be assessed with the information that they provided. They should use their data to explain the methods that they have used in a form which is open to scientific assessment.
137	Each endpoint should be analysed and interpreted separately.	x			Each endpoint will be considered for toxicological and pathological implications. Yet, single endpoints - i.e. differences observed in single endpoints, respectively - might not sufficiently explain nor display toxicological and pathological effects or patterns.

138	Complete description of the statistical approach for the analysis of the different endpoints should include and not be limited to: the hypothesis tested, the method selected, the verification of the assumptions (normality, independency, homogeneity of variance), the description of the model chosen, if any.		x		Data will be screened for outliers and any errors corrected (but genuine outliers are not to be removed). Descriptive statistics (means and standard deviations) will be calculated separately for each sex for each endpoint. The null hypothesis that there are no differences among treatment means will be tested using a two-way analysis of variance without interaction (to remove any block effect) separately for each sex. Residuals plots will be used to evaluate assumptions of normality of residuals and homogeneity of variances. Where neccessary a suitable scale transformation will be used. Analyses will be done with and withoput any outliers. If the conclusions are the same, the outlier will be retained. If the outlier changes the conclusions then this will need further investigation (e.g. looking at correlations between characters). Where statistically significant ($p<0.05$) effects are detected Dunnett's test will be used to compare the treated group means with the controls. Data will be presented graphically as differences between control and treated groups with a 95% confidence interval. The results will need to be interpreted by an experienced toxicologist as some false positive results are expected given the large number of endpoints to be measured. See Study Plan A and B.
140	It is not clearly justified why the statistical analysis of all the endpoints should be separated by gender. If gender would be included as a main factor in the statistical model, then the dependent variable could be adjusted for gender as a covariate. If the problem lies in unequal variances, i.e., heteroscedasticity in			x	We understand that this is how it is done in industry. There are advantages in analysing all endpoints in the same way. We want to keep the statistical analysis as simple as possible, provided power is not lost. However, part of the aims of the project are to evaluate ways of analysing and presenting the data, so the raw data will be made available to anyone who wants it. (see Q&A 136)

	either the male or the female group or both, then a statistical analysis free of any homoscedasticity assumption should be applied and described in the study plan.				
141	The statistical analysis should not be defined on a case-by-case basis, rather it should be determined in advance, during the design of the study.	x		x	A statistical analysis will be carried out using the usual industry standard statistical methods (i.e ANOVA, test of homogeneity of variance, transformations where appropriate, post hoc comparisons etc.). These methods are understood and defined *a priori*. The results will be presented as tables of means and standard deviations with indications of statistical significance in the usual manner. However, the raw data will also be available for any further statistical analyses as part of continued research into methods of analyzing such experiments.
142	The historical data that will be used during the course of the study, both for the power analysis and during the evaluation of results, should be properly described and evaluated for their quality.			x	There are no historic data for toxicity test with GM whole food/feed for the lab. The historical data can not be used in a power analysis. It is not possible to do a power analysis separately for each endpoint.

143	A statement should be included that adverse effects cannot be identified on the basis of statistical significance alone, and that toxicologically relevant findings may not attain statistical significance for isolated endpoints.			x	Statistical significance does not imply toxicological importance. The latter has to be assessed by a toxicologist taking into account the route and level of administration, the dose level, species differences and likely human exposure. Conversely, lack of statistical significance can't be taken alone as evidence of safety. An experiment may not have sufficient power to detect some toxicologically important adverse effects due to small sample size, inadequately controlled variation and species and strain differences. The best that can be done is to look hard for any adverse effects. If these are not found then the test compound can provisionally be assumed to be safe.
	References				
144	Check software references	x			done
145	Check OECD TG408 reference	x			done
	Reporting				
146	Detail on a summary of the study, test and control material data, methods and procedures, appropriate individual animal, and summary data tables, graphs of weekly body weights and food consumption for both sexes, and an interpretation and discussion of the study results. Also, it should contain the results of the diet		x		Data generated in the course of and in conjunction with the feeding trials will be made available at the project web site as far as logistically possible (see Q&A 136). Results will be presented and discussed by further stakeholder consultations. Report topics are summarized in the Study Plan A and B.

	analyses and the histopathological processing and microscopic examination, as well as the name of the conducting laboratories. the report should be comprehensive, attempting to define the differences between the control diet and test diet groups.				

Part II: List of open questions

Q/A	Comment/suggestion - answers pending
	General
3	Make use of existing best practices (see ESTP comments)
	On the protocol
	Staff, GLP, data recording
17	Include the qualifications of the pathologist performing the histopathology evaluation. Does this individual possess appropriate board certification as a veterinary pathologist?
	Objective
	test and control crops
	Test System
50	Diet preparation: description of the milling procedure, targeted particle size.
55	Certification: GM-free diet (except for the event to be tested).
60	Diet preparation: 16% proteins is too high and should be adapted to the rat's life stage.
	Experimental design
72	Suggestion to provide a mechanism for replacing animals either prior to or shortly after initiation of dosing. Replacement animals are typically selected from the remaining pre-test animals, and the reasons for replacement appropriately documented in the study records.
	Periodical Observations
97	Clinical pathology: reference of sampling on citrate tubes.
101	Calculate globulin and albumin/globulin ratios from the parameters already included in the protocol.
	Metabolomics
108	Metabolomics: details on sample size, number, identification, and storage conditions; shipping contact information; shipping conditions; and the analyses to be conducted.
	Pathology
115	More details on the histopathology diagnosis should be provided (e.g., the number of pathologists involved, the number of slides per tissue per depth).
117	Pituitary gland, prostate, thyroid glands (including parathyroid glands) should be weighed.
118	Other tissues should be defined more precisely instead of using general terminology; e.g. accessory sex organs; cervix, epididymides, seminal vesicles, ventral prostate, and vagina. For the same reasons specific lymph nodes (mandibular and mesenteric); one peripheral nerve (sciatic); and sections from the cervical, thoracic, and lumbar spinal cord; stomach (forestomach and glandular); small intestine (duodenum, jejunum, ileum); large intestine (cecum, colon, rectum); skin (including mammary gland); and eyes with optic nerve and harderian glands should also be clearly indicated. which salivary gland should be sampled (submaxillary, parotid)?

122	Some organs should be deleted from the list of organs (lungs, pancreas, gallbladder!) to be collected (and weighed), others added (pituitary gland, prostate, thyroid glands including parathyroid glands). Pituitary gland should be separated from brain. Skeletal muscle adjacent to the peripheral nerve.
127	Include in the protocol staining and embedding procedures.
128	Follow uniform trimming guideline for histological processing of the tissues sampling intestines (see EuropaBio comments)
129	Specify slide processing.
133	Quality of the analysis: audit of lab, peer review of raw data / slides, additional expert for independent review, offer by ESTP.
134	The peer review pathologist should also be included in the professional staff.
	Data evaluation and statistical analysis
139	Check for inconsistences in terms of proposed methodology in the data evaluation and statistical analysis section.
	References
	Reporting
147	Peer Review Statement