



(19)中華民國智慧財產局

(12)發明說明書公開本

(11)公開編號：TW 201723480 A

(43)公開日：中華民國 106 (2017) 年 07 月 01 日

(21)申請案號：105136492

(22)申請日：中華民國 105 (2016) 年 11 月 09 日

(51)Int. Cl.：

G01N33/50 (2006.01)

C40B20/04 (2006.01)

G06F19/18 (2011.01)

C12N15/29 (2006.01)

(30)優先權：2015/11/10

美國

62/253,213

(71)申請人：陶氏農業科學公司 (美國) DOW AGROSCIENCES LLC (US)

美國

(72)發明人：王伯豪 WANG, POHAO (TW)；斯里拉姆 斯瑞德哈蘭 SRIRAM, SHREEDHARAN (IN)；沙士翠 丹特 拉克希米 SASTRY-DENT, LAKSHMI (US)

(74)代理人：惲軼群；劉法正

申請實體審查：無 申請專利範圍項數：14 項 圖式數：1 共 44 頁

(54)名稱

用於預測轉殖基因靜默之風險的方法及系統

METHODS AND SYSTEMS FOR PREDICTING THE RISK OF TRANSGENE SILENCING

(57)摘要

本發明係基於使用植物基因組中 H3K9me2 甲基化程度來預測轉殖基因靜默、轉殖基因穩定性、及/或轉殖基因表現程度。所提供者為用於產生全基因組 H3K9me2 圖譜及其與一指定閾值一起用於預測基因靜默的方法及/或系統。提供於此之該等方法及/或系統可被使用於高通量設置中，用以在與現有技術相比一相對短的時間期間內篩選大量轉型事件。

The present invention is based on the use of the H3K9me2 methylation levels in plant genomes to predict transgene silencing, transgene stability, and/or transgene expression level. Provided are methods and/or systems for generating whole-genome H3K9me2 maps and its use with an assigned threshold value for predicting gene silencing. The methods and/or systems provided herein can be used in high-throughput setting for screening large number of transformed event in a relatively short period of time as compared to existing technologies.

【發明摘要】**【中文發明名稱】**

用於預測轉殖基因靜默之風險的方法及系統

【英文發明名稱】

METHODS AND SYSTEMS FOR PREDICTING THE RISK OF TRANSGENE SILENCING

【中文】

本發明係基於使用植物基因組中H3K9me2甲基化程度來預測轉殖基因靜默、轉殖基因穩定性、及/或轉殖基因表現程度。所提供者為用於產生全基因組H3K9me2圖譜及其與一指定閾值一起用於預測基因靜默的方法及/或系統。提供於此之該等方法及/或系統可被使用於高通量設置中，用以在與現有技術相比一相對短的時間期間內篩選大量轉型事件。

【英文】

The present invention is based on the use of the H3K9me2 methylation levels in plant genomes to predict transgene silencing, transgene stability, and/or transgene expression level. Provided are methods and/or systems for generating whole-genome H3K9me2 maps and its use with an assigned threshold value for predicting gene silencing. The methods and/or systems provided herein can be used in high-throughput setting for screening large number of transformed event in a relatively short period of time as compared to existing technologies.

【指定代表圖】 (無)

【代表圖之符號簡單說明】

(無)

【特徵化學式】

(無)

【發明說明書】

【中文發明名稱】

用於預測轉殖基因靜默之風險的方法及系統

【英文發明名稱】

METHODS AND SYSTEMS FOR PREDICTING THE
RISK OF TRANSGENE SILENCING

【技術領域】

【0001】本發明係有關於用於預測轉殖基因靜默之風險的方法及系統。

【先前技術】

【0002】發明背景

在1990年代，如大豆或玉米植物之植物的基因組被成功地以轉殖基因轉型。過去二十年間，眾多的方法學被發展用以轉型植物的基因組，其中一轉殖基因被穩定地整合進入植物的基因組。植物轉型方法學的演進導致可成功地引入包含一農藝性狀的一轉殖基因於植物基因組內的能力。引入昆蟲抗性及除草劑耐受性狀於植物內提供生產者一新穎且方便的用於控制昆蟲及廣譜雜草之技術創新，其在農業栽培方法中係無可比擬的。

【0003】現今的轉型方法學依賴隨機插入轉殖基因於植物的基因組中。由於基因轉殖事件可能隨機地整合於基因轉錄序列中，如此的事件可能中斷內生性性狀的表現且改變該植物的生長與發育。此外，該等基因轉殖事件可能不加選擇地整合進入該基因組容易基因靜默的位置，最終在基因轉殖植物的第一或其後之世代中造成轉殖基因表

現的降低或全然抑制。最後，該隨機地整合轉殖基因於植物基因組中需要相當大的努力與花費在辨識該基因轉殖事件的位置並選擇按設計執行且不對該植物造成農藝衝擊的基因轉殖事件。

【0004】因此，需要對評估轉殖基因之整合位置就基因表現的有效性及/或對預測轉殖基因靜默風險有益的發明。

【發明內容】

【0005】發明概要

本發明係有關於用於預測轉殖基因靜默風險、及/或預測轉殖基因表現程度、及/或預測轉殖基因整合位置的穩定性及/或有效性的方法與系統。在一方面，所提供者為用於預測轉殖基因靜默風險的方法。該方法包含：

- (a) 產生一植物基因組的組蛋白甲基化/乙醯化剖繪(profile)資料；
- (b) 匯集步驟(a)的組蛋白甲基化/乙醯化剖繪資料至一組蛋白甲基化/乙醯化資料庫，該資料庫顯示組蛋白甲基化/乙醯化剖繪的峰數(peak numbers)；
- (c) 分析至少一轉殖基因插入位置的序列；及
- (d) 將步驟(c)的轉殖基因插入位置的序列與步驟(b)的組蛋白甲基化/乙醯化資料庫比對；

其中(i)若是該轉殖基因插入位置落入該甲基化/乙醯化剖繪的峰零(peak zero)或峰一(peak one)，則無轉殖基因靜默的風險，或

(ii)若是該轉殖基因插入位置落入該組蛋白甲基化/乙醯化剖繪的峰二(peak two)或以上，則有顯著的轉殖基因靜默風險。

【0006】於一實施態樣中，該組蛋白甲基化/乙醯化剖繪資料係以一染色質免疫沉澱定序(ChIP-seq)測定法產生。在另一實施態樣中，該組蛋白甲基化/乙醯化剖繪資料係與選自於由下列所組成之群組的組蛋白甲基化/乙醯化有關聯：H3K4me2、H3K4me3、H3K9/14ac、H3K9me2、H3K9me3、H3K27me1、H3K27me3及H4K20me3、及其等之多種組合。在另一實施態樣中，該組蛋白甲基化/乙醯化剖繪資料係與選自於由下列所組成之群組的組蛋白甲基化/乙醯化有關聯：H3K9me2、H3K9me3、H3K27me1、H3K27me3及H4K20me3、及其等之多種組合。在另一實施態樣中，該組蛋白甲基化/乙醯化剖繪資料係與H3K9me2甲基化有關聯。在另一實施態樣中，該植物係選自於大豆、玉米、油菜(canola)、棉花、小麥及稻米。在另一實施態樣中，該轉殖基因插入位置的序列係藉由直接定序(direct sequencing)獲得。

【0007】在另一方面，所提供者為用於預測轉殖基因靜默風險的電腦化系統。該系統包含：

- (a) 一個組蛋白甲基化/乙醯化資料庫，其顯示一植物基因組中的組蛋白甲基化/乙醯化剖繪之峰數；
- (b) 一個輸入模組，其讓一使用者輸入至少一轉殖基因插入位置的序列；及

- (c) 一個輸出模組，其顯示轉殖基因靜默風險之預測，其中該預測係基於比對步驟(b)所輸入的序列與步驟(a)之組蛋白甲基化/乙醯化剖繪。

【0008】於一實施態樣中，該組蛋白甲基化/乙醯化剖繪資料係以一染色質免疫沉澱定序(ChIP-seq)測定法產生。在另一實施態樣中，該組蛋白甲基化/乙醯化剖繪資料係與選自於由下列所組成之群組的組蛋白甲基化/乙醯化有關聯：H3K4me2、H3K4me3、H3K9/14ac、H3K9me2、H3K9me3、H3K27me1、H3K27me3及H4K20me3、及其等之多種組合。在另一實施態樣中，該組蛋白甲基化/乙醯化剖繪資料係與選自於由下列所組成之群組的組蛋白甲基化/乙醯化有關聯：H3K9me2、H3K9me3、H3K27me1、H3K27me3及H4K20me3、及其等之多種組合。在另一實施態樣中，該組蛋白甲基化/乙醯化剖繪資料係與H3K9me2甲基化有關聯。在另一實施態樣中，該植物係選自於大豆、玉米、油菜、棉花、小麥及稻米。在另一實施態樣中，該轉殖基因插入位置的序列係藉由直接定序獲得。

【0009】在另一方面，所提供者為一種供用於一用以預測轉殖基因靜默風險之電腦化系統中的方法。該方法包含：

- (a) 經由一使用者輸入至少一轉殖基因插入位置的序列至文中所提供之該系統；及
- (b) 接收來自文中所提供之該系統就轉殖基因靜默

風險預測的輸出，其中該預測係基於比對所輸入的轉殖基因插入位置的序列與該系統中之組蛋白甲基化/乙醯化剖繪。

【圖式簡單說明】

【0010】圖1顯示出大豆基因特徵(基因本體(gene body)、上游2kb、下游1kb)中為不同表現類別(高、中高、中、中低、低)之函數的H3K9me2豐度。

【實施方式】

【0011】較佳實施例之詳細說明

本發明係基於使用植物基因組中H3K9me2甲基化程度以預測轉殖基因靜默、轉殖基因穩定性、及/或轉殖基因表現程度。所提供者為用於產生全基因組H3K9me2圖譜及其與一指定閾值一起用於預測基因靜默的方法及/或系統。提供於此之該等方法及/或系統可被使用於高通量設置中，用以在與現有技術相比一相對短的時間期間內篩選大量轉型事件。

【0012】轉殖基因性能(performance)及遺傳性轉殖基因穩定性對於性狀產物概念的成功係重要的。數種途徑在過去已被使用以估算轉殖基因表現及性能，其包括量化之表現分析技術，該等技術包括例如RT-PCR、DNA甲基化、及隔代穩定性測試。這些途徑在該性狀發育過程期間耗費時間及資源，且需要多代的基因轉殖事件追蹤，且可能潛在地導致挑選到不好的基因轉殖事件來促進下游產物發展。

【0013】表觀遺傳標記係指DNA及與其有關聯的染色質蛋白質的酵素媒介化學修飾。表觀遺傳標記典型地並不會改變DNA的初級序列，但其等對調控基因組功能係重要的。DNA修飾(包括胞嘧啶甲基化、組蛋白尾及組蛋白核的轉譯後修飾、及核小體的定位)，可能影響染色質的轉錄狀態及其他功能性層面。組蛋白修飾係一涉及轉錄調控機制之細胞系統。活化型(例如H3K4me3及H3K9/14ac)以及壓抑型(例如H3K27me3、H3K9me3、H3K9me2、及H4K20me3)組蛋白標記兩者皆在先前被識別。特別是組蛋白H3中的一個特定離胺酸(K9)之雙甲基化，一般被稱為H3K9me2，已被描繪為一指示出關閉型染色質或基因組層級之轉錄壓抑的表觀遺傳標記。所提供者為方法與系統，其等係用於使用H3K9me2作為一診斷工具以瞭解在一整合的轉殖基因周邊之基因組位置的染色質狀態且預測基因轉殖事件中之基因靜默的潛在風險。所提供之該等方法與系統的適用性包括例如在性狀發育過程中之事件分類(sorting)或選擇。具體地，提供大豆根及枝中的全基因組範圍(genome-wide)之H3K9me2剖繪資料作為一例。在一些實施態樣中，在一轉殖基因插入位置之基因座(locus)處的H3K9me2甲基化程度可被使用以辨識/預測具有較高穩定表現機率之事件。

【0014】在一些實施態樣中，為每個作物植物產生全基因組H3K9me2圖譜。在一些實施態樣中，根據此等全基因組H3K9me2圖譜產生一資料庫。在一些實施態樣

中，此資料庫可根據所提供的方法及/或系統用以預測在我們有興趣之基因座中的染色質結構。

【0015】除非另外指出，所有使用於此的技術及科學用語具有如同其等對於熟習本發明技藝者所會具有的相同意義。針對本技藝中的定義及用語，實施者可特別地參考 Sambrook et al. *Molecular Cloning: A Laboratory Manual* (Second Edition), Cold Spring Harbor Press, Plainview, N.Y., 1989 及 Ausubel FM et al. *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, N.Y., 1993。必須瞭解的是，本發明並不侷限於所敘述之特定方法學、流程(protocol)、及試劑，因其等係可變化的。

【0016】於此所使用的用語「約」係指大於或小於所述數值或數值範圍可達百分之十，而非意圖指定任何數值或數值範圍至僅此較廣定義。各數值或數值範圍前加上用語「約」也意圖包含該所述之絕對數值或數值範圍的實施態樣。

【0017】於此所使用的用語「載體」係指一DNA片段，典型地係雙股的，其可具有一外來DNA片段插入其中。該載體可為例如質體或病毒來源，其典型地編碼一可選擇的或可篩選的標記或轉殖基因。該載體係用於運輸外來或異源DNA至一合適的宿主細胞。一旦在宿主細胞中，該載體可獨立地複製或與該宿主染色體DNA同時複製。任擇地，該載體可靶定外來或異源DNA的插入至宿主染色體中。

【0018】於此所使用的用語「轉殖基因載體」係指包含一插入的DNA片段之載體，「轉殖基因」係被轉錄為mRNA或複製為一RNA於宿主細胞內。用語「轉殖基因」不僅指會轉變成RNA之插入的DNA部分，也指那些為了RNA之轉錄或複製所必須的載體部分。轉殖基因典型地包含一感興趣的基因，但並不必須一定要包含一含有可製造蛋白質的開讀框(open reading frame)之多核苷酸序列。

【0019】於此所使用的用語「轉型的」或「轉型」係指引入DNA至一細胞。用語「轉型體」或「基因轉殖」係指已轉型的或已經歷一轉型程序之植物細胞、植物、及類似物。該被引入之DNA通常是呈包含一插入的DNA片段之載體的形式。

【0020】於此所使用的用語「基因轉殖植物」係指一植物，其之基因組已藉由穩定整合重組DNA而改變。一基因轉殖植物包括從原始轉型(originally-transformed)植物細胞再生之一植物，及從後繼世代或轉型植物雜交而來的子代基因轉殖植物。

【0021】於此所使用的用語「重組DNA」係指經受基因工程且在細胞外建構的DNA，其包括包含自然發生DNA或cDNA或合成DNA的DNA。

【0022】於此所使用的用語「可選擇的標記」或「可選擇的標記基因」係指一基因，其可選地用於植物轉型以，例如，保護該植物之細胞免於一選擇性因子或提供對一選擇性因子的抗性/耐受性。只有那些接受一功能性可選擇

的標記的細胞或植物才有能力在具有一選擇性因子的條件下分裂或生長。選擇性因子之範例可包括，如抗生素，其包括奇黴素 (spectinomycin)、新黴素、康黴素 (kanamycin)、巴龍黴素 (paromomycin)、建它黴素 (gentamicin) 及濕黴素 (hygromycin)。這些可選擇的標記包括新黴素磷酸轉移酶 (npt II) 之基因，其表現一種賦予對抗生素康黴素之抗性的酵素，及相關的抗生素新黴素、巴龍黴素、建它黴素及 G418 之基因，或濕黴素磷酸轉移酶 (hpt) 之基因，其表現一種賦予對濕黴素之抗性的酵素。其他可選擇的標記基因可包括之基因編碼除草劑抗性，包括 Bar (對 BASTA® (草銨膦銨鹽 (glufosinate ammonium)) 或膦蘇菌素 (phosphinothricin, PPT) 之抗性)；乙醯乳酸合成酶 (ALS，對如磺醯尿素 (SUs)、咪唑啉酮 (imidazolinones, IMIs)、三唑并嘧啶 (triazolopyrimidines, TPs)、嘧啶氧代苯甲酸 (pyrimidinyl oxybenzoates, POBs)、及磺醯胺基羰基三唑啉酮 (sulfonamino carbonyl triazolinones) 之防止支鏈胺基酸合成第一步的抑制劑之抗性)；嘉磷塞 (glyphosate)、2,4-D 及金屬抗性或敏感性。用語「標記陽性」係指植物已被轉型以包括該可選擇的標記基因。

【0023】多種可選擇的或可偵測的標記可被併入所選之表現載體以允許辨識及選擇轉型的植物或轉型體。許多方法可被利用以確認選擇標記在轉型植物中的表現，包括如 DNA 定序及 PCR (聚合酶鏈反應)；南方印漬術；RNA

印漬術；免疫方法，用以偵測載體所表現之蛋白質，如媒介朊蘇菌素抗性的沉澱蛋白質、或其他蛋白質諸如報導基因 β -葡萄糖醛酸酶(β -glucuronidase, GUS)、螢光素酶、綠螢光蛋白(GFP)、DsRed、 β -半乳糖苷酶、氯黴素乙醯基轉移酶(chloramphenicol acetyltransferase, CAT)、鹼性磷酸酶及類似物(見 Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Third Edition, Cold Spring Harbor Press, N.Y., 2001, 其內容以其整體併入於此作為參考)。

【0024】可選擇的標記基因被利用以選擇轉型的細胞或組織。可選擇的標記基因包括編碼抗生素抗性的基因，諸如那些編碼新黴素磷酸轉移酶II(NEO)及濕黴素磷酸轉移酶(HPT)者，以及賦予對除草化合物之抗性的基因。除草劑抗性基因一般編碼一對除草劑不敏感之經修飾的目標蛋白質，或一在除草劑可於植物中作用之前降解或解毒該除草劑的酵素。例如，對嘉磷塞之抗性或已藉由使用編碼突變型目標酵素5-烯醇丙酮莽草酸-3-磷酸合成酶(5-Enolpyruvylshikimate-3-phosphate synthase, EPSPS)的基因獲得。EPSPS之基因及突變體已被揭露於美國專利號 4,940,835、5,188,642、5,310,667、5,633,435、5,633,448及6,566,587，其等之內容以其等整體併入作為參考。對草胺磷胺鹽、溴苯腈(bromoxynil)及2,4-二氯苯氧乙酸鹽(2,4-dichlorophenoxyacetate, 2,4-D)的抗性已藉由使用編碼可解毒該等個別除草劑之磷

蘇菌素乙醯基轉移酶、腈酶(nitrilase)或2,4-二氯苯氧乙酸鹽單氧酶的細菌基因獲得。草銨膦抗性/耐受性之酵素/基因已被揭露於美國專利號 5,273,894、5,276,268、5,550,318及5,561,236，其等之內容以其等整體併入作為參考。2,4-D抗性之酵素/基因已在先前被揭露於美國專利號 6,100,446 及 6,153,401，以及專利申請案 US 2009/0093366 (AAD-1) 及 WO 2007/053482 (AAD-12)，其等之內容以其等整體併入於此作為參考。腈酶之酵素/基因已在先前被揭露於美國專利號 4,810,648，其內容以其整體併入作為參考。

【0025】其他除草劑會抑制生長點或分生組織，包括咪唑啉酮或磺醯尿素，而針對該等除草劑之乙醯羧基酸合成酶(acetohydroxyacid synthase, AHAS)及乙醯乳酸合成酶(ALS)之抗性/耐受性的基因已被描述。AHAS及突變體之基因及突變體已被揭露於美國專利號4,761,373、5,304,732、5,331,107、5,853,973及5,928,937，其等之內容以其等整體併入作為參考。ALS之基因及突變體已被揭露於美國專利號5,013,659及5,141,870，其等之內容以其等整體併入作為參考。

【0026】嘉磷塞抗性基因分別包括突變型5-烯醇丙酮莽草酸-3-磷酸合成酶(EPSPs)基因(經由引入重組核酸及/或天然EPSPs基因之多種形式的活體內誘變)、aroA基因、及嘉磷塞乙醯基轉移酶(GAT)基因。用於其他膦醯基化合物的抗性基因包括草銨膦(來自包括吸水鏈絲菌(*Streptomyces*

hygroscopicus) 及綠色產色鏈絲菌 (*Streptomyces viridichromogenes*) 之鏈絲菌屬物種的膾蘇菌素乙醯基轉移酶(PAT)基因)、及吡啶氧或苯氧基丙酸(pyridinoxy or phenoxy proprionic acids)及環己烯酮(cyclohexones)(編碼 ACCase抑制物之基因)。乙醯輔酶A羧酶(ACCase)之除草劑抗性/耐受性基因已被描述於美國專利 5,162,602 及 5,498,544，其等之內容以其等整體併入作為參考。

【0027】編碼一突變型aroA基因的一DNA分子可在 ATCC登錄號39256下獲得，且該突變型基因的核苷酸序列被揭露於Comai之美國專利號4,769,061、Kumada等人之歐洲專利申請案號0 333 033、及Goodman等人之美國專利號4,975,374，其等揭露賦予對諸如L-膾蘇菌素之除草劑之抗性的麩胺合成酶基因的核苷酸序列。PAT基因的核苷酸序列係被提供於Leemans等人之歐洲申請案號 0 242 246。且 DeGreef 等人於 Bio/Technology 7:61 (1989)描述表現編碼PAT活性之嵌合bar基因之基因轉殖植物的生產。賦予對包括西殺草(Sethoxydim)及氟吡禾靈(haloxyfop)之苯氧基丙酸及環己烯酮之抗性的範例基因為 Acc1-S1、Acc1-S2 及 Acc1-S3 基因，其等被 Marshall 等人描述於 Theon. Appl. Genet. 83:435 (1992)。可賦予嘉磷塞抗性的GAT基因被描述於 Castle 等人的WO 2005012515。賦予對2,4-D、fop及吡啶氧基生長素(pyridyloxy auxin)除草劑之抗性的基因被描述於 WO 2005107437及美國專利申請案序號11/587,893。

【0028】其他除草劑會抑制光合作用，包括三嗪(triazine)(psbA及1s+基因)或苄腈(腈酶基因)。Przibila等人在 Plant Cell 3:169 (1991) 描述單胞藻屬(*Chlamydomonas*)以編碼突變型psbA基因的質體轉型。腈酶基因的核苷酸序列揭露於Stalker的美國專利號4,810,648，而含有這些基因的DNA分子可在ATCC登錄號53435、67441及67442下獲得。編碼一羰胺基硫S-轉移酶的DNA之選殖及表現被Hayes等人描述於 Biochem. J. 285:173 (1992)。

【0029】為了本發明的目的，可選擇的標記基因包括但並不侷限於編碼以下之基因：新黴素磷酸轉移酶II (Fraley *et al.* (1986) CRC Critical Reviews in Plant Science, 4:1-25)；氰胺水合酶(Maier-Greiner *et al.* (1991) Proc. Natl. Acad. Sci. USA, 88:4250-4264)；天冬胺酸激酶；二氫吡啶二羧酸合成酶(dihydrodipicolinate synthase)(Perl *et al.* (1993) Bio/Technology, 11:715-718)；色胺酸去羧酶(Goddijn *et al.* (1993) Plant Mol. Bio., 22:907-912)；二氫吡啶二羧酸合成酶及去敏的(desensitized)天門冬胺酸激酶 (Perl *et al.* (1993) Bio/Technology, 11:715-718)；bar基因(Toki *et al.* (1992) Plant Physiol., 100:1503-1507及Meagher *et al.* (1996) and Crop Sci., 36:1367)；色胺酸去羧酶(Goddijn *et al.* (1993) Plant Mol. Biol., 22:907-912)；新黴素磷酸轉移酶(NEO)(Southern *et al.* (1982) J. Mol. Appl. Gen.,

1:327)；濕黴素磷酸轉移酶(HPT或HYG)(Shimizu *et al.* (1986) *Mol. Cell Biol.*, 6:1074)；二氫葉酸還原酶(DHFR)(Kwok *et al.* (1986) *PNAS USA* 4552)；磷蘇菌素乙醯基轉移酶(DeBlock *et al.* (1987) *EMBO J.*, 6:2513)；2,2-二氯丙酸去鹵酶(Buchanan-Wollatton *et al.* (1989) *J. Cell. Biochem.* 13D:330)；乙醯羧酸合成酶(Anderson等入之美國專利號4,761,373；Haughn *et al.* (1988) *Mol. Gen. Genet.* 221:266)；5-烯醇丙酮-莽草酸-磷酸合成酶(*aroA*) (Comai *et al.* (1985) *Nature* 317:741)；鹵代芳基腈酶(haloarylnitrilase) (Stalker等人的已公開之PCT申請案WO87/04181)；乙醯輔酶A羧酶(Parker *et al.* (1990) *Plant Physiol.* 92:1220)；二氫蝶酸合成酶(dihydropteroate synthase) (*sul I*) (Guerineau *et al.* (1990) *Plant Mol. Biol.* 15:127)；及32 kD光系統II多肽(*psbA*) (Hirschberg *et al.* (1983) *Science*, 222:1346)。

【0030】也被包括者為編碼對以下之抗性的基因：氯黴素 (Herrera-Estrella *et al.* (1983) *EMBO J.*, 2:987-992)；胺甲喋呤(Herrera-Estrella *et al.* (1983) *Nature*, 303:209-213；Meijer *et al.* (1991) *Plant Mol Bio.*, 16:807-820 (1991)；濕黴素 (Waldron *et al.* (1985) *Plant Mol. Biol.*, 5:103-108；Zhijian *et al.* (1995) *Plant Science*, 108:219-227 及 Meijer *et al.* (1991) *Plant Mol. Bio.* 16:807-820)；鏈黴素(Jones *et al.* (1987) *Mol. Gen. Genet.*, 210:86-91)；奇黴素(Bretagne-Sagnard *et al.*

(1996) *Transgenic Res.*, 5:131-137)；博萊黴素(Hille *et al.* (1986) *Plant Mol. Biol.*, 7:171-176)；磺醯胺(Guerineau *et al.* (1990) *Plant Mol. Bio.*, 15:127-136)；溴苯腈(Stalker *et al.* (1988) *Science*, 242:419-423)；2,4-D (Streber *et al.* (1989) *Bio/Technology*, 7:811-816)；嘉磷塞 (Shaw *et al.* (1986) *Science*, 233:478-481)；及膦蘇菌素 (DeBlock *et al.* (1987) *EMBO J.*, 6:2513-2518)。本揭露中所引述的所有參考文獻以其等整體併入於此作為參考，除非另有聲明。

【0031】 以上所列之可選擇的標記及報導基因並不意欲作為限制。任何報導物或可選擇的標記基因係被此發明所包含。若有必要，如此的基因可被此技藝中已知的方法定序。

【0032】 該等報導物及可選擇的標記基因被合成以供在植物中有最佳化的表現。亦即，該基因的編碼序列已被修飾以增進在植物中的表現。該合成的標記基因被設計以在植物中以一較高的程度被表現導致較高的轉型效率。用於基因之合成最佳化的方法在此技藝中是可得的。事實上，數個基因已被最佳化以增加植物中基因產物的表現。

【0033】 該標記基因的序列可被最佳化以用於表現於特定的植物物種，或任擇地可被修飾以用於在植物家族中有最佳化的表現。植物所偏好的密碼子可決定自在有興趣的特定植物物種之最大量表現的蛋白質中有最高頻率的密碼子。參見如EPA 0359472、EPA 0385962、WO

91/16432、Perlak *et al.* (1991) *Proc. Natl. Acad. Sci. USA*, 88:3324-3328、以及Murray *et al.* (1989) *Nucleic Acids Research*, 17: 477-498、美國專利號5,380,831、以及美國專利號5,436,391，其等併入於此作為參考。以此方式，該等核苷酸序列可被最佳化以用於在任何植物中表現。已公認該基因序列的全部或任何部分可被最佳化或合成。亦即，完全最佳化或部分最佳化之序列也可被利用。

【0034】再者，數個利用農桿菌媒介轉型系統的轉型策略已被發展。例如，二元載體(binary vector)策略係基於一個二質體系統，其中T-DNA所在之質體不同於Ti質體的其餘者。在一個共整合(co-integration)策略中，該T-DNA之一小部分如同外來基因被置於相同的載體中，該載體接著與該Ti質體重組。

【0035】於此所使用的用語「植物」包括雙子葉植物及單子葉植物。雙子葉植物的範例包括菸草、阿拉伯芥、大豆、番茄、木瓜、油菜、向日葵、棉花、苜蓿、馬鈴薯、葡萄藤、樹豆、豌豆、薔薹屬、雞豆、甜菜、油菜子、西瓜、甜瓜、胡椒、花生、南瓜、蘿蔔、菠菜、西葫蘆、青花菜、甘藍、紅蘿蔔、花椰菜、芹菜、大白菜、小黃瓜、茄子、及萵苣。單子葉植物的範例包括玉蜀黍、稻米、小麥、甘蔗、大麥、裸麥、高粱、蘭花、竹子、香蕉、香蒲、百合、燕麥、洋蔥、小米、及黑小麥。

【0036】於此所使用的用語「植物」也包括一完整植物及一植物之任何後代、細胞、組織、或部分。用語「植

物部分」包括一植物的任何部分，包括，例如而無限制：種子(包括成熟種子及不成熟種子)、一植物插條(cutting)、一植物細胞、一植物細胞培養、一植物器官(例如花粉、胚、花、果實、枝、葉、根、莖、及外植體)。一植物組織或植物器官可為一種子、癒傷組織、或任何其他被組織為一結構或功能單元的植物細胞群組。一植物細胞或組織培養可有能力再生具有該細胞或組織從而獲得的植物之生理及形態特徵的一植物，且可有能力再生具有與該植物實質相同基因型的一植物。相對地，一些植物細胞並沒有能力再生以產生植物。在一植物細胞或組織培養中之可再生的細胞可為胚、原生質體、分生細胞、癒傷組織、花粉、葉、花藥、根、根尖、鬚絲(silk)、花、種仁、穗、穗軸、莢、或柄。

【0037】植物部分包括可收穫的部分及利於繁殖子代植物的部分。利於繁殖的植物部分包括，例如而無限制：種子、果實、一插條、一苗、一塊莖、及一砧木。一可收穫的植物部分可為一植物任何有用的部分，包括，例如而無限制：花、花粉、苗、塊莖、葉、莖、果實、種子、及根。

【0038】於此所使用的用語「所描述的植物細胞」或「轉型的植物細胞」係指一植物細胞，其經以穩定整合的、非自然的重組DNA轉型，例如藉由農桿菌媒介轉型或藉由使用被重組DNA包覆的微粒衝擊(bombardment)或其他方法。本發明的一植物細胞可為一原始轉型的植物細胞，

其存在為一微生物或為一子代植物細胞而可再生為分化的組織，例如可再生為一具有穩定整合的、非自然的重組DNA之基因轉殖植物，或從一子代基因轉殖植物取得的種子或花粉。

【0039】於此所使用的用語「共通(consensus)序列」係指同源蛋白質之胺基酸序列比對(alignment)的一保留區中的胺基酸人工序列，例如，如藉由CLUSTALW比對(功能性)同源蛋白質的胺基酸序列而決定者。

【0040】於此所使用的用語「同源物」係指一核酸或在一展現相同生物功能之蛋白質群組中的一蛋白質，例如，屬於相同蛋白質家族的蛋白質或於本發明基因轉殖植物中提供一共同增益性狀的相似核酸。同源物係由同源基因表現。參照同源基因，同源物包括異種同源物，即在不同物種中表現的基因，其等係藉由種化從共同祖先基因演化而來且編碼維持相同功能的蛋白質；但並不包括同種同源物，即因複製而相關但已演化成編碼不同功能蛋白質的基因。同源基因包括自然發生的對偶基因及人工創造的變異體。遺傳密碼的簡併(Degeneracy)提供了可能性來以一不同鹼基取代一基因的蛋白質編碼序列的至少一鹼基而不會造成該基因所製造的多肽之胺基酸序列被改變。當最佳地比對，同源基因相較於被辨識為與在植物細胞中表現時給予增益性狀有關聯的基因之全長，具有至少50%、60%、70%、80%、85%、90%、95%、96%、97%、98%或99%或更多的序列同一性。在本發明的一方面，同

源基因具有一核酸或胺基酸序列相似性，其與揭露於此的蛋白質、核苷酸及同源物的一共通序列具有至少85%、90%、95%、96%、97%、98%或99%同一性。

【0041】同源物可藉由比較胺基酸序列被辨識，例如，手動地或藉由使用一運用已知的同源基礎搜尋演算法的電腦化工具，諸如那些通常已知且稱為BLAST、FASTA、及Smith-Waterman者。一局部序列比對程式(例如：基本局部比對搜尋工具(BLAST))可以使用於搜尋一序列資料庫以尋找相似序列、及用以量測序列基礎相似性的總期望值(E-值)。因為一與一特定生物碰巧有最佳E-值的蛋白質並不必然為一異種同源物，即具有相同功能，或為唯一的異種同源物，為了辨識異種同源物，一交互查詢被用於過濾具顯著E-值的碰巧序列。該交互查詢需要對照與所查詢的蛋白質之序列相似的基礎生物之胺基酸序列資料庫搜尋該等顯著的碰巧物。當該交互查詢的最佳位元是該查詢蛋白質本身或是在種化後由一複製基因所編碼的蛋白質時，碰巧物可被辨識為一異種同源物。一進一步方面，對本發明基因轉殖植物有效的DNA所編碼之同源物係為因在一天然序列中缺失或插入一或多個胺基酸而異於一已揭露蛋白質的那些蛋白質。

【0042】於此所使用的用語「百分比同一性」或「%同一性」係指DNA或蛋白質片段序列在如核苷酸序列或胺基酸序列之序列比對視窗中從頭到尾不變的程度。就一序列與一參考序列比對的「同一性分率」係為該等序列共有

的同一組分之數目除以該比對的長度，該長度不包括比對演算法所引入的間隙。「百分比同一性」為該同一性分率乘以100。百分比同一性係相較於比對長度計算，較佳地使用如BLASTp的局部比對演算法。

【0043】於此所使用的用語「功能活性」或「功能上活性」係指供根據本發明使用之蛋白質/酵素，其等具有提供壓力耐受性的能力而可以導致產量增加。轉移該功能活性至植物或細菌系統可涉及一核酸序列，其編碼本發明蛋白質之胺基酸序列、整合至適合於宿主的一蛋白質表現載體中，該載體會存在於該宿主中。一種用以獲得編碼一具有功能活性的蛋白質之核酸序列的途徑為，利用從如在此所揭露的蛋白質胺基酸序列推斷而來的資訊，從製造感興趣蛋白質的細菌物種中分離天然遺傳物質。該天然序列可被最佳化以在植物中表現。最佳化的多核苷酸也可基於蛋白質序列而被設計。

【0044】於此所使用的用語「控制序列」及「調控序列」係可交換的且係指在植物中用於轉錄/基因表現的核酸序列。「控制序列」或「調控序列」可包括但不侷限於啟動子、操縱子、強化子、複製起始點、核糖體結合位置、終止及多腺核苷酸化訊號。

【0045】於此所使用的用語「啟動子」係指用於起始轉錄的調控DNA。一「植物啟動子」係一在植物細胞內可起始轉錄的啟動子，不論其起源是否為植物細胞，例如，已熟知農桿菌T-DNA啟動子在植物細胞中具功能性。因

此，植物啟動子包括從植物、植物病毒及如農桿菌及慢生根瘤菌 (*Bradyrhizobium*) 細菌之細菌所獲得的啟動子 DNA。在發育控制之下的啟動子範例包括在包括葉、根、或種子的特定組織中優先起始轉錄的啟動子。如此的啟動子係稱為「組織偏好的」。只在特定組織中起始轉錄的啟動子係稱為「組織特定的」。一「細胞類型」特定啟動子主要驅動在一或多個器官之特定細胞類型中的表現，例如在根或葉中的維管束細胞。一「可誘導的」或「可壓抑的」啟動子係一在環境或化學控制下的啟動子。可藉由可誘導的啟動子影響轉錄之環境條件範例包括缺氧條件、或某些化學品、或光的存在。組織特定的、組織偏好的、細胞類型特定的、及可誘導的啟動子屬於「非持續型 (non-constitutive)」啟動子類別。一「持續型」啟動子係指在大多數條件下及在大多數組織中有活性的啟動子。

【0046】在植物細胞中有活性的多數啟動子在先前已被描述。這些包括存在於植物基因組中的啟動子以及從其他來源而來的啟動子，包括如在美國專利號 5,164,316 及 5,322,938 所揭露的在根癌農桿菌 (*Agrobacterium tumefaciens*) 的瘤誘導質體上所帶有的胭脂鹼 (nopaline) 合成酶 (NOS) 啟動子及章魚鹼 (octopine) 合成酶 (OCS) 啟動子，及由花椰菜嵌紋病毒 (cauliflower mosaic virus) 而來的 CaMV35S 啟動子。由植物基因得到的有用啟動子係發現於美國專利號 5,641,876，其揭露一種稻米肌動蛋白啟動子；美國專利號 7,151,204，其揭露一種玉米葉綠體

醛縮酶啟動子及一種玉米醛縮酶(FDA)啟動子；及美國專利申請公開案2003/0131377，其揭露一種玉米烟草胺(nicotianamine)合成酶啟動子。在植物細胞中作用的這些及數種其他啟動子已被熟習此技藝者知曉，且可得用於此處所描述之重組多核苷酸來提供基因轉殖植物細胞中所欲基因的表現。

【0047】再者，該等啟動子可被改變以包含多個「強化子序列」用來協助提高基因表現。這樣的強化子在本技藝中為已知。藉由以如此建構體(construct)包括一強化子序列，所選擇蛋白質之表現可被強化。這些強化子通常被發現於作用於真核細胞之啟動子中的轉錄起始之5'，但經常可以被插入於編碼序列的上游(5')或下游(3')。在某些情況下，這些5'強化元件為內含子(intron)。特別有效作為強化子者為稻米肌動蛋白1(見例如美國專利號5,641,876)及稻米肌動蛋白2基因的5'內含子、玉米醇脫氫酶基因內含子、玉米熱休克蛋白70基因內含子(見例如美國專利號5,593,874)及玉米皺縮(shrunken)1基因內含子。亦見美國專利申請公開案2002/0192813A1，其揭露有利於設計有效植物表現載體的5'、3'及內含子元件。

【0048】在一些實施態樣中，在植物種子組織中充足的表現係所欲來影響種子組成的改良。用於種子組成修飾的範例啟動子包括從諸如下列之種子基因而來的啟動子：如揭露於美國專利號5,420,034的奈平(napin)、如揭露於美國專利號6,433,252的玉米L3油體蛋白(maize L3

oleosin)、如由Russell *et al.* (1997) *Transgenic Res.* 6(2): 157-166 揭露的玉米蛋白 Z27(zein Z27)、如由Belanger *et al* (1991) *Genetics* 129:863-872揭露的球蛋白1(globulin 1)、如由Russell (1997, 上文)揭露的穀蛋白1(glutelin 1)、及如由Stacy *et al.* (1996) *Plant Mol Biol.* 31(6):1205-1216揭露的過氧化物氧化還原酶抗氧化劑(peroxiredoxin antioxidant, Perl)。

【0049】於此所使用的用語「可操作地連接」係指在一個重組DNA建構體中兩個或更多個DNA片段的關聯，以致其中一個如蛋白質編碼DNA的功能被其他如啟動子所控制。

【0050】重組DNA建構體係以本技藝中具有通常知識者所熟知的方法組合，且典型地包含一個可操作地連接至DNA的啟動子，該DNA的表現提供強化農藝性狀。其他建構體組分可包括額外的調控元件，其包含5'前導物(leader)及用於強化轉錄的內含子、3'非轉譯區(例如，多腺核苷酸化訊號及位置)、及用於轉運(transit)或訊號肽的DNA。

【0051】在此描述的重組DNA建構體一般也包括一個典型包含多腺核苷酸化訊號及位置的3'元件。熟知的3'元件包括那些從根癌農桿菌基因而來的如nos 3'、tml 3'、tmr 3'、tms 3'、ocs 3'、tr73'，例如揭露於美國專利號 6,090,627 者；從如小麥(普通小麥(*Triticum aestivum*))熱休克蛋白17(Hsp1 73)、小麥泛素基因、小

麥果糖-1，6-二磷酸酶基因、稻米穀蛋白基因、稻米乳酸脫氫酶基因及稻米 β 微管蛋白基因之植物基因而來的3'元件，所有該等3'元件揭露於美國專利申請公開案2002/0192813；以及豌豆(*Pisum sativum*)核酮糖二磷酸羧酶基因(rbs 3')及來自宿主植物內部基因的3'元件。

【0052】建構體及載體也可包括一個用於靶定一基因至植物胞器的轉運胜肽，特別是至葉綠體、白色體或其他色素體胞器。葉綠體轉運胜肽的使用描述可見如美國專利號5,188,642及美國專利號5,728,925。

【0053】於此所使用的用語「表現」係指一蛋白質在植物細胞中被表現或製造於當其同族的(cognate)DNA被轉錄為mRNA而被轉譯為該蛋白質。於此所使用的用語「壓制的」係指一蛋白質的表現或活性下降。典型地，一蛋白質在植物細胞中被壓制於當該蛋白質的數量及/或活性在該植物細胞中下降。該蛋白質的存在或活性可被降低達上至且包括蛋白質表現及/或活性全然喪失的任一數量。

【0054】於此所使用的用語「性狀」係指植物、或特定的植物材料或細胞的生理的、形態的、生化的、或物理的特徵。在某些情況下，此特徵對人眼係可見的，包括種子或植物大小，或可藉由生化技術而被量測，包括偵測種子或葉的蛋白質、澱粉、或油含量，或藉由觀察一個代謝或生理的過程(如藉由量測二氧化碳的攝取)，或藉由觀察一基因或多個基因的表現程度(如藉由利用北方(Northern)分析、RT-PCR、微陣列基因表現測定法)，

或報導基因表現系統，或藉由包括壓力耐受性、產量、或病原體耐受性的農業觀察。

【0055】於此所使用的用語「過表現」係指一基因在對於該基因之任何發育或時間階段，相較於在野生型(wild-type)植物、細胞或組織中的表現，在一植物、植物細胞或植物組織中有較大的表現程度。過表現可發生於例如當該等基因在一強大表現訊號的控制下，該訊號包括描述於此的啟動子之一或此技藝中所知的花椰菜嵌紋病毒35S轉錄起始區。過表現可發生於植物的各處或在植物的特定組織中，其係取決於所使用的啟動子。

【0056】於此所使用的「非基因序列」或「非基因基因組序列」係在一植物的核基因組中發現的一天然DNA序列，其具有至少1 Kb的長度，且缺乏任何開讀框、基因序列、或基因調控序列。再者，該非基因序列並不包含任何內含子序列(即內含子從非基因的定義中排除)。該非基因序列不能被轉錄或轉譯為蛋白質。許多植物基因組包含非基因區域，其中最多如95%的基因組可為非基因的，且這些區域可包括主要重複性DNA(repetitive DNA)。

【0057】於此所使用的「基因區域」係定義為包含一個編碼RNA及/或多肽之開讀框的多核苷酸序列。該基因區域也可包含任何牽涉該開讀框之表現調控的可辨識鄰接5'及3'非編碼核苷酸序列上至約編碼區上游2 Kb以及編碼區下游1 Kb，但也可能更上游或下游。一基因區域更包括任何可能存在於該基因區域內的內含子。再者，基因區域

可包含一單一基因序列、或分插有短跨距(short span)(少於1 Kb)非基因序列的多基因序列。

【0058】於此所使用的用語「低甲基化(hypomethylation)」或「低甲基化的(hypomethylated)」參照一DNA序列定義在一特定DNA序列中甲基化DNA核苷酸殘基之降低狀態。典型地，該降低的甲基化係有關於甲基化的腺嘌呤或胞嘧啶殘基數量，相對於存在於如大豆或玉米之植物的基因組中的非基因序列中所發現的平均甲基化程度。

【0059】用語「在基因區域近側位置」當參照一非基因序列使用係定義該非基因序列與基因區域的相對位置。特別地，在一40 Kb鄰域(即所選擇的最佳化大豆基因組基因座序列的任一端40 Kb以內)中的基因區域數量被分析。在該40 Kb鄰域以內之基因區域數量範圍可從最少1個基因至最多18個基因。

【0060】於此所使用的用語「重組的證據」係有關於在任一對橫跨一個包含所選擇的序列之染色體區域的基因組標記之間減數分裂的重組頻率。該等重組頻率係基於標記之間的基因距離(以分摩根(cM)為單位)與標記之間的物理距離(以兆鹼基(Mb)為單位)的比率而被計算。為了使一個選擇的序列具有重組的證據，該選擇的序列當使用從多定位群體(multiple mapping populations)所產生的高解析度標記資料集來檢測時，必須包含至少一個介於兩個側翼(flanking)於該選擇序列之標記之間的重組事件。

【0061】於此所使用的用語「相對位置數值」係一經計算的數值，其定義一基因組基因座與其對應染色體中節的距離。針對每一個選擇的序列，從該選擇序列的自然位置至該選擇序列所在的染色體之中節的基因組距離被量測(以Bp為單位)。該選擇序列在該染色體中的相對位置係被以其至中節的基因組距離相對於其所位於之特定染色體臂長度(以Bp為單位量測)的比率來呈現。這些用於最佳的非基因的基因組基因座之相對位置數值可為了不同植物而產生，用於資料集的相對位置數值的範圍可從最小0至最大0.99682基因組距離比率。

【0062】於此所使用的用語「外源DNA序列」係任何核酸序列，其被從其天然位置移除並插入一新的位置而改變了側翼於該被移除之核酸序列的序列。例如，一個外源DNA序列可包含一從另一物種而來的序列。

【0063】「重組」係指兩個多核苷酸之間遺傳資訊的交換過程，其包括但不侷限於藉由非同源端連接(NHEJ)的供體捕獲(donor capture)及同源重組。於此所使用的「同源重組(HR)」係指如此之交換的特殊形式，其例如在細胞中雙股斷裂修復期間通過同源介導(homology-directed)的修復機制發生。這個過程需要核苷酸序列同源性，使用一「供體」分子來模板修復一「標的」分子(即經歷雙股斷裂的核苷酸序列)，且其有各種稱謂如「非交換基因轉換(non-crossover gene conversion)」或「短束基因轉換(short tract gene conversion)」，因其

導致遺傳資訊從供體轉移至標的。不希望受限於任何特定的理論，如此的轉移可涉及在斷裂的標的與供體之間形成之異源雙股DNA的誤配(mismatch)校正，及/或「合成依賴之股黏著(annealing)」，其中供體係用來重新合成將成為標的的一部分之遺傳資訊，及/或相關過程。這樣特殊化的HR通常導致標的分子序列的改變，使得供體多核苷酸序列的部分或全部被併入標的多核苷酸中。對於HR介導的整合，供體分子包含有至少2個與基因組同源的區域(同源臂)，其長度至少為50-100個鹼基對。見例如美國專利公開案號20110281361，其內容併入於此作為參考。

【0064】於此所使用的用語「基因表現」係指包含在一基因的資訊轉換成一基因產物。一基因產物可為一基因的直接轉錄產物(如mRNA、tRNA、rRNA、反義RNA、干擾RNA、核糖核酸酵素、結構RNA、或任何其他類型的RNA)或藉由mRNA轉譯所產生的蛋白質。基因產物也包括經修飾的RNA等，該修飾是藉由諸如加帽(capping)、多腺核苷酸化、甲基化、及編輯(editing)之過程，以及藉由例如甲基化、乙醯化、磷酸化、泛蛋白化、ADP核糖基化、肉豆蔻酸化(myristilation)、及醯化修飾之蛋白質。

【0065】於此所使用的用語「實質同源的」或「實質同源」，有關於一鄰接核酸序列，其係指鄰接核苷酸序列可在嚴苛條件下雜交至參考核酸序列。例如，實質同源於一參考核酸序列的核酸序列係那些在嚴苛條件下(例如所提中度嚴苛條件等，上文)雜交至該參考核酸序列的核酸序

列。實質同源的序列可具有至少80%序列同一性。例如，實質同源的序列可具有從約80%至100%序列同一性，諸如約81%、約82%、約83%、約84%、約85%、約86%、約87%、約88%、約89%、約90%、約91%、約92%、約93%、約94%、約95%、約96%、約97%、約98%、約98.5%、約99%、約99.5%、及約100%。實質同源的特性與特定的雜交密切相關。例如，當在特定結合係所欲的條件下，例如在嚴苛雜交條件下，有足夠程度的互補性以防止核酸非特定地結合至非標的序列時，一核酸分子可特定地雜交。

【0066】在某些情況，「同源的」可被用以指一第一基因與一第二基因藉由從一共同祖先DNA序列承襲而來的相關性。在這樣的情況下，用語「同源」係指藉由種化事件分離之基因間的相關性(見異種同源物)或藉由基因複製事件分離之基因間的相關性(見同種同源物)。在其他情況下，「同源的」可被用以指一或多個多核苷酸序列之間的序列同一性程度，在這樣的情況下，該一或多個多核苷酸序列並不必須承襲自一共同祖先DNA序列。那些本技藝中具有技術者知曉用語「同源的」之可互換性且領會該用語的適當運用。

【0067】雖本發明已參照特定方法與實施態樣來描述，將瞭解可做出各種修改及改變而不背離本發明。於此所記載之所有公開文件被特意地併入於此作為參考，以達描述及揭露可能被用以結合本發明的組成物及方法學的目的。所有記載的專利、專利申請案、及於參考網頁的序列資訊、及公開的資料庫也作為參考併入。

實例

實例1

【0068】根據在此所提供的方法及/或系統，資料庫及/或圖譜資料可用於藉由剖繪在轉殖基因插入基因座附近的局部染色質狀態，特別是H3K9me2的豐度(abundance)，來估算或預測轉殖基因性能及/或穩定性。如此處所提供，一閾值可被指定以就發生有高度富集H3K9me2之基因組區域的基因轉殖整合預測轉殖基因靜默的潛在風險。為了估算該局部染色質結構，一個大豆(*Glycine max*) cv Maverick的全基因組範圍組蛋白3離胺酸(K)9二甲基化(H3K9me2)圖譜被建構。H3K9me2經常在包括中節及攜帶端粒之重複性序列的靜默異染色質區域被觀察到。此全基因組範圍H3K9me2圖譜允許我們在大豆基因組內的任何基因座預測染色質結構(打開的或緊縮的)並提供對潛在靜默的估算。

【0069】大豆(*Glycine max*) cv Maverick的枝及根組織在v1(約在栽植後12天)期被收穫，接著使用1%甲醛交聯(cross-link)。該交聯的組織被用於使用H3K9me2抗體(Millipore 17-648)的染色質免疫沉澱(ChIP)測定法，其中兔子的正常IgG(Millipore Cat. 12-370)被使用作為一負控制。三個獨立的從各組織而來的ChIP測定法被執行且該經ChIP的DNA被準備以使用HiSeq來定序。定序讀數(read)係列於表1。

表 1.來自 HiSeq 的 ChIP 定序讀數						
樣本 名稱	標示	物種	條碼	總讀數 #	定位 讀數#	定位 比率
x486_7	枝-1 (1)	大豆 (Gmax_109)	索引 5	76,305,548	57,296,302	75.09 %
x486_8	枝-2 (2)	大豆 (Gmax_109)	索引 6	98,536,615	78,863,536	80.03 %
x486_9	枝-2 (3)	大豆 (Gmax_109)	索引12	90,263,804	74,632,245	82.68 %
x486_10	根-3 (1)	大豆 (Gmax_109)	索引 4	91,350,018	68,443,285	74.92 %
x486_11	根-3 (2)	大豆 (Gmax_109)	索引12	99,169,687	74,526,487	75.15 %
x486_12	根-4 (3)	大豆 (Gmax_109)	索引 6	92,090,396	69,694,991	75.68 %

【0070】ChIP 定序 (ChIP-Seq) 資 料 分 析 - 來 自 H3K9me2 ChIP 定序實驗的未處理定序資料係使用 Bowtie 對 Glycine Max Williams82 基 因 組 比 對 。 MACS(ChIP 定序之模型基底分析 (model-based analysis))之峰尋找演算法可被用於辨識背景之上的 H3K9me2富集區域，其中一個「峰」代表一個基因組延伸 (genomic stretch)，該基因組延伸被預測為相較於其基因組鄰域為H3K9me2富集。另一方面，一個「峰頂」代表在預測的「峰」中具有最大富集的基因組位置。

【0071】根據此實例，ChIP定序資料在技術性重複中係一致的。為了減少該ChIP定序資料所導致的變異及雜訊，三次ChIP定序的技術性重複被各自執行。為了檢查各回合之ChIP拉下物(pull-downs)是否一致，並反映在大豆基因組中真實的H3K9me2修飾，來自各技術性重複的拉下

物可被比較。在該三次技術性重複中，該ChIP定序資料係互相一致，代表來自各回合ChIP實驗之拉下物係一致的。因此，使用於此實例的ChIP定序流程(protocol)在跨根與枝之技術性重複樣本為高度可重複的。

實例2

【0072】 H3K9me2富集係與異染色質區域及基因靜默相關聯 - H3K9me2係與異染色質(緊密地緊縮的染色質結構)及一個會導致真核生物轉錄壓抑的高度保守之組蛋白標記相關聯。為了決定H3K9me2是否在大豆中與異染色質區域及基因靜默相關聯，H3K9me2在基因區域中的富集(以在一基因區域內所預測的H3K9me2「峰」之數目量測-較高的峰數代表較大程度的H3K9H3K9me2富集)被調查，並且與相對應的基因表現程度相比(使用RNAseq技術量測，以標準化讀數計算為單位)。在此，我們觀察到具零個或一個H3K9me2峰之基因區域展現出顯著的基因表現範圍。另一方面，具多於兩個H3K9me2峰之基因區域與相對較低的基因表現程度相關聯。此指出在大豆中H3K9me2的較高富集係與較低的基因表現程度高度相關聯。

【0073】 除了瞭解基因區域中的H3K9me2含量，我們還進一步辨識轉位元(transposable element)中較高程度的H3K9me2，其經常與基因組中高度重複且轉錄壓抑的區域相關聯。我們比較大豆基因組中轉位元與基因間的H3K9me2富集分佈。平均來說，大部分轉位元的H3K9me2富集比起普通基因高得多。此資料進一步實證了H3K9me2

富集與大豆基因組的壓抑狀態高度相關聯。

【0074】為了建立我們對H3K9me2豐度如何與基因表現相關聯的瞭解，我們基於基因的表現程度將該等基因分類為五種不同表現類別-高、中高、中、中低、及低-，且接著從各個類別中選擇500個基因的代表數目(見圖1)。針對每個類別中的每個基因，遍及基因區域的H3K9me2豐度被計算。該基因區域接著被分為三區段 - 基因本體、上游2kb、及下游1kb。每區段中H3K9me2豐度的平均數值係針對各表現類別繪製於圖1。高及中高表現類別可基於在所有三個基因區段中的H3K9me2豐度與低及中低表現類別區分。此對於如何可以使用H3K9me2資料來預測一特定感興趣之轉殖基因的表現程度提供了一個獨特的觀點及統計模型。總結來說，以上資料指出H3K9me2也是一個大豆中保守的壓抑型組蛋白標記，且H3K9me2的豐度與大豆中異染色質區域及轉錄壓抑相關聯。我們推斷出一個連結到可接受之表現程度的閾值，其中該基因區域具有一個峰在定位至大豆基因組的H3K9me2剖繪資料。根據此實例，一個用於大豆基因組的資料庫基於H3K9me2剖繪資料被建立。

實例 3

【0075】使用大豆領先事件(lead event)來估算H3K9me2富集，以驗證基於H3K9me2剖繪資料之表現程度的預測。為了評估施用H3K9me2圖譜作為一個估算轉殖基因靜默潛在風險的工具之實際使用，九個來自Enlist及

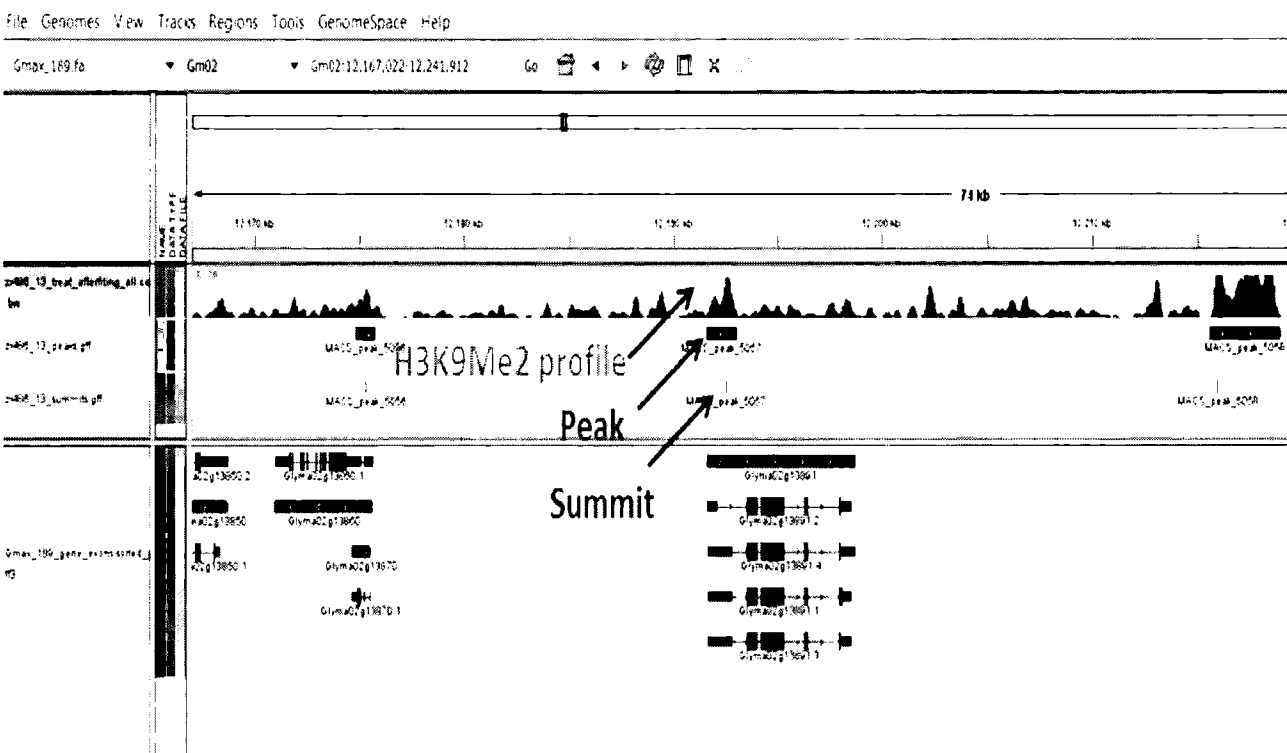
昆蟲抗性建構體的大豆領先事件被選擇且列於表2。因為這些領先事件的插入位置是已知的，在轉殖基因插入位置附近(5kb)的H3K9me2富集剖繪被評估以預測基因靜默的風險。因為所有領先事件係基於穩定的轉殖基因表現及穩定性被選擇，預期該轉殖基因係插入於一個具有低H3K9me2富集(峰數小於等於1)的基因座。在該轉殖基因插入位置的5kb窗中，九個相異的事件中，有六個事件顯示零個峰且三個事件顯示一個H3K9me2峰。該低H3K9me2富集對應於穩定的轉殖基因表現。此表示H3K9me2剖繪提供了一個有效的工具以預測基因靜默的風險。

表2.商業大豆事件中轉殖基因插入位置附近5kb的H3K9me2富集			
大豆事件	染色體	起始	終止
pDAB4468.04.16.1	Gm 04	46004353	46004354
pDAB4464.11.04.1	Gm 15	876318	876319
pDAB4472.16.06.1	Gm 08	8345349	8345350
pDAB9582.812.09.1	Gm 05	2810964	2810965
pDAB9582.814.19.1	Gm 02	10027415	10027416
pDAB9582.146.06.1	Gm 17	7737798	7737799
pDAB9582.140.11.1	Gm 18	57058706	57058707
pDAB9582.816.15.1	Gm 03	46300211	46300212
pDAB9582.816.12.1	Gm 04	48406612	48406613

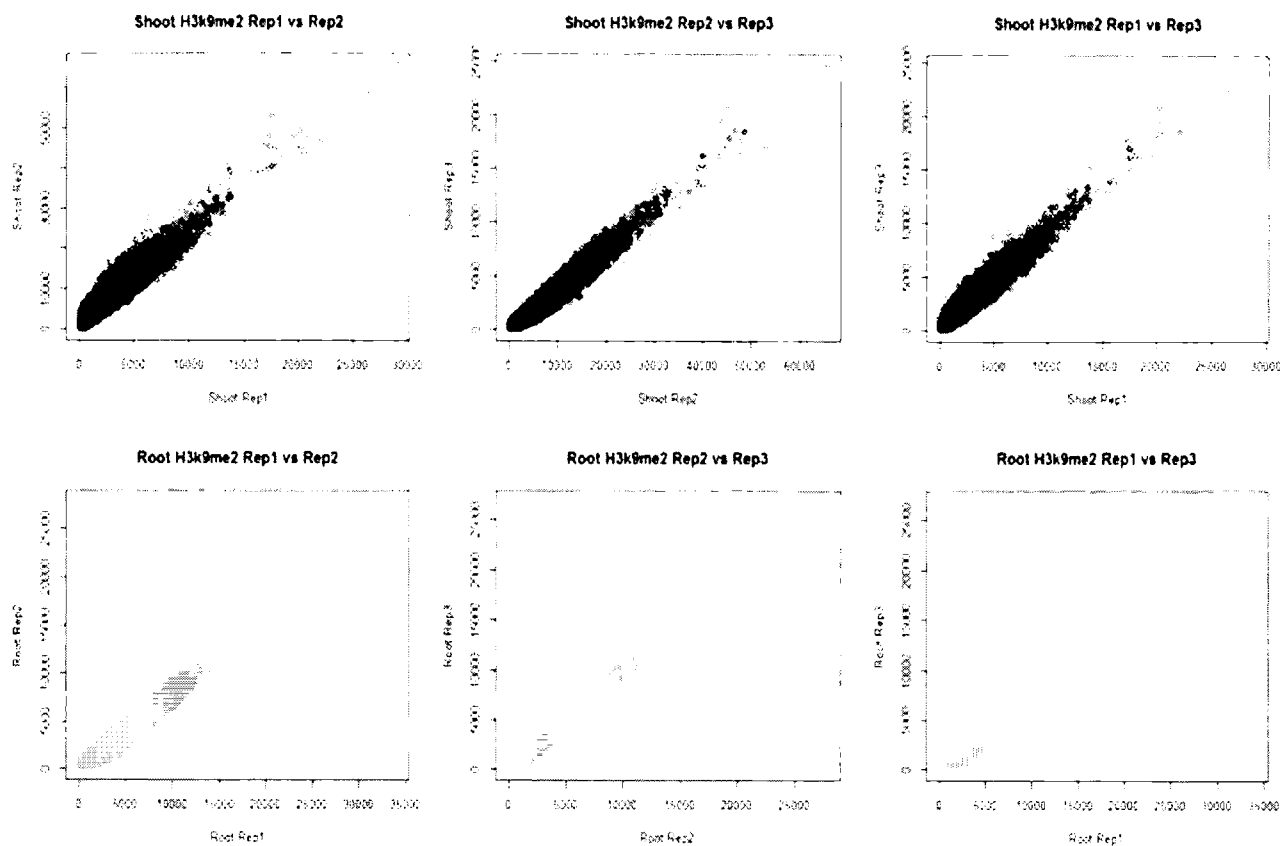
【符號說明】

(無)

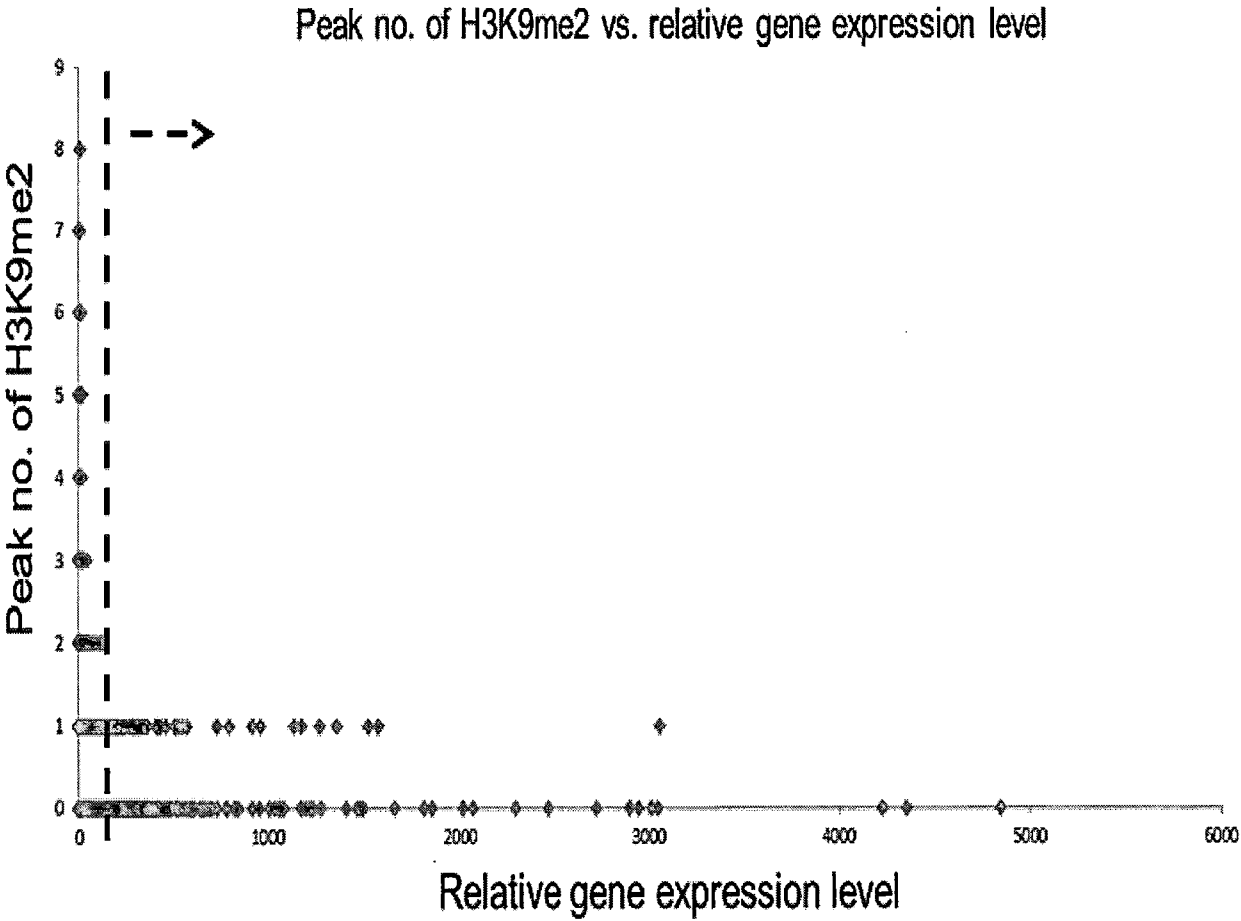
Appendix



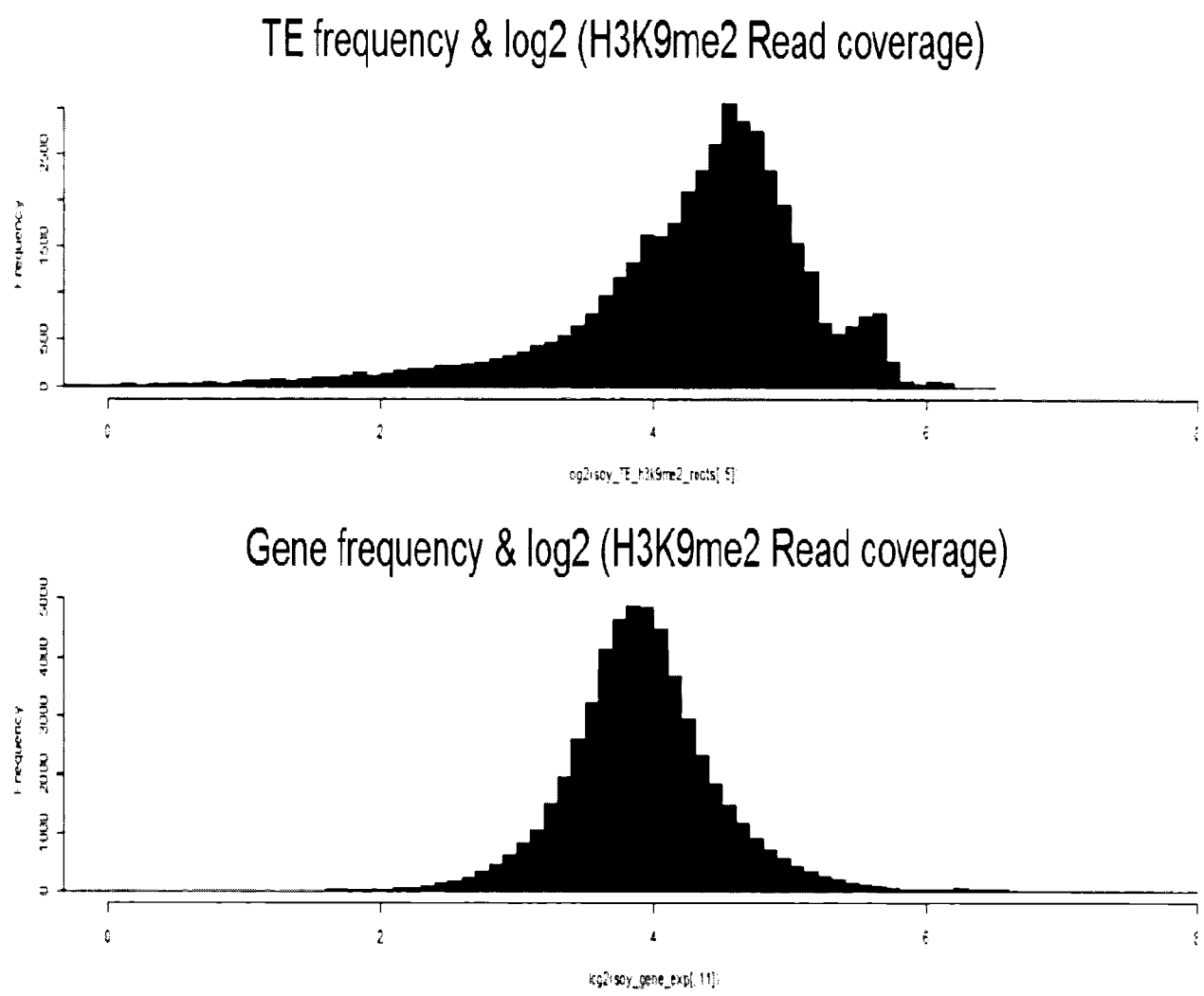
Appendix Figure 1: A genome browser visualization to show genome wide H3K9me2 profiling data mapped to the Soybean Williams 82 genome and an example of an identified H3K9me2 enrichment region (peak and summit).



Appendix Figure 2. ChIP-seq comparison among three technical replicates. These plots support that the CHIPseq protocols are highly replicable across root and shoot technical replicate samples.



Appendix Figure 3. Comparison of peak numbers of H3K9me2 within genic region and the corresponding relative gene expression



Appendix Figure 4. Distribution of H3K9me2 enrichment among transposon elements and genes.

【發明申請專利範圍】

【第1項】 一種用於預測轉殖基因靜默風險的方法，其包含：

(a) 產生一植物基因組的組蛋白甲基化/乙醯化剖繪(profile)資料；

(b) 匯集步驟(a)的組蛋白甲基化/乙醯化剖繪資料至一組蛋白甲基化/乙醯化資料庫，該資料庫顯示組蛋白甲基化/乙醯化剖繪的峰數(peak numbers)；

(c) 分析至少一轉殖基因插入位置的序列；及

(d) 將步驟(c)的轉殖基因插入位置的序列與步驟(b)的組蛋白甲基化/乙醯化資料庫比較；

其中(i)若是該轉殖基因插入位置落入該組蛋白甲基化/乙醯化剖繪的峰零(peak zero)或峰一(peak one)，則無轉殖基因靜默的風險，或

(ii)若是該轉殖基因插入位置落入該組蛋白甲基化/乙醯化剖繪的峰二(peak two)或以上，則有顯著的轉殖基因靜默風險。

【第2項】 如請求項1之方法，其中該組蛋白甲基化/乙醯化剖繪資料係使用一染色質免疫沉澱定序(ChIP-seq)測定法產生。

【第3項】 如請求項1之方法，其中該組蛋白甲基化/乙醯化剖繪資料係與選自於下列所組成之群組的組蛋白甲基化/乙醯化相關聯：H3K4me2、H3K4me3、H3K9/14ac、H3K9me2、H3K9me3、H3K27me1、H3K27me3 及

H4K20me3、及其等之多種組合。

【第4項】 如請求項1之方法，其中該組蛋白甲基化/乙醯化剖繪資料係與選自於下列所組成之群組的組蛋白甲基化/乙醯化相關聯：H3K9me2、H3K9me3、H3K27me1、H3K27me3及H4K20me3、及其等之多種組合。

【第5項】 如請求項1之方法，其中該組蛋白甲基化/乙醯化剖繪資料係與H3K9me2甲基化相關聯。

【第6項】 如請求項1之方法，其中該植物係選自於大豆、玉米、油菜、棉花、小麥及稻米。

【第7項】 如請求項1之方法，其中該轉殖基因插入位置的序列係藉由直接定序獲得。

【第8項】 一種用於預測轉殖基因靜默風險的電腦化系統，其包含：

(a) 一個組蛋白甲基化/乙醯化資料庫，其顯示一植物基因組中的組蛋白甲基化/乙醯化剖繪之峰數；

(b) 一個輸入模組，其讓一使用者輸入至少一轉殖基因插入位置的序列；及

(c) 一個輸出模組，其顯示轉殖基因靜默風險之預測，其中該預測係基於比較步驟(b)所輸入的序列與步驟(a)之組蛋白甲基化/乙醯化剖繪。

【第9項】 如請求項8之系統，其中該組蛋白甲基化/乙醯化剖繪資料係使用一染色質免疫沉澱定序(ChIP-seq)測定法產生。

【第10項】 如請求項8之系統，其中該組蛋白甲基化/

乙醯化剖繪資料係與選自於下列所組成之群組的組蛋白甲基化/乙醯化相關聯：H3K4me2、H3K4me3、H3K9/14ac、H3K9me2、H3K9me3、H3K27me1、H3K27me3及H4K20me3、及其等之多種組合。

【第11項】如請求項8之系統，其中該組蛋白甲基化/乙醯化剖繪資料係與選自於下列所組成之群組的組蛋白甲基化/乙醯化相關聯：H3K9me2、H3K9me3、H3K27me1、H3K27me3及H4K20me3、及其等之多種組合。

【第12項】如請求項8之系統，其中該組蛋白甲基化/乙醯化剖繪資料係與H3K9me2甲基化相關聯。

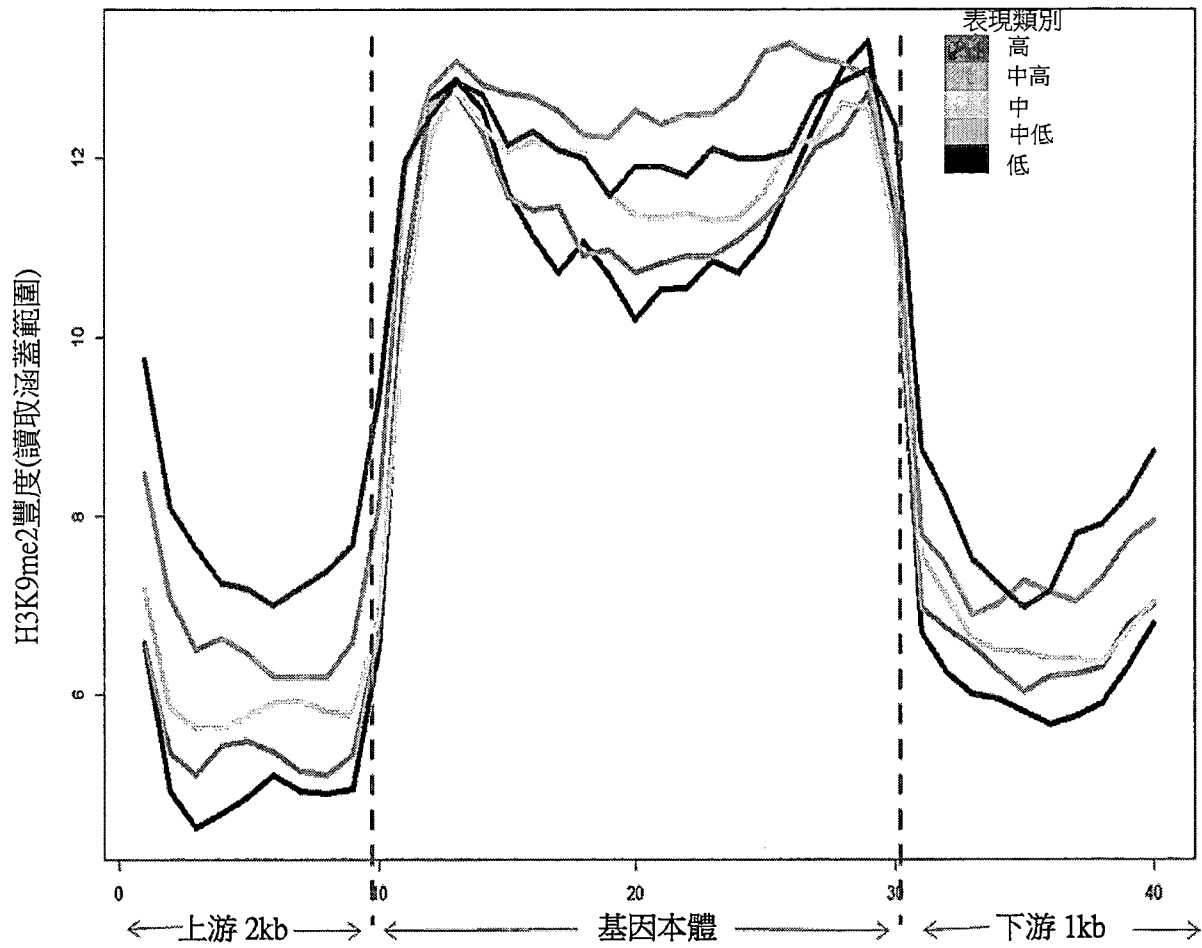
【第13項】如請求項8之系統，其中該植物係選自於大豆、玉米、油菜、棉花、小麥及稻米。

【第14項】一種供用於一用以預測轉殖基因靜默風險之電腦化系統中的方法，其包含：

(a) 經由一使用者輸入至少一轉殖基因插入位置的序列至如請求項8之系統；及

(b) 接收從如請求項8之系統就轉殖基因靜默風險預測的輸出，其中該預測係基於比較所輸入的轉殖基因插入位置的序列與該系統中之組蛋白甲基化/乙醯化剖繪。

【發明圖式】



【圖1】