Issues to be Addressed by Technical Experts to Define Molecular Marker Sets for Establishing Thresholds for ISF EDV Arbitration

1) Sampling of individual varieties

How many individuals of a variety should be profiled? Should the molecular profiles of a single variety be examined using individual profiles of DNA extracted from individual plants or can DNA from several plants of the same variety be examined in one bulk extract? If bulks can be used then what is the minimum number of individuals that should form the bulk for each variety? And what is the maximum number of individuals for each bulk? Should tissues from individual plants be bulked before DNA isolation? Or should DNA from individual plants be bulked after extraction from individual plants? How many DNA bulks does a variety need? The breeding system of the crop will need to be taken into account. The error rate of the marker system will likely have to be taken into consideration, i.e. the ability to detect rare alleles in a bulk sample. Is it important to be able to detect rare alleles or is there a threshold of allele frequency below which alleles need not be detected? If so, what is that threshold? Should there be at least an initial study of heterogeneity using a representative sample of varieties? How will data from heterozygous loci (for inbred lines and varieties of self pollinated crops) be dealt with? Should there be minimum standards required for homozygosity?

2) Gauging discrimination power of markers

Which varieties should be used to establish a set of varieties that will be used to gauge the discriminative power of candidate molecular marker loci, the genomic coverage of markers, and the number of markers that should constitute a panel for determining information useful for establishing EDV status? Will related species (germplasm) be included for analysis? Which varieties would constitute a survey of the relevant breadth of genetic diversity likely to be encountered in future EDV determinations? Is it important to use varieties that have already been established as distinct, uniform and stable? Is it important to include some varieties that could be regarded as EDVs due to known close pedigree relationships and/or morphological similarity to other varieties included in the set of “standard” varieties? Is it important to include varieties of well-known pedigree relationships (i.e. publicly bred inbred lines of known pedigree)? Is it necessary to include varieties that represent different eras in the history of breeding of the crop? Is it important to include varieties that are of significant importance to the industry (grown on large area) at different periods although they may not necessarily represent unique germplasm?

3) Type of molecular marker system

What are the acceptable types of molecular markers? Is it essential that the marker is co-dominant (detecting both alleles at one locus)? Will non-PCR based markers be excluded? If more than one marker type is potentially usable then what criteria would be used to prioritise the choice of marker type? Is it important to understand the genetic control of the markers? What criteria and standards should be used to determine identity, type, number, and genomic coverage of molecular markers? Should markers come from expressed regions of DNA, from anonymous coding regions, from regions of repeat DNA, or from a combination of chromosomal regions? Is map information needed? Should each chromosome arm be sampled? What density of chromosomal coverage is required?

4) Eligibility of markers

What minimum criteria must markers pass to be considered as candidates for inclusion in the set used for EDV? How to determine? Must profiles be interpretable in terms of alleles? What minimum discrimination power (polymorphic index content) is required? Do certain criteria rule out eligibility, e.g. a certain level of stutter bands? For polyploid crops (such as cotton) will the monomorphic fragment be included in data analysis although polymorphism is revealed by the same primers (at the same or different locus)? Are markers available publicly or against licenses, or not available to third parties?

5) PCR conditions

Should specific PCR conditions be specified? If so, what are these? E.G. addition of +A. Is it possible to use one or a very few PCR conditions for the same type of molecular marker system (e.g. SSR)? Should PCR based markers be amplified singly in individual tubes or wells or can they be amplified in a multiplexed format to increase throughput and provide cost efficiencies? If Yes, then should the, multiplex be carried out before PCR or only after PCR amplification of individual loci?

6) Gel running conditions

Should PCR based markers that require molecular size to be measured be electrophoresed singly in individual gel lanes or capillaries or can they be run in a multiplexed fashion to increase running speed and provide more efficient running costs? If markers can be multiplexed then are there specific criteria that must be observed? If so, what are these criteria?

7) Recording data

What molecular weight standards are needed as a basis from which to measure molecular sizes of amplified fragments? What threshold criteria should be used to determine whether an amplified product is of significant magnitude to warrant recording? What are the procedures and equipment that should be used to objectively record molecular weights? How should data be recorded - as molecular sizes or as discrete binned data? If data should be recorded in terms of discrete bins then what are the criteria for establishing bin sizes? What standards should be used to validate marker identity and to check repeatability of scoring and databasing?

8) Measuring distances

Which statistical methods should be used to measure genetic distances? Should the data be analysed as qualitative data (presence vs. absence) or as genetic data (alleles at specific loci)? How are data used from loci that exhibit two or more (a possibility if varieties are comprised of bulked individuals) alleles per locus as compared to loci where a single amplified product is recorded?

9) Data quality

Will two (or more?) laboratories conduct experiments to ensure the data quality? How will those experiments be planned to ensure there is no bias in data generation, scoring or interpretation? What means (if any) are needed to ensure or to at least evaluate consistency of data generation, scoring and interpretation?

10) Data use and availability

How will the data be used? Will the data be accessible to the public, or only available on a restricted basis, and to whom?