APPENDIX I:
MOLECULAR GENETIC CHARACTERIZATION DATA

INTRODUCTION

In July of 1998, regulatory officials of the Canadian Food Inspection Agency (CFIA), Health Canada, and the United States Department of Agriculture's Animal and Plant Health Inspection Service (USDA-APHIS) met to compare, and harmonize where possible, aspects of molecular genetic characterization that are part of their review processes for transgenic plants. Agreement on common requirements and acceptable analytical approaches for molecular genetic characterization will facilitate the submission of supporting data by developers seeking regulatory approval to incorporate such plants into agricultural production or commerce in both countries. This Appendix is one of the outcomes of this meeting. The Appendix summarizes and identifies similarities and differences in the critical elements of the molecular genetic characterization of transgenic plants considered during the review process by these participating agencies. Molecular genetic characterization is only part of the information considered during assessments of such plants conducted prior to commercialization.

The scope of this document is limited to consideration of the transformation process and vectors used during transformation; the genetic material that was potentially delivered to the recipient plant; the identification, inheritance, and expression of the genetic material in the transgenic plant, and the production of new proteins encoded by the introduced genetic material. This document does <u>not</u> address specific types of techniques nor quality assurance practices (e.g., good laboratory practices) that are used to generate molecular genetic characterization data.

The agencies found very substantial areas of agreement in the types of molecular genetic characterization data they require to be submitted and considered. In addition to the specific data sets reviewed, the participants of both countries reaffirmed that reviews are still conducted on a case-by-case basis which allows for reviewing additional or fewer data sets, depending upon the individual case and the regulatory authority of the individual agencies. The use of the word "may" in this document is intended to reflect some of this flexibility in determining when data sets will be considered as an appropriate part of the entire application package. Therefore, consultations between regulatory agencies and individual applicants are considered to be an important part of the overall application process in making such determinations.

The critical elements of the molecular genetic characterization of transgenic plants described below apply to the review process of the participating agencies in both Canada and the United States, except where noted. The contents of this document will be reviewed and amended as necessary by these agencies. The glossary which follows has been included to provide definition to certain terms within the context of this document.

GLOSSARY

carrier DNA DNA used to expedite the preparation or the transformation of genetic material into a plant but which is itself not part of the construct.

coding region A DNA sequence which can be translated to produce a protein. Synonymous with open reading frame.

construct An engineered DNA fragment (eg. plasmid) which contains, but is not limited to, the DNA sequences to be integrated into a target plant's genome.

database citations Publicly accessible sources of nucleotide or protein sequence information. Four commonly used databases and their website addresses are:

> **GenBank**: An annotated collection of all publicly available DNA sequences maintained by the National Institute of Health (NIH). http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html
>
> **DNA Data Bank of Japan**: The officially certified DNA bank of Japan, which collects DNA sequences from researchers. http://www.ddbj.nig.ac.jp/fromddbj-e.html
>
> **EMBL Nucleotide Sequence**: A database of DNA and RNA sequences collected from the scientific literature, patent applications, and directly submitted from researchers and sequencing groups. http://www.ebi.ac.uk/ebi_docs/embl_db/ebi/topembl.html
>
> The **SWISS-PROT Protein Sequence Data Bank**: A database of protein sequences produced collaboratively by Amos Bairoch (University of Geneva) and the EBI. http://www.ebi.ac.uk/swissprot/

insert That part of a construct (see above) which is integrated into the recipient plant's genome.

non-coding region DNA sequences which lie outside of an open reading frame and which are not translated to become part of a protein. These might include scaffold attachment regions, promoters, leader sequences, enhancers, introns, terminators, and any other sequences that are used for gene expression either in the plant or other hosts. such as origins of replication, transposable elements, T-DNA borders, *lox* sequences, etc..

stability	The ability of the transgenic trait to be expressed in the transformed plant line and plant lines derived therefrom in a consistent, reliable, and predictable manner.
trait(s)	The phenotypic characteristic(s) conferred to the recipient plant by the transgenic insert.
vector	An automonously replicating DNA molecule into which foreign DNA is inserted and then propagated in a host cell.

MOLECULAR GENETIC CHARACTERIZATION OF TRANSGENIC PLANTS

1 THE TRANSFORMATION SYSTEM

1.1 Description of the transformation method

1.1.1 Describe and provide references for the transformation method, e.g. *Agrobacterium*-mediated transformation or direct transformation by methods such as particle bombardment, electroporation, PEG transformation of protoplasts, etc.

1.1.2 For direct transformation methods, describe the nature and source of any carrier DNA used.

1.1.3 For *Agrobacterium*-mediated transformation, provide the strain designation of the *Agrobacterium* used during the transformation process, and indicate how the Ti plasmid based vector was disarmed, and whether *Agrobacterium* was cleared from the transformed tissue.

1.1.4 For transformation systems other than *Agrobacterium*, provide the following information:

1.1.4.1 Does the system utilize a pathogenic organism or nucleic acid sequences from a pathogen?

1.1.4.2 How were any pathogenesis-related sequences removed prior to transformation?

1.1.4.3 Did the transformation process involve the use of helper plasmids or a mixture of plasmids? If so, describe these in detail.

1.2 Description of the genetic material potentially delivered to the recipient plant material (the modification/constructs).

1.2.1 Provide a summary of all genetic components which comprise the vector including coding regions, and non-coding sequences of known function (see Table 1). For each genetic component provide a citation where these functional sequences were described, isolated, and characterized (publicly available database citations are acceptable) and indicate:

1.2.1.1 The portion and size of the sequence inserted.

1.2.1.2 The location, order, and orientation in the vector.

1.2.1.3 The function in the plant.

1.2.1.4 The source (scientific and common, or trade name, of the donor organism).

1.2.1.5 If the genetic component is responsible for disease or injury to plants or other organisms, and is a known toxicant, allergen, pathogenicity factor, or irritant.

1.2.1.6 If the donor organism is responsible for any disease or injury to plants or other organisms, produces toxicants, allergens or irritants or is related to organisms that do.

1.2.1.7 If there is a history of safe use of the source organism or components thereof.

1.2.2 If there has been a significant modification that affects the amino acid sequence of genes designed to be expressed in the plant, provide the citation. If the modified amino acid sequence has not been published, provide the complete sequence highlighting the modifications. Modifications that affect only a few amino acids can simply be stated without providing the complete sequence. Indicate whether the modifications are known or expected to result in changes in post-translational modifications or sites critical to the structure or function of the gene product.

1.2.3 Provide a detailed map of the vector (see Figure 1) with the location of sequences described above that is sufficient to be used in the analysis of data supporting the characterization of the DNA, including as appropriate the location of restriction sites and/or primers used for PCR and regions used as probes.

2 INHERITANCE AND STABILITY OF INTRODUCED TRAITS WHICH ARE FUNCTIONAL IN THE PLANT

2.1 For plants which are either male or female fertile or both, provide data that demonstrates the pattern and stability of inheritance and expression of the new transgene traits. If the new trait can not be directly measured by an assay, it may be necessary to examine the inheritance of the DNA insert directly, and expression of the RNA.

2.2 For plants which are either infertile or for which it is difficult to produce seed (such as vegetatively propagated male-sterile potatoes), provide data to demonstrate that the transgene trait is stably maintained and expressed during vegetative propagation over a number of cycles that is appropriate to the crop.

3 CHARACTERIZATION OF THE DNA INSERTED IN THE PLANT

3.1 For all coding regions, provide data that demonstrate if complete or partial copies are inserted into the plant's genome. Coding regions may include truncated sense constructs,

sequences engineered to be nontranslatable, antisense constructs, and constructs containing ribozymes, regardless of whether or not the coding region is designed or expected to be expressed in the transgenic plant. For Canadian submissions, information may be required indicating the number of copies which have been inserted, including integration of partial copies; and for allopolyploid plants, information indicating into which parental genome insertion has occurred.

3.2 For noncoding regions associated with the expression of coding regions:

3.2.1 Data should demonstrate whether or not plant promoters are inserted intact with the coding regions whose expression they are designed to regulate.

3.2.2 DNA analysis may be necessary for introns, leader sequences, terminators, and enhancers of plant-expressible cassettes.

3.2.3 DNA analysis may be necessary for promoters and other regulatory regions associated with bacteria-expressible cassettes.

3.3 For noncoding regions which have no known plant function and are not associated with expression of coding regions:

3.3.1 DNA analysis may be required for some sequences of known function (e.g., *ori V* and *ori-322*, *bom*, T-DNA borders of *Agrobacterium*, and bacterial transposable elements).

3.3.2 DNA analysis is not required for any remaining sequences of the plasmid backbone.

4 PROTEIN AND RNA CHARACTERIZATION AND EXPRESSION

4.1 For all complete coding regions inserted, provide data that demonstrates whether the protein is or is not produced as expected in the appropriate tissues consistent with the associated regulatory sequences driving its expression (e.g., if the gene is inducible, determine if the gene is expressed in the appropriate tissues under induction conditions). For virus resistant plants where the transgenes are derived from a viral genome, in addition to transgene protein analysis, determine transgene RNA levels in tissues consistent with the associated regulatory regions driving expression of the transgene. The following exceptions also apply:

4.1.1 If the protein concentration is below the limits of detection, mRNA data may be substituted.

4.1.2 Protein analysis for products of genes used only as selectable markers may be waived under certain circumstances, e.g. when there is at least one complete copy of a selectable

marker gene present and the effective expression of the selectable marker gene is verified by the process used to select the transformed tissue.

4.1.3 For plants modified to express nontranslatable mRNA, truncated sense constructs, antisense constructs, or constructs containing ribozymes, since the function of these genetic constructs is to specifically alter the accumulation of a specific mRNA or protein present in the transgenic plant, provide data on the level of the target protein only (e.g. native tomato fruit polygalacturonase would be the target protein of antisense polygalacturonase to achieve altered fruit ripening). If the target protein levels are below levels of detection, determine target mRNA levels.

4.2 When a fragment of a coding region designed to be expressed in a plant is detected, determine whether a fusion protein could be produced and in which tissues it may be located.

4.3 Protein or RNA characterization may not be required for fragments of genetic constructs not expected to be functional in the plant (e.g., fragments of selectable marker genes driven by bacterial promoters.)

Table 1: Example of a Table Describing the DNA Components of a Vector

Summary of DNA Components in PV-STBT02 (from Table III.1 from APHIS petition # 94-257-01p)

Genetic Element	Size[1], Kb	Function and Source
RB	0.36	A restriction fragment from the pTiT37 plasmid containing the 24 bp nopaline-type T-DNA right border used to initiate the T-DNA transfer from *Agrobacterium tumefaciens* to the plant genome (Depicker *et al.*, 1982)
E35S	0.62	The cauliflower mosaic virus (CaMV) promoter (Odell *et al.*, 1985) with the duplicated enhancer region (Kay *et al.*, 1987).
cryIIIA	1.8	The gene which confers resistance to CPB. The gene encodes an amino acid sequence identical to the CPB control protein (referred to as the *B.t.t.* Band 3 protein) found in *B.t.t.* as described by Perlak *et al.* (1993).
E9 3'	0.63	A 3' nontranslated region of the pea ribulose-1,5-bisphosphate carboxylase small subunit (rbcS) E9 gene (Coruzzi *et al.*, 1984), which functions to terminate transcription and direct polyadenylation of the *cryIIIA* mRNA.
NOS 3'	0.26	A 3' nontranslated region of the nopaline synthase gene which functions to terminate transcription and direct polyadenylation of the *nptII* mRNA (Depicker *et al.*, 1982; Bevan *et al.*, 1983).
nptII	0.79	The gene isolated from Tn5 (Beck *et al.*, 1982) which encodes for neomycin phosphotransferase type II. Expression of this gene in plant cells confers resistance to kanamycin and serves as a selectable marker for transformation (Fraley *et al.*, 1983).
35S	0.32	The 35S promoter region of the cauliflower mosaic virus (CaMV) (Gardner *et al.*, 1981; Sanders *et al.*, 1987).
LB	0.45	A restriction fragment from the octopine Ti plasmid, pTi15955, containing the 24 bp T-DNA left border used to terminate the transfer of the T-DNA from *Agrobacterium tumefaciens* to the plant genome (Barker *et al.*, 1983).
ori V	1.3	Origin of replication segment for ABI *Agrobacterium* derived from the broad-host range plasmid RK2 (Stalker *et al.*, 1981).
ori-322/rop	1.8	A segment of pBR322 which provides the origin of replication for maintenance of the PV-STBT02 plasmid in *E. coli*, the replication of primer (*rop*) region and the *bom* site for the conjugational transfer into the *Agrobacterium tumefaciens* cells (Bolivar *et al.*, 1977; Sutcliffe, 1978).
aad	0.93	A fragment isolated from transposon Tn7 containing a 0.79 kb gene which encodes for the enzyme streptomycin adenylyltransferase that allows for bacterial selection on spectinomycin or streptomycin (Fling *et al.*, 1985).

1. Sizes are approximations.

Figure 1: Example of a detailed map of a plasmid vector (from APHIS petition #94-257-01p).

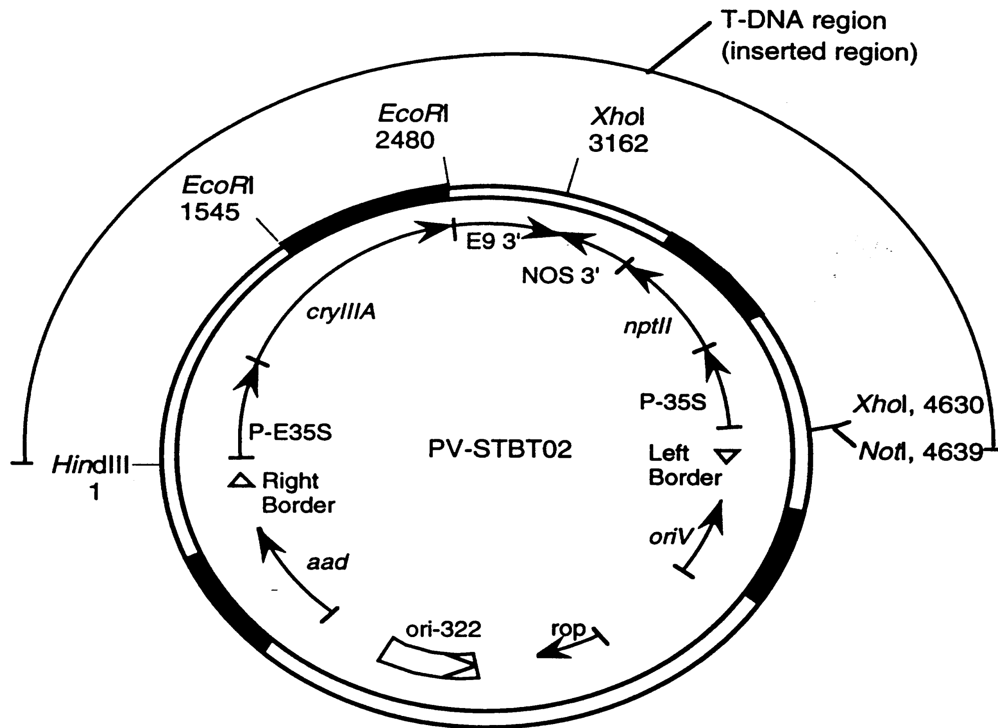

Figure III.1. Plasmid map of PV-STBT02.
Restriction sites, and their locations in base pairs, utilized during Southern analyses are shown. The region which served as the T-DNA is marked and its delineating right and le borders are denoted by open arrows. The blackened regions denote the positions of homology for PCR probes used during Southern analyses as described in Section V.A. Cleavage sites for *Hin*dIII, *Eco*RI, *Xho*I and *Not*I restriction endonucleases are shown. A description of the genetic elements appears in Table III.1.