

(19)



Евразийское
патентное
ведомство

(21)

201892288

(13)

A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2019.04.30

(51) Int. Cl. C12N 15/82 (2006.01)
C12N 9/02 (2006.01)

(22) Дата подачи заявки
2017.04.12

(54) ЯДЕРНО-КОДИРУЕМАЯ МУЖСКАЯ СТЕРИЛЬНОСТЬ КАК РЕЗУЛЬТАТ МУТАЦИИ ЦИТОХРОМ Р450 ОКСИДАЗЫ

(31) 10 2016 106 656.7

(32) 2016.04.12

(33) DE

(86) PCT/EP2017/058815

(87) WO 2017/178541 2017.10.19

(71) Заявитель:

КВС ЗААТ СЕ (DE)

(72) Изобретатель:

Борхардт Дитрих, Чарнецки Олаф,
Мехелке Вольфганг (DE)

(74) Представитель:

Вашук Т.В., Емельянова В.А.,
Королева С.В. (BY)

(57) Изобретение касается растений, имеющих рецессивный ядерно-кодируемый мужской стерильный фенотип, и коррелирующего с ним локуса гена (gst), включая ген, который ответственен за fertильный/стерильный фенотип и который мутирован в стерильном фенотипе. Изобретение также обеспечивает способы идентификации генотипа, коррелирующего с экспрессией признаков, указанных выше, и соответствующие генетические инструменты, такие как зонды для гибридизации и олигонуклеотиды. Описано также использование растений, получаемых по данному изобретению, в селекции гибридов и производстве продуктов из возобновляемых сырьевых материалов, таких как биоэтанол, биогаз и продукты на основе сахара.

A1

201892288

201892288

A1

ЯДЕРНО-КОДИРУЕМАЯ МУЖСКАЯ СТЕРИЛЬНОСТЬ КАК РЕЗУЛЬТАТ МУТАЦИИ ЦИТОХРОМ Р450 ОКСИДАЗЫ

Область техники

Настоящее изобретение относится к области упрощения трудоемких селекционных программ с помощью методов молекулярной биологии, генной инженерии и маркерных технологий. В частности, изобретение обеспечивает растения, проявляющие гомозиготный ядерно-кодируемый мужски стерильный фенотип в результате спонтанной мутации участка гена в ядерном геноме, причем в отличие от ЦМС (*цитоплазматической мужской стерильности*) мутация достигается посредством экспрессии рецессивного признака, что исключает необходимость получения стерильных и фертильных генотипов в селекционной программе. В этой связи настоящее изобретение обеспечивает растения, в частности растения сахарной свеклы и картофеля, в которых мутация, являющаяся результатом вышеупомянутой экспрессии признака, детектируется в гене цитохром P450 оксидазы (CYPgst) при помощи маркерных технологий, а также соответствующий способ идентификации данной мутации. Кроме того, изобретение обеспечивает белок CYPgst, молекулу ДНК, содержащую не мутировавший ген, обеспечивающий вышеупомянутую экспрессию признака, рекомбинантную молекулу ДНК, обеспечивающую получение гена дикого типа, промотор для специфических экспрессий такого гена или гетерологичный ген в цветках и/или плодах растений, и/или нуклеотидную последовательность, кодирующую ингибиторы гена CYPgst, а также соответствующие векторы и клетки-хозяева.

Настоящее изобретение также касается растений, модифицированных посредством генной инженерии, которые имеют рецессивный ядерно-кодируемый мужски стерильный фенотип в результате ингибирования экспрессии гена CYPgst, соответствующих ингибиторов, а также способов ингибирования гена и способов восстановления фертильности. Изобретение также касается использования таких растений в селекции гибридов, селекции на устойчивость и/или производстве семян. Также изобретение включает семена или потомков, органы, растительные части, ткани или клетки растений по данному изобретению, а также их применение.

Предпосылки создания изобретения

Для контролируемого получения изменения генетики растений сахарной свеклы (*Beta vulgaris*, подвид *vulgaris*), а также всех других культивируемых растений, осуществляется скрещивание. У еще закрытых цветков родительской формы вручную удаляются пыльники, несущие пыльцу. Опыление также проводится вручную путем нанесения пыльцы донора на рыльца родительской формы. Как альтернативный вариант, опыление может также осуществляться после удаления пыльников у цветков родительской формы путем помещения пыльцы донора в пространственной близости от родительской формы для цветения и доставки пыльцы. В любом случае это связано с процессом ведения трудоемких протоколов, который весьма предрасположен к ошибкам, так как если удаляется только часть пыльников цветка, то может иметь место самоопыление.

В настоящее время (комерческие) гибридные семена часто получают путем скрещивания между компонентами родительской формы генотипов, проявляющих ЦМС (*цитоплазматическую мужскую стерильность*), и последующего возвратного скрещивания, чтобы увеличить генетическую часть компонентов родительской формы. Получаемые в результате мужски стерильные компоненты родительской формы можно затем выращивать в большом количестве, при этом опыление происходит вследствие того, что доноры пыльцы растут рядом (*топкроссный* метод). Поскольку доминантный ген ЦМС наследуется по материнской линии, то для использования линий с ЦМС одновременно должны обеспечиваться фертильные линии-закрепители стерильности (не являющиеся линиями О-типов из ЦМС), чтобы могло иметь место опыление ЦМС-линий. Это требует большого объема работ по планированию и значительных производственных усилий, а также сложной логистики. Например, для получения семян достаточно высокого качества коммерчески значимые сорта сахарной свеклы в настоящее время являются тройными гибридами. Создание гибридов в селекционных программах также является дорогостоящим и трудозатратным и пока связано с наличием преград на этом пути.

Отмечены линии или генотипы многочисленных растений, проявляющие естественно встречающуюся ядерно-кодируемую мужскую стерильность (*mc: мужская стерильность*). Она обычно вызывается спонтанной мутацией гена в ядерном геноме, которая происходит в результате экспрессии рецессивного признака. Использование генотипов, являющихся ядерно-кодируемыми мужски стерильными по фенотипу, подходит для упрощения процессов селекции и/или применения этих процессов для получения гибридных семян. Таким образом, фенотипы с мужской стерильностью обладают тем преимуществом, что их использование не требует ручного удаления пыльников, а также исключает необходимость одновременного получения фертильных и стерильных генотипов, поскольку в свою очередь гетерозиготные генотипы в каждом репродуктивном цикле расщепляются на фертильные и стерильные индивиды посредством самоопыления.

Исходя из вышеизложенного, целью настоящего изобретения является обеспечение средств и способов для использования ядерно-кодируемой мужской стерильности у культурных растений, в частности, у растений сахарной свеклы и картофеля. Данная цель достигается посредством вариантов осуществления изобретения, указанных в описании и формуле изобретения.

Краткое изложение сущности изобретения

Настоящее изобретение относится к области упрощения трудоемких селекционных программ, маркерной технологии и генной инженерии. Изобретение обеспечивает растения, проявляющие рецессивный ядерно-кодируемый мужски стерильный фенотип в результате мутации сегмента ДНК, содержащего цитохром Р450 оксидазы (*CYPgst*). Ген *CYPgst* и мутация идентифицируются путем использования маркерных технологий и методов молекулярной биологии. Поскольку мутация остается интактной вследствие экспрессии рецессивного признака, и гетерозиготные генотипы в результате самоопыления в ходе каждого репродуктивного цикла затем расщепляются на фертильные и стерильные генотипы, то не требуется одновременное получение фертильных линий-закрепителей стерильности. Полученные данные можно использовать для создания трансгенных растений, имеющих рецессивный ядерно-кодируемый мужски стерильный фенотип, а также для восстановления фертильности.

Таким образом, настояще изобретение касается вариантов осуществления, представленных в следующих параграфах [1] - [36] и иллюстрируемых примерами.

[1] Растение, в частности культурное растение, проявляющее рецессивный ядерно-кодируемый мужски стерильный фенотип, при этом фенотип обусловливается мутацией гена эндогенной цитохром Р450 оксидазы (*CYPgst*) или отсутствием, пониженным содержанием или пониженнной активностью функционального белка *CYPgst*, кодируемого геном дикого типа *CYPgst*, по сравнению с соответствующим (мужски фертильным) растением дикого типа, причем не мутировавший ген *CYPgst* представляет собой а) ген *BvCYPgst Beta vulgaris*, который предпочтительно содержит нуклеотидную последовательность SEQ ID No. 1 или 2 или кодирует аминокислотную последовательность, представленную в SEQ ID No. 3, либо его гомолог, аналог или ортолог, б) ген *StCYPgst Solanum tuberosum*, который предпочтительно содержит нуклеотидную последовательность SEQ ID No. 12 или 13 или кодирует аминокислотную последовательность, представленную в SEQ ID No. 14, либо его гомолог, аналог или ортолог, с) ген *ZmCYPgst Zea mays*, который предпочтительно содержит нуклеотидную последовательность SEQ ID No. 9 или 10 или кодирует аминокислотную последовательность, представленную в SEQ ID No. 11, либо его гомолог, аналог или ортолог.

[2] Растение согласно параграфу [1], являющееся гетерозиготным по мутации и мужски фертильным или гомозиготным по мутации и мужски стерильным, в котором развитие функциональной пыльцы супрессируется, предпочтительно супрессируется полностью, у стерильных растений.

[3] Растение согласно параграфу [1] или [2], в котором ген *CYPgst* экспрессируется по меньшей мере в закрытых цветках и плодах.

[4] Растение согласно любому из параграфов [1] - [3], в котором мутация предотвращает транскрипцию и/или трансляцию функционального белка, при этом предпочтительно мутация касается делеции, присоединения, инсерции или замещения в кодирующей нуклеотидной последовательности гена *CYPgst*, последовательности сигнала для сплайсинга или регуляторной последовательности, предпочтительно в последовательности промотора гена *CYPgst*. В предпочтительном варианте осуществления изобретения молекула нуклеиновой кислоты имеет мутацию, которую можно также встретить в нуклеотидной последовательности SEQ ID No. 8, в отличие от гена дикого типа согласно SEQ ID No. 1. В частности, мутация может представлять собой делецию между нуклеотидными позициями 1560 и 2095 SEQ ID No. 1 или соответствующими позициями SEQ ID No. 12 или 9. Делеция может содержать по меньшей мере 20, 30 или 50 последовательных пар оснований, предпочтительно по меньшей мере 100, 150, 200 или 250 последовательных пар оснований и наиболее предпочтительно по меньшей мере 300, 400 или 500 последовательных пар оснований. В наиболее предпочтительном варианте осуществления изобретения молекула нуклеиновой кислоты содержит нуклеотидную последовательность согласно SEQ ID No. 8. В другом предпочтительном варианте осуществления изобретения молекула нуклеиновой кислоты имеет точечную мутацию в нуклеотидной последовательности SEQ ID No. 1 (Таблица 1), предпочтительно между нуклеотидными позициями 1560 и 2095 SEQ ID No. 1.

[5] Растение согласно параграфу [4], являющееся растением *Beta vulgaris*, предпочтительно растением *Beta vulgaris*, подвид *vulgaris*, в котором делецию можно детектировать посредством определения отсутствия одного или обоих маркерных локусов

sle5983d14 (продукт амплификации праймера, имеющего последовательности SEQ ID No. 4 и 5) и sle5983d17 (продукт амплификации праймера, имеющего последовательности SEQ ID No. 6 и 7), а также посредством определения присутствия убиквитарного маркера.

[6] Растение согласно любому из параграфов [1] - [5], в котором ген локализуется в *Beta vulgaris*, предпочтительно *Beta vulgaris*, подвид *vulgaris*, в сегменте хромосомы 1 между маркерными локусами sxn215s01 и sle3305s02, при этом последовательность маркера sxn215s01, показанная в SEQ ID No. 24, и последовательность маркера sle3305s02, показанная в SEQ ID No. 26, указывают на присутствие локуса *gst*, и последовательность маркера sxn215s01, показанная в SEQ ID No. 25, и последовательность маркера sle3305s02, показанная в SEQ ID No. 27, указывают на референсную последовательность.

[7] Растение согласно параграфу [6], в котором длина последовательности составляет приблизительно от 50 до 5000 kbp, от 100 до 1000 kbp, более предпочтительно от 100 до 500 kbp и наиболее предпочтительно от 200 до 250 kbp.

[8] Растение согласно любому из параграфов [1] - [7], в котором не мутировавший ген представляет собой функциональный ген *BvCYPgst Beta vulgaris*, предпочтительно *Beta vulgaris*, подвид *vulgaris*, или функциональный гомологичный, аналогичный или ортологичный ген другого культивируемого(CYPgst) или культурного растения.

[9] Растение согласно параграфу [8], в котором гомологичный, аналогичный или ортологичный ген представляет собой ген *Zea mays*, который предпочтительно содержит нуклеотидную последовательность SEQ ID No. 9 или 10 или кодирует аминокислотную последовательность, представленную в SEQ ID No. 11, ген *Solanum tuberosum*, который предпочтительно содержит нуклеотидную последовательность SEQ ID No. 12 или 13 или кодирует аминокислотную последовательность, представленную в SEQ ID No. 14, ген *Triticum aestivum*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 15, ген *Helianthus annuus*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 16, ген *Hordeum vulgare*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 17, ген *Brassica napus*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 18, ген *Brassica oleracea*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 19, ген *Brassica rapa*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 20, ген *Glycine max*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 21, ген *Gossypium*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 22, или ген *Sorghum bicolor*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 23.

[10] Растение согласно любому из параграфов [1] - [9], в котором не мутировавший ген (ген дикого типа) имеет нуклеотидную последовательность, выбиаемую из группы, состоящей из:

(а) нуклеотидной последовательности, представленной нуклеотидной последовательностью SEQ ID NO: 1 или SEQ ID NO: 2, либо ее функционального фрагмента (см. Рис. 4А и 4В);

- (b) нуклеотидной последовательности, которая кодирует аминокислотную последовательность, представленную в SEQ ID NO: 3;
- (c) нуклеотидной последовательности, способной гибридизоваться в жестких условиях с нуклеотидной последовательностью, комплементарной нуклеотидной последовательности согласно (a) или (b);
- (d) нуклеотидной последовательности, которая кодирует аминокислотную последовательность, имеющую различия с аминокислотной последовательностью SEQ ID NO: 3, проявляющиеся в форме аминокислотных делеций, замещений, присоединений и/или инсерций в аминокислотной последовательности, и которая предпочтительно идентична по меньшей мере на 60% аминокислотной последовательности по всей ее длине;
- (e) нуклеотидной последовательности, которая кодирует белок, обладающий одной и той же ферментативной активностью, что и белок, кодируемый нуклеотидной последовательностью согласной любой из нуклеотидных последовательностей (a) – (d); и
- ((f) нуклеотидной последовательности, содержащей по меньшей мере 200 или 400, предпочтительно по меньшей мере 600 или 800, наиболее предпочтительно по меньшей мере 1000 последовательных нуклеотидов промотора последовательности нукleinовой кислоты SEQ ID No. 1 с нуклеотидными позициями от 1 до 1518, предпочтительно от 518 до 1518, наиболее предпочтительно от 1318 до 1518, или последовательности, которая гибридизуется в этом участке, причем нуклеотидная последовательность способна контролировать экспрессию гена, и гетерологичная молекула нукleinовой кислоты оперативно связана с нуклеотидной последовательностью, особенно в закрытых цветках или плодах.

Нуклеотидная последовательность согласно (c), (d) или (e) представляет собой, например, нуклеотидную последовательность, которая имеет последовательность нуклеотидов, представленную в SEQ ID No. 12 или SEQ ID No. 13, или ее функциональный фрагмент, либо которая кодирует аминокислотную последовательность, представленную в SEQ ID No. 14. Также нуклеотидная последовательность согласно (c), (d) или (e) представляет собой, например, нуклеотидную последовательность, которая имеет последовательность нуклеотидов, представленную в SEQ ID No. 9 или SEQ ID No. 10, или ее функциональный фрагмент, либо которая кодирует аминокислотную последовательность, представленную в SEQ ID No. 11. Кроме того, нуклеотидная последовательность согласно (c), (d) или (e) представляет собой, например, нуклеотидную последовательность, которая кодирует аминокислотную последовательность, выбираемую из группы, состоящей из последовательностей SEQ ID No. 15-23.

[11] Растение согласно любому из параграфов [1] - [10], характеризующееся тем, что оно является инбредным или гибридным растением.

[12] Растение согласно любому из параграфов [1] - [11], характеризующееся тем, что оно является растением рода *Zea*, *Solanum*, *Triticum*, *Triticale*, *Helianthus*, *Secale*, *Hordeum*, *Brassica*, *Brachypodium*, *Glycine*, *Gossypium*, *Sorghum*, *Saccharum*, *Setaria*, *Aegilops*, *Oryza*, *Daucus*, *Eucalyptus*, *Erythranthe*, *Genlisea*, *Musa*, *Avena*, *Nicotiana*, *Coffea*, *Vitis*, *Cucumis*, *Morus*, *Cruciiflora*, *Cardamine*, *Lepidium*, *Capsella*, *Olimarabidopsis*, *Arabis*, *Raphanus*, *Eruca*, *Citrus*, *Jatropha*, *Populus* или *Beta*, предпочтительно растение вида *Zea mays*, *Solanum tuberosum*, *Triticum aestivum*, *Triticum durum*, *Triticum spelta*, *Helianthus annuus*, *Secale cereal*,

Hordeum vulgare, Hordeum bulbosum, Brassica napus, Brassica oleracea, Brassica rapa, Brassica juncacea, Brassica nigra, Glycine max, Gossypium sp., Sorghum bicolor, Triticale, Saccharum officinarum, Setaria italica, Oryza sativa, Oryza minuta, Oryza australiensis, Oryza alta, Brachypodium distachyon, Hordeum marinum, Aegilops tauschii, Daucus glochidiantus, Daucus pusillus, Daucus muricatus, Daucus carota, Eucalyptus grandis, Erythranthe guttata, Genlisea aurea, Musa sp., Avena sp., Nicotiana sylvestris, Nicotiana tabacum, Nicotiana tomentosiformis, Solanum lycopersicum, Coffea canephora, Vitis vinifera, Cucumis sativus, Morus nobilis, Arabidopsis thaliana, Arabidopsis lyrata, Arabidopsis arenosa, Crucihimalaya himalaica, Crucihimalaya wallichii, Cardamine flexuosa, Lepidium virginicum, Capsella bursa-pastoris, Olimarabidopsis pumila, Arabis hirsute, Raphanus sativus, Eruca vesicaria sativa, Citrus x sinensis, Jatropha curcas, Populus trichocarpa или Beta vulgaris.

[13] Молекула нуклеиновой кислоты или рекомбинантная молекула ДНК, содержащая нуклеотидную последовательность, представленную в параграфе [10].

[14] Рекомбинантная молекула ДНК согласно параграфу [13], содержащая (i) промотор, имеющий нуклеотидную последовательность, представленную в параграфе [10] (f), которая оперативно связана с гетерологичной молекулой нуклеиновой кислоты, либо (ii) кодирующую нуклеотидную последовательность, представленную в параграфе [10] (a) – (e), которая оперативно связана с гетерологичным промотором, и предпочтительно способная управлять экспрессией нуклеотидной последовательности, особенно в закрытых цветках или плодах.

[15] Рекомбинантная молекула ДНК, содержащая нуклеотидную последовательность, кодирующую shРНК (*малую РНК, образующую шильки*), siРНК (*малую интерферирующую РНК*), отрицательно-полярную РНК, положительно-полярную РНК или двухцепочечную РНК, которая приводит к ингибираванию экспрессии функционального (не мутировавшего) гена CYPgst после ее экспрессии в растительной клетке или после ее введения в растительную клетку. В предпочтительном варианте осуществления изобретения нуклеотидная последовательность состоит по меньшей мере из 15, 16, 17, 18, 19 или 20, предпочтительно по меньшей мере из 21, 22, 23, 24 или 25, более предпочтительно по меньшей мере из 30, 35, 40, 45 или 50, наиболее предпочтительно по меньшей мере из 100, 200, 300, 500 или 1000 последовательных нуклеотидов SEQ ID No. 1, 2, 9, 10, 12 или 13 в положительно- или отрицательно-полярной ориентации, либо по меньшей мере одного экзона 1 SEQ ID No. 1 из нуклеотидов в позиции 1762-2679 и экзона 2 SEQ ID No. 1 из нуклеотидов в позиции 3507-4142. Экзон 1 SEQ ID No. 12 простирается от нуклеотидной позиции 1762 до 2032, экзон 2 SEQ ID No. 12 протирается от нуклеотидной позиции 2449 до 2161 и экзон 3 SEQ ID No. 12 протирается от нуклеотидной позиции 4032 до 4694. Экзон 1 SEQ ID No. 9 простирается от нуклеотидной позиции 2001 до 2927 и экзон 2 SEQ ID No. 9 простирается от нуклеотидной позиции 3018 до 3683. В еще одном предпочтительном варианте осуществления изобретения нуклеотидная последовательность состоит по меньшей мере из 15, 16, 17, 18, 19 или 20, предпочтительно по меньшей мере из 21, 22, 23, 24 или 25, более предпочтительно по меньшей мере из 30, 35, 40, 45 или 50 и наиболее предпочтительно по меньшей мере из 100, 200, 300, 500 или 1000 последовательных нуклеотидов, которые способны гибридизоваться с нуклеотидной последовательностью, представленной в параграфе [10].

[16] Молекула нуклеиновой кислоты, содержащая нуклеотидную последовательность, представленную в параграфе [10], с мутацией в форме делеции, присоединения, инсерции или замещения, в которой данная мутация не приводит к синтезу функциональных белков CYPgst. Предпочтительно мутация происходит в кодирующей нуклеотидной последовательности гена CYPgst, локализуемой в сигнале для сплайсинга, или в регуляторной последовательности гена CYPgst, предпочтительно в промоторе гена CYPgst. В предпочтительном варианте осуществления изобретения молекула нуклеиновой кислоты имеет мутацию, которая можно также обнаружить в нуклеотидной последовательности согласно SEQ ID No. 8. В частности, мутация может представлять собой делецию между нуклеотидными позициями 1560 и 2095 SEQ ID No. 1 или между соответствующими позициями в SEQ ID No. 12 или 9. Делеция может иметь длину по меньшей мере в 20, 30 или 50 последовательных пар оснований, более предпочтительно по меньшей мере в 100, 150, 200 или 250 последовательных пар оснований и наиболее предпочтительно по меньшей мере в 300, 400 или 500 последовательных пар оснований. В особенно предпочтительном варианте осуществления изобретения молекула нуклеиновой кислоты содержит нуклеотидную последовательность согласно SEQ ID No. 8. В еще одном предпочтительном варианте осуществления изобретения молекула нуклеиновой кислоты имеет точечную мутацию в нуклеотидной последовательности SEQ ID No. 1 согласно Таблице 1, предпочтительно между нуклеотидными позициями 1560 и 2095 SEQ ID No. 1.

[17] Молекула нуклеиновой кислоты, состоящая по меньшей мере из 15, 16, 17, 18, 19 или 20, предпочтительно по меньшей мере из 21, 22, 23, 24 или 25, более предпочтительно по меньшей мере из 30, 35, 40, 45 или 50, наиболее предпочтительно по меньшей мере из 100, 200, 300, 500 или 1000 последовательных нуклеотидов SEQ ID No. 1, 2, 9, 10, 12 или 13 в положительно- и/или отрицательно-полярной ориентации, или по меньшей мере одного экзона последовательностей SEQ ID No. 1, 9 или 12 в положительно- и/или отрицательно-полярной ориентации. Экзон 1 SEQ ID No. 1 простирается от нуклеотидной позиции 1762 до 2679, экзон 2 SEQ ID No. 1 простирается от нуклеотидной позиции 3507 до 4142. Экзон 1 SEQ ID No. 12 простирается от нуклеотидной позиции 1762 до 2032, экзон 2 SEQ ID No. 12 простирается от нуклеотидной позиции 2449 до 3161 и экзон 3 SEQ ID No. 12 простирается от нуклеотидной позиции 4032 до 4694. Экзон 1 SEQ ID No. 9 простирается от нуклеотидной позиции 2001 до 2927 и экзон 2 SEQ ID No. 9 простирается от нуклеотидной позиции 3018 до 3683. В специальном случае осуществления изобретения длина молекулы нукleinовой кислоты превышает размер по меньшей мере одного интрана SEQ ID No. 1, SEQ ID No. 9 или SEQ ID No. 12, т.е. молекула нукleinовой кислоты содержит последовательно расположенные i) по меньшей мере один нуклеотид с 3'-конца экзона 1 SEQ ID No. 1 (предпочтительно последний нуклеотид экзона 1 SEQ ID No. 1 в направлении 5' - 3'; соответствующий нуклеотиду в позиции 583 SEQ ID No. 2) и по меньшей мере один нуклеотид с 5'-конца экзона 2 SEQ ID No. 1 (предпочтительно первый нуклеотид экзона 2 SEQ ID No. 1 в направлении 5' - 3'; соответствующий нуклеотиду в позиции 584 SEQ ID No. 2), ii) по меньшей мере один нуклеотид с 3'-конца экзона 1 SEQ ID No. 12 (предпочтительно последний нуклеотид экзона 1 SEQ ID No. 12 в направлении 5' - 3'; соответствующий нуклеотиду в позиции 271 SEQ ID No. 13) и по меньшей мере один нуклеотид с 5'-конца экзона 2 SEQ ID No. 12 (предпочтительно первый нуклеотид экзона 2 SEQ ID No. 12 в направлении 5' - 3'; соответствующий нуклеотиду в позиции 272 SEQ ID No. 13), iii) по меньшей мере один нуклеотид с 3'-конца экзона 2 SEQ ID No. 12 (предпочтительно последний нуклеотид экзона 2 SEQ ID No. 12 в направлении 5' - 3'; соответствующий

нуклеотиду в позиции 984 SEQ ID No. 13) и по меньшей мере один нуклеотид экзона 3 SEQ ID No. 12 в направлении 5' - 3'; соответствующий нуклеотиду в позиции 985 SEQ ID No. 13), либо iv) по меньшей мере один нуклеотид с 3'-конца экзона 1 SEQ ID No. 9 (предпочтительно последний нуклеотид экзона 1 SEQ ID No. 9 в направлении 5' - 3'; соответствующий нуклеотиду в позиции 987 SEQ ID No. 10) и по меньшей мере один нуклеотид с 5'-конца экзона 2 SEQ ID No. 9 (предпочтительно первый нуклеотид экзона 2 SEQ ID No. 9 в направлении 5' - 3'; соответствующий нуклеотиду в позиции 928 SEQ ID No. 10). В еще одном предпочтительном варианте осуществления изобретения нуклеотидная последовательность состоит по меньшей мере из 15, 16, 17, 18, 19 или 20, предпочтительно по меньшей мере из 21, 22, 23, 24 или 25, более предпочтительно по меньшей мере из 30, 35, 40, 45 или 50 и наиболее предпочтительно по меньшей мере из 100, 200, 300, 500 или 1000 последовательных нуклеотидов, и способна специфически гибридизоваться с нуклеотидной последовательностью, представленной в [10] или [16].

[18] Олигонуклеотид длиной предпочтительно не более 50 нуклеотидов, содержащий молекулу нукleinовой кислоты согласно параграфу [17] или молекулу нукleinовой кислоты, способную специфически гибридизоваться с нуклеотидной последовательностью согласно SEQ ID No. 8, и/или предпочтительно имеющий одну из следующих нуклеотидных последовательностей:

- i) SEQ ID No. 4, 6 или их комплемент, либо
- ii) SEQ ID No. 5, 7 или их комплемент.

[19] Вектор, предпочтительно растительный вектор, содержащий молекулу ДНК или молекулу нукleinовой кислоты согласно любому из параграфов [13]- [16] или молекулу нукleinовой кислоты согласно параграфу [17].

[20] Вектор согласно параграфу [19], в котором молекула ДНК или молекула нукleinовой кислоты в виде трансгена способна экспрессировать функциональный CYPgst и предпочтительно генетически связана с другим трансгеном, предотвращающим перенос молекулы ДНК или молекулы нукleinовой кислоты через пыльцу, при этом предпочтительно вектор или трансген также содержит экспрессионную кассету, которая маркирует семена, предпочтительно флуоресцентными метками.

[21] Клетка-хозяин, предпочтительно растительная клетка, содержащая рекомбинантную молекулу ДНК или молекулу нукleinовой кислоты согласно любому из параграфов [13] - [16], или вектор согласно параграфу [19] или [20].

[22] Белок CYPgst, кодируемый нуклеотидной последовательностью, представленной в параграфе [10], или его функциональный и/или иммунологически активный фрагмент. Белок CYPgst предпочтительно представляет собой а) аминокислотную последовательность, выбираемую из группы, состоящей из SEQ ID No. 3, SEQ ID No. 11 и SEQ ID No. 14-23, или б) аминокислотную последовательность, которая по меньшей мере на 80%, 82%, 84%, 86% или 88%, предпочтительно по меньшей мере на 90%, 91%, 92%, 93%, 94% или 95%, наиболее предпочтительно по меньшей мере на 96%, 97%, 98%, 99% или 99,5% идентична аминокислотной последовательности SEQ ID No. 3, предпочтительно по всей ее длине.

[23] Антитело, которое специфически связывается с белком CYPgst или его фрагментом согласно параграфу [22].

[24] Набор, содержащий молекулу ДНК или молекулу нуклеиновой кислоты согласно любому из параграфов [13] - [16], нуклеиновую кислоту согласно параграфу [17], олигонуклеотид согласно параграфу [18], вектор согласно параграфу [19] или [20], белок CYPgst или его фрагмент согласно параграфу [22] и/или антитело согласно параграфу [23], а также потенциально реагенты, предназначенные для процедур, основанных на детекции нуклеиновых кислот, или для иммунологической детекции.

[25] Способ получения растения, в частности культурного растения, проявляющего гомозиготный рецессивный ядерно-кодируемый мужски стерильный фенотип, при котором ингибируется экспрессия гена CYPgst.

[26] Способ согласно параграфу [25], характеризующийся тем тем, что он включает в себя этап введения рекомбинантной молекулы ДНК согласно параграфу [15], молекулы нуклеиновой кислоты согласно параграфу [16] или [17] или вектора согласно параграфу [19] или [20], например, посредством агробактериальной трансформации, Т-ДНК мечения, гомологичной рекомбинации, мутагенеза, такого как TILLING, и направленного мутагенеза, например, с использованием нуклеаз «цинковые пальцы», TALE-нуклеаз (эффекторные нуклеазы, подобные активаторам транскрипции) и системы CRISPR/Cas, приводящий к ингибированию экспрессии гена, например, посредством iPHK, косупрессии, либо вследствие введенной мутации.

[27] Способ восстановления fertильности растения согласно любому из параграфов [1] - [12] или растения, которое можно получать посредством способа согласно параграфу [25] или [26], включающий введение в растение функционального гена CYPgst.

[28] Способ согласно параграфу [27], при котором ген CYPgst вводится с помощью рекомбинантной ДНК согласно параграфу [13] или [14], вектора согласно параграфу [19] или [20], либо посредством скрещивания растения, несущего ген CYPgst дикого типа или функциональный ген CYPgst, предпочтительно в гомозиготном состоянии. Как вариант, отбор по присутствию гена CYPgst дикого типа или функционального гена CYPgst можно проводить в следующем поколении.

[29] Растение, содержащее растительные клетки согласно параграфу [21], и/или которое можно получать способом согласно любому из параграфов [25] - [28].

[30] Орган, растительная часть, ткань или клетка растения согласно любому из параграфов [1] - [12] или [19].

[31] Семена или потомки растения согласно любому из параграфов [1] - [12] или [29], характеризующиеся тем, что семена или потомки содержат мутацию, представленную в любом из параграфов [1] - [12], и/или рекомбинантную молекулу ДНК или молекулу нуклеиновой кислоты согласно любому из параграфов [13] - [16] или вектор согласно параграфу [19] или [20].

[32] Способ идентификации растения согласно любому из параграфов [1] - [12] или [29] путем детекции мутации гена CYPgst или маркера, связанного с мутацией.

[33] Использование молекулы ДНК или молекулы нуклеиновой кислоты согласно любому из параграфов [13] - [16], молекулы нуклеиновой кислоты согласно параграфу [17], олигонуклеотида согласно параграфу [18], вектора согласно параграфу [19] или [20], белка CYPgst или его фрагмента согласно параграфу [22], антитела согласно параграфу [23] и/или наборов согласно параграфу [24], предназначенных для получения растения согласно любому из параграфов [1] - [12] или [29], получения рецессивного ядерно-кодируемого мужски стерильного растения, растения с восстановленной fertильностью, гибридного растения, в селекционных программах на устойчивость или для производства семян.

[34] Использование молекулы ДНК согласно параграфу [14] или промотора, представленного в параграфе [14], для специфической экспрессии гетерологичных молекул нуклеиновой кислоты в цветках и/или плодах растений.

[35] Использование растения согласно любому из параграфов [1] - [12] или [29], органа, растительной части, ткани или клетки согласно параграфу [30], семян или потомков согласно параграфу [31] или растения, которое можно идентифицировать согласно параграфу [32] или получать посредством осуществления использования согласно параграфу [33] или [34] или тканей, клеток, потомков или семян для производства пищевых продуктов, активных веществ, лекарственных средств и их прекурсоров, диагностических продуктов, косметики, чистых химикатов, сахара, патоки, биоэтанола или биогаза.

[36] Пищевые продукты, корма или активные вещества, содержащиеся в растении согласно любому из параграфов [1] - [12] или [29], органе, растительной части, ткани или клетке согласно параграфу [30], семенах или потомках согласно параграфу [31] или в растении, которое идентифицируется при помощи способа согласно параграфу [32] или которое можно получать посредством осуществления использования согласно параграфу [33] или [34], либо его тканей, клеток, потомков или семен.

[37] Использование растения согласно любому из параграфов [1] - [12] или [29] для селекции или получения растения-потомка, при которых для рекуррентного отбора используется ядерно-кодируемый мужски стерильный фенотип.

Некоторые термины, использованные в настоящей заявке, обсуждаются ниже более подробно.

Выражения «сегмент хромосомы», «хромосомный сегмент», а также их вариации, используются взаимозаменяемо, если не указано иное, и означают конкретный хромосомный сегмент ДНК конкретной хромосомы, содержащий по меньшей мере один ген.

«Ген CYPgst» или «ген CYPgst дикого типа», кодируемый «белком CYPgst», играет роль в образовании жизнеспособной пыльцы, причем мутация CYPgst вызывает мужскую стерильность у растений вследствие супрессии образования функциональной пыльцы. Это было экспериментально показано на примере сахарной свеклы (*Beta vulgaris*, подвид *vulgaris*). Сравнение гомологий позволило установить сходство с геном CYP703 *Arabidopsis thaliana*, который, как следует из существующего уровня техники (Morant *et al.*, The Plant Cell, 19 (2007), 1473-1487), выполняет основную функцию в синтезе спорополленина и предпочтительно катализирует преобразование среднцепочечных насыщенных жирных кислот в соответствующие простые гидроксилированные жирные кислоты, предпочтительно в процессе гидроксилирования лауриновой кислоты в позиции С-7. Элиминация гена

CYP703 в *Arabidopsis thaliana* приводит к частичной мужской стерильности. Количество пыльцы действительно сокращалось, однако функциональная пыльца все еще могла образовываться. В связи с этим ген *Beta vulgaris*, подвид *vulgaris*, по-видимому, выполняет иную функцию, так как его элиминация приводит к мужской стерильности растения. Безотносительно какой-либо теории, можно предположить, что ген CYPgst должен быть отнесен к другой классификации генов CYP, либо что у культивируемых растений, в частности у культурных растений, таких как сахарная свекла, он выполняет иную функцию или имеет иное значение в отличие от низшего модельного растения *Arabidopsis thaliana*; см. обсуждение в Примере 1. Специалист в данной области техники может получить информацию о других белках CYPgst из баз данных, используя другие поисковые параметры и компьютерные программы для скрининга гомологичных последовательностей или сравнения последовательностей. Возможную оценку того, выполняет ли идентифицированный ген ту же функцию, что и ген CYPgst *Beta vulgaris*, подвид *vulgaris*, можно проводить путем восстановления функции CYPgst в мужски стерильном растении сахарной свеклы посредством гетерологической экспрессии идентифицированного гена, т.е. путем восстановления фертильности с помощью трансгена.

Безотносительно какой-либо теории, представляется также возможным, что ген CYPgst *Beta vulgaris*, подвид *vulgaris*, и ген CYP703 *Arabidopsis thaliana* относятся к одному и тому же семейству CYP и, таким образом, выполняют ту же или по меньшей мере сходную функцию в синтезе спорополленина, однако у культивируемых растений, которые годами направленной селекции и скрещивания оптимизировались относительно урожайности, устойчивости к вредителям и абиотическим стрессовым факторам и т.п., а также содержания растительных веществ, способность компенсировать потерю спорополленина утрачена и поэтому в результате отсутствия спорополленина образование пыльцы супрессируется, что делает растения мужски стерильными.

Согласно изобретению термин «локус *gst*» относится к геномной ДНК растения, в частности культурного растения, где мутация обусловливается рецессивно-наследуемой ядерно-кодирующей мужской стерильностью, причем мутация затрагивает ген цитохром Р450 оксидазы (CYPgst) и содержится в локусе *gst* растения, подвергшегося мутации. В частности, при возникновении мутации у гомозиготных растений содержание или активность функционального белка CYPgst оказывается ниже, чем у соответствующего (мужски фертильного) растения, содержащего локус дикого типа (растение дикого типа), либо он вообще отсутствует. Обычно мутация гена CYPgst предотвращает транскрипцию и/или трансляцию функционального белка CYPgst.

Термин «тесно сцепленный» следует понимать, как означающий, что расстояние между двумя flankирующими локусами (например, двумя маркерами (маркерными локусами)) на генетической карте составляет менее 15 cM, менее 12 cM, менее 10 cM, менее 8 cM, менее 7 cM, менее 6 cM, менее 5 cM, менее 4 cM, менее 3 cM, менее 2 cM, менее 1 cM, менее 0,5 cM, менее 0,2 cM, менее 0,1 cM, менее 0,05 cM.

Термин «гибридизировать» или «гибридизация» означает процесс, при котором одноцепочечная молекула нуклеиновой кислоты соединяется с насколько возможно комплементарной цепочкой нуклеиновой кислоты, т.е. при этом образуются пары оснований. Стандартные методы гибридизации описаны, например, Sambrook *et al.*, Molecular Cloning; A Laboratory Manual 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001.

Предпочтительно это означает, что по меньшей мере 60%, более предпочтительно по меньшей мере 65%, 70%, 75%, 80% или 85%, наиболее предпочтительно 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% или 99% оснований молекулы нуклеиновой кислоты образуют пары оснований с насколько возможно комплементарной цепочкой нуклеиновой кислоты. Возможность такого соединения зависит от жесткости условий гибридизации. Термин «жесткость» относится к условиям гибридизации. Условия высокой жесткости означают условия, при которых образование пар оснований затруднено; условия низкой жесткости означают условия, при которых образование пар оснований проходит легче. Жесткость условий гибридизации зависит, например, от концентрации соли, ионной силы и температуры. Как правило, жесткость условий можно повышать путем повышения температуры и/или снижения содержания соли. Термин «жесткие условия гибридизации» означает условия, при которых гибридизация происходит только между гомологичными молекулами нуклеиновой кислоты и гомологичными генами. Термин «условия гибридизации», таким образом, относится не только к условиям, превалирующим непосредственно при соединении цепей нуклеиновых кислот, но также к условиям, превалирующим во время последующих этапов развития. Примерами жестких условий гибридизации являются условия, при которых, прежде всего, гибридизуются те молекулы нуклеиновой кислоты, которые обладают по меньшей мере 70%, предпочтительно по меньшей мере 75%, по меньшей мере 80%, по меньшей мере 85%, по меньшей мере 90% или по меньшей мере 95% идентичностью. К условиям жесткой гибридизации относится, например, гибридизация в 4 x SSC при 65°C и последующая многократная отмычка в 0,1 x SSC при 65°C приблизительно в течение 1 часа. Используемый здесь термин “условия жесткой гибридизации” может также означать гибридизацию при 68°C в 0,25 M фосфат натрия, pH 7,2, 7% SDS, 1 mM EDTA и 1% BSA в течение 16 часов и последующую двукратную отмычку в 2 x SSC и 0,1% SDS при 68°C. Предпочтительно, чтобы гибридизация осуществлялась в жестких условиях.

Термин «комплементарная» нуклеотидная последовательность применительно к нуклеиновой кислоте в форме двухцепочечной ДНК означает, что первая цепочка ДНК является комплементарной второй цепочке ДНК в отношении регулирования пар оснований нуклеотидов, соответствующих основаниям первой цепочки.

Термин «(молекулярный) маркер» означает нуклеотидную последовательность, которая служит в качестве референсной точки или точки ориентации. Маркер для детекции рекомбинации должен быть пригодным для мониторинга различий или полиморфизмов внутри растительной популяции. Что касается маркеров, эти различия встречаются на уровне ДНК и представляют собой различия в полинуклеотидных последовательностях, например, такие как SSR (*простые повторяющиеся последовательности*), RFLP (*полиморфизмы длин рестрикционных фрагментов*), FLP (*полиморфизмы длины фрагмента*) или SNP (*простые нуклеотидные полиморфизмы*). Маркеры могут происходить из геномных или экспрессированных нуклеиновых кислот, например, сплайсированной РНК, сДНК или EST-последовательностей, и могут также относиться к нуклеиновым кислотам, используемым в качестве зонда или пар праймеров, и, как таковые, быть пригодными для амплификации фрагмента последовательности с использованием методов на основе ПЦР. Маркеры, касающиеся генетических полиморфизмов между частями популяции, можно детектировать с помощью разработанных методов, известных из уровня техники (An Introduction to Genetic Analysis, 7th edition, Griffiths, Miller, Suzuki *et al.*, 2000). Эти методы

включают, например, секвенирование ДНК, амплификацию специфических последовательностей на основе ПЦР, детекцию RFLP, детекцию полинуклеотидных полиморфизмов с использованием аллель-специфической гибридизации (ASH) и детекцию SSR, SNP или RFLP. Помимо этого, известны методы детекции EST (экспрессирующиеся секвенированные последовательности) и RAPD (случайно амплифицированная полиморфная ДНК). В зависимости от контекста термин «маркер» может также означать специфическую позицию хромосомы в геноме вида, где может локализоваться специфический маркер (например, SNP). В настоящем изобретении маркеры также используются для детекции делеций.

Термин «культурное растение» охватывает как культивируемые, так и дикорастущие растения. Растения, называемые культурными растениями, используются, прямо или косвенно, в любой форме людьми, например, в качестве пищевых продуктов, стимуляторов или лекарственных средств, а также в качестве лесоматериалов или в качестве кормов для скота.

«Культивируемое растение» в отличие от дикорастущих растений представляет собой растение, которое люди, сажают, выращивают и защищают, и которое можно использовать в качестве культурного или декоративного растения. Генетической основой существования культивируемых растений являются точечные мутации, соматические мутации, хромосомные мутации и полиплоидизация. Эти мутации лежат в основе отбора. Они формируют природный или искусственно продленный (увеличение частоты мутаций, скрещивание, обработка колхицином, методы генной инженерии) начальный материал для эволюции, контролируемой человеком. Культивируемые растения включают пищевые растения, технические растения (например, прядильные растения), кормовые растения и декоративные растения. Важными признаками данных культивируемых растений являются увеличение размера растений и, в частности, используемых органов, потеря горечи, устойчивость к вредителям и/или более высокое содержание питательных веществ.

«Оперативно связанный» означает связанный в общей молекуле нуклеиновой кислоты таким образом, что связанные элементы расположены и ориентированы относительно друг друга так, что может происходить транскрипция молекулы нуклеиновой кислоты. ДНК, оперативно связанная с промотором, находится под транскрикционным контролем такого промотора.

«Растение» в контексте изобретения может быть растением любого вида, выбираемым из двудольных или однодольных растений. Предпочтительны растения, используемые в сельском хозяйстве, садоводстве и для получения биоэнергии (биоэтанола, биогаза и т.п.). Растения, используемые в настоящем изобретении, предпочтительно различаются своими запасающими органами, включающими клубни, корни, семена, зерна, плоды и т.п. К таким растениям относятся, например, вида *Zea mays*, *Solanum tuberosum*, *Triticum aestivum*, *Triticum durum*, *Triticum spelta*, *Helianthus annuus*, *Secale cereal*, *Hordeum vulgare*, *Hordeum bulbosum*, *Brassica napus*, *Brassica oleracea*, *Brassica rapa*, *Brassica juncea*, *Brassica nigra*, *Glycine max*, *Gossypium sp.*, *Sorghum bicolor*, *Triticale*, *Saccharum officinarum*, *Setaria italica*, *Oryza sativa*, *Oryza minuta*, *Oryza australiensis*, *Oryza alta*, *Brachypodium distachyon*, *Hordeum marinum*, *Aegilops tauschii*, *Daucus glochidiantus*, *Daucus pusillus*, *Daucus muricatus*, *Daucus carota*, *Eucalyptus grandis*, *Erythranthe guttata*, *Genlisea aurea*, *Musa sp.*, *Avena sp.*, *Nicotiana sylvestris*, *Nicotiana tabacum*, *Nicotiana tomentosiformis*, *Solanum lycopersicum*, *Coffea*

canephora, *Vitis vinifera*, *Cucumis sativus*, *Morus notabilis*, *Crucihibimalaya himalaica*, *Crucihibimalaya wallichii*, *Cardamine flexuosa*, *Lepidium virginicum*, *Capsella bursa-pastoris*, *Olimarabidopsis pumila*, *Arabis hirsute*, *Raphanus sativus*, *Eruca vesicaria sativa*, *Citrus sinensis*, *Jatropha curcas*, *Populus trichocarpa* или *Beta vulgaris*. Растение согласно изобретению предпочтительно является растением рода *Beta*, в частности, растением сахарная свекла (*Beta vulgaris*, подвида *vulgaris*).

Растительные «органы» означают, например, листья, растительные стебли, стволы, корни, вегетативные почки, меристемы, зародыши, пыльники, семязачатки, семена или плоды, в частности зерна. Термин «растительная часть» или “растительные части” включает, но не ограничивается этим, растительный стебель, листья, цветки, соцветия, корни, плоды и семена, а также пыльцу. Кроме того, растительные “части” также означают комбинацию нескольких органов, например, цветка и семени, либо часть органа, например, поперечный разрез стебля растения. Примерами растительных “тканей” являются каллусные ткани, запасающие ткани, меристематические ткани, ткани листьев, ткани побегов, ткани корней, ткани растительных центров или репродуктивные ткани, а также образовательные ткани, основные ткани (называемые паренхима), проводящие ткани, механические ткани и покровные ткани (называемые эпидермис). Этот перечень, однако, не является исчерпывающим. Термин растительные “клетки” следует понимать как означающий, например, выделенные клетки с наличием клеточной стенки, либо их агрегаты или протопласты.

В контексте настоящего изобретения термин «регуляторная последовательность» означает нуклеотидную последовательность, которая влияет на специфичность и/или уровень экспрессии, например, регуляторная последовательность может придавать специфическую тканеспецифичность. Такая регуляторная последовательность может локализоваться как вверх по течению от точки инициации транскрипции минимального промотора, так и вниз по течению от нее, например, как в лидерной последовательности, которая транскрибирована, но не транслирована, или внутри интрона.

«Промотор» - это нетранслируемый сегмент ДНК обычно вверх по течению от кодирующего участка, который содержит сайт связывания РНК-полимеразы и инициирует транскрипцию ДНК. Промотор также содержит другие элементы, которые функционируют в качестве регуляторов экспрессии генов (например, cis-регуляторные элементы). «Ядерный или минимальный промотор» - это промотор, который содержит по меньшей мере основные элементы, необходимые для инициации транскрипции (например, ТАТА-бокс или инициатор).

«Трансгенное растение» означает растение, в геном которого интегрирован по меньшей мере один полинуклеотид, предпочтительно гетерологичный полинуклеотид. Предпочтительно интеграция полинуклеотида является стабильной, что означает, что полинуклеотид остается стабильным в растении, экспрессируется и может также стабильно наследоваться потомками. Стабильное введение полинуклеотида в геном растения также включает его интеграцию в геноме растения предшествующего поколения, причем полинуклеотид может далее стабильно передаваться последующим поколениям. Термин «гетерологичный» означает, что введенный полинуклеотид происходит, например, от клетки или организма иного генетического фона того же вида или другого вида, либо является гомологичным по отношению к прокариотической или эукариотической клетке-хозяину, и в

таком случае локализуется в иной генетической среде и поэтому отличается от потенциально естественно встречающегося соответствующего полинуклеотида. Гетерологичный полинуклеотид может присутствовать в добавление к соответствующему эндогенному гену.

Настоящее изобретение далее иллюстрируется примерами вариантов его осуществления со ссылками на фигуры чертежей и последовательности:

Рис. 1 А, С) цветки фертильных растений сахарной свеклы (*Beta vulgaris*, подвид *vulgaris*) и **В, Д)** цветки мужски стерильных растений сахарной свеклы, фенотипы которых происходят от Донора C311 [2043_K5]. **А, В** – закрытые цветки, чашелистики и лепестки которых удалены вручную. Четко видны жизнеспособные пыльники фертильного генотипа (**А**), обозначенные светлым (желтым) цветом. Пыльники стерильных генотипов четко обозначены темным (коричневым) цветом. Во время стадии цветения пыльники фертильных генотипов раскрываются и выделяется пыльца (**С**), а пыльники стерильных генотипов не продолжают развиваться и не содержат пыльцы.

Рис. 2:Модель гена, аннотированного в RefBeet 1.2, тип BvCYPgst (g6845.tl), в референсном генотипе KWS2320. Белок длиной 517 аминокислот кодируется экзонами, общая длина которых составляет 1554 бр. Генотипы, выраженные в мужски стерильном фенотипе, содержат делецию протяженностью 533 бр, часть, включающую 5'-UTR и первый экзон гена. В таком случае правильная транскрипция мРНК и трансляция функционального белка является невозможной.

Рис. 3: Выравнивание фрагмента геномной ДНК протяженностью 4721 бр, кодирующей модель гена сахарной свеклы BvCYPgst (g6845.tl) из стерильных и фертильных генотипов. Последовательность стерильных генотипов содержит делецию протяженностью 533 бр.

Рис. 4: Анализ последовательности фрагмента геномной ДНК протяженностью 4721 бр, кодирующей модель гена сахарной свеклы BvCYPgst (g6845.tl) из стерильных и фертильных генотипов. **А)** – последовательность геномной ДНК гена CYPgst *Beta vulgaris*, подвид *vulgaris*, включающая участок предполагаемого промотора, а также 5'-UTR и 3'-UTR. Участок предполагаемого промотора отмечен «жирным», 5'-UTR и 3'-UTR подчеркнуты, экзон 1 отмечен «жирным» и подчеркнут, экзон 2 обозначен курсивом и подчеркнут, и инtron обозначен курсивом. Представленная последовательность соответствует последовательности SEQ ID No. 1. Функциональные участки гена располагаются следующим образом: предполагаемый промотор – 1...518; 5'-UTR – 1519...1761; транскрибированный участок – 1519...4275; экзон – 1762...2679; инtron – 2680...3506; экзон – 3507...4142; 3'-UTR – 4143...4275. **В)** – последовательность сДНК гена CYPgst *Beta vulgaris*, подвид *vulgaris*, включающая 5'-UTR и 3'-UTR. 5'-UTR и 3'-UTR подчеркнуты, экзон 1 отмечен «жирным» и подчеркнут, и экзон 2 обозначен курсивом и подчеркнут. Представленная последовательность соответствует последовательности SEQ ID No. 2. Функциональные участки сДНК располагаются следующим образом: 5'-UTR – 1...243; экзон – 244...1161; экзон – 1162...1787; 3'-UTR – 1798...1930. **С)** – аминокислотная последовательность гена CYPgst *Beta vulgaris*, подвид *vulgaris*. Представленная последовательность соответствует последовательности SEQ ID No. 3. **Д)** последовательность геномной ДНК мутировавшего гена CYPgst *Beta vulgaris*, подвид *vulgaris*, включающая участок предполагаемого промотора и 3'-UTR. Участок предполагаемого промотора отмечен «жирным», 3'-UTR подчеркнут,

процессированный экзон 1 отмечен «жирным» и подчеркнут, экзон 2 обозначен курсивом и подчеркнут, и инtron обозначен курсивом. Представленная последовательность соответствует последовательности SEQ ID No. 8. Функциональные участки мутировавшего гена CYPgst располагаются следующим образом: предполагаемый промотор – 1...1353; 5'-UTR – 1519...1761; транскрибированный участок – 1354...3542; процессированный экзон – 1354...1938; инtron – 1939...2755; экзон – 2756...3394; 3'-UTR = 3395...3542.

Рис. 5: Анализ экспрессии гена *BvCYPgst* (*GST*, g6845.tl) при помощи ПЦР в реальном времени. РНК получена из различных тканей фертильных растений. Экспрессия гена *GST* установлена путем сравнения с экспрессией гена g4645.tl; n.d. означает, что экспрессия не была детектирована. В представленном эксперименте *GST* более сильно экспрессировался в закрытых цветках. При сравнении экспрессия гена *GST* не была детектирована в закрытых цветках стерильных генотипов.

Подробное описание изобретения

Настоящее изобретение обеспечивает получение растения в результате мутации в сегменте ДНК ядерного генома, содержащего ген цитохром Р450 оксидазы (CYPgst), которое проявляет ядерно-кодируемый мужски стерильный фенотип. Оно отличается тем, что мутация вызывается экспрессией рецессивного признака. Получаемое таким образом растение можно использовать для упрощения трудоемких селекционных программ. Для идентификации гена, отвечающего за данную экспрессию признака, используется сахарная свекла (*Beta vulgaris*, подвид *vulgaris*), как показано в Примерах 1 и 2 и на Рис. 1-5. Упомянутый ген классифицируется как член семейства цитохром Р450 оксидаз (CYP) по его структурным признакам, выявленным на основе анализа последовательности, и имеет суффикс «gst», характеризующий фенотип, наблюдаемый в его мутантах с ядерной мужской стерильностью. Поскольку ген идентифицируется в сахарной свекле, то также используется префикс «*Bv*» в тех случаях, когда в примерах дается ссылка на конкретный ген.

В общем, настоящее изобретение касается растения, в частности культивируемого или культурного растения, проявляющего рецессивный ядерно-кодируемый мужски стерильный фенотип, которое отличается тем, что фенотип обусловливается мутацией гена эндогенной цитохром Р450 оксидазы (CYPgst) или отсутствием, пониженным содержанием или пониженной активностью функционального белка CYPgst, кодируемого геном дикого типа CYPgst, по сравнению с соответствующим (мужски фертильным) растением дикого типа, при этом не мутировавший ген CYPgst представляет собой ген *BvCYPgst Beta vulgaris*, который предпочтительно содержит одну из нуклеотидных последовательностей SEQ ID No. 1 или 2 или кодирует аминокислотную последовательность, представленную в SEQ ID No. 3, либо его гомолог, аналог или ортолог. Как показано выше и объяснено на примерах, у растений можно идентифицировать другие белки CYPgst или кодирующие их гены, т.е. гомологи, аналоги и ортологи, посредством традиционных подходов биоинформатики (поиски по базам данных и компьютерные программы для скрининга гомологичных последовательностей), при этом можно предположить, что инициируется мутация того же фенотипа, что и у растений сахарной свеклы. Растение по данному изобретению также отличается тем, что не мутировавший ген CYPgst представляет собой ген *StCYPgst Solanum tuberosum*, который предпочтительно содержит нуклеотидную последовательность SEQ ID

No. 12 или 13 или кодирует аминокислотную последовательность, представленную в SEQ ID No. 14, или его гомолог, аналог или ортолог, и не мутировавший ген CYPgst представляет собой ген *ZmCYPgst Zea mays*, который предпочтительно содержит нуклеотидную последовательность SEQ ID No. 9 или 10 или кодирует аминокислотную последовательность, представленную в SEQ ID No. 11, или его гомолог, аналог или ортолог.

В контексте изобретения термин «гомолог(и)» означает, что релевантные гены (из двух различных видов растений) по существу выполняют ту же самую функцию и имеют общего предка, и что их нуклеиновые кислоты и кодирующие аминокислотные последовательности по существу идентичны. Имеется, однако, много генов, гомологичных друг другу, однако выравнивание последовательностей белков не приводит к получению полезного спаривания. Напротив, термин «аналогичные» гены или белки указывает на то, что они выполняют идентичные или сходные функции, но имеют различную структуру, т.е. не имеют общего предка. В таком случае их нуклеиновые кислоты или кодирующие аминокислотные последовательности не совсем идентичны друг другу, либо, в лучшем случае, идентичны только в определенных функциональных доменах.

В контексте секвенирования генома для аннотации термин «гомолог» определяется более конкретно. Термины «ортолог» и «паралог» также сюда относятся. Ортологи – это гены, которые возникают в результате видеообразования. Паралоги – это гены, которые возникают в результате дупликации.

Согласно данному изобретению ген, как правило, представляет собой гомолог, аналог или ортолог, если он способен комплементировать мужски стерильный фенотип, который встречается в референсном гене CYPgst у растений сахарной свеклы (*BvCYPgst*) и/или обеспечивает инициацию направленной мутации в релевантном гене, либо вызывает изменения биологической активности генетических продуктов фенотипа у растения с мужской стерильностью, кодируемом гомологом или аналогом, от которого ген был получен. Соответственно, релевантный гомолог или аналог гена CYPgst по данному изобретению, показанный на этом примере, предпочтительно отличается тем, что он способен комплементировать мужски стерильный фенотип, наблюдаемый у мутантов CYPgst сахарной свеклы, т.е. он способен восстанавливать fertильность фенотипа. Дополнительно или как вариант, гомолог или аналог CYPgst предпочтительно отличается тем, что растение с мужски стерильным фенотипом получают путем ингибирования их экспрессии или биологической активности генетического продукта, кодируемого гомологом или аналогом. Мужски стерильный фенотип предпочтительно обладает признаками, характерными, например, для мутантов CYPgst у растений сахарной свеклы, в частности признаками, описанными в примерах; см. варианты осуществления изобретения выше.

Соответствующие методы и технологии комплементации в генетике известны специалисту в данной области; см., например, Napoli *et al.*, и Plant Physiology 120 (1999), 615-622, где описана мутация в инбредных штаммах петунии, которые также имеют мужски стерильный фенотип, который можно элиминировать посредством трансгенной комплементации с помощью функциональной халкон-сингазы A сДНК, что позволяет заключить, что ген A халкон-сингазы по существу ответственен за мужски стерильный фенотип или что фенотип с мужской стерильностью вызывается мутацией этого гена.

У Jeong *et al.*, J. Exp. Bot. 65 (2014), 6693-6709, мужская стерильность у так называемых томатов-мутантов *ms10*³⁵ комплементируется и элиминируется посредством комплементации и трансгенной экспрессии различных генов-кандидатов. Способы получения мужской стерильности у трансгенных растений путем ингибиования целевого гена, в данном случае *CYPgst*, также известны специалисту в данной области из предшествующего уровня техники, см., например, международную заявку WO 1996/017945 и следующие варианты осуществления изобретения.

Таким образом, согласно варианту осуществления настоящего изобретения растение, проявляющее рецессивный ядерно-кодируемый мужски стерильный фенотип, отличается тем, что данный фенотип обусловлен мутацией гена эндогенной цитохром Р450 оксидазы (*CYPgst*). Растение согласно изобретению также отличается отсутствием, пониженным содержанием или пониженной активностью функционального белка *CYPgst*, кодируемого геном дикого типа *CYPgst*, по сравнению с соответствующим мужски фертильным растением дикого типа. Геномная последовательность мутированного гена, которая больше не способна транслироваться, представлена в SEQ ID No. 8, но является только примером, и настоящее изобретение не ограничивается этим. В частности, изобретение касается растения, относящегося к культивируемым и культурным растениям.

Инактивация гена *CYP703 Arabidopsis thaliana* (*CYP703A2*), чтобы уменьшить количество образуемой пыльцы и, таким образом, вызвать частичную мужскую стерильность известна из предшествующего уровня техники (Morant *et al.*, The Plant Cell, 19 (2007), 1473-1487). Это можно объяснить тем, что спорополленин, являющийся основным компонентом внешних стенок пыльцы, отсутствует или структурно изменен. Хотя и представляется возможным, что ген *CYPgst* выполняет иную функцию, так как мутация вызывает рецессивную ядерно-кодируемую мужскую стерильность, и пыльца не образуется, вероятность того, что ген *CYPgst* и ген *CYP703 Arabidopsis thaliana* принадлежат к тому же самому семейству генов и, таким образом, выполняют ту же или по меньшей мере сходную функцию в синтезе спорополленина, нельзя исключать. Безотносительно какой-либо теории, можно предположить, что способность компенсировать потерю спорополленина утрачена у культивируемых растений, которые годами направленной селекции и скрещивания оптимизировались относительно урожайности, устойчивости к вредителям, устойчивости к абиотическим стрессовым факторам, а также содержания растительных веществ, и, таким образом, отсутствие спорополленина супрессирует образование пыльцы, что делает растение мужски стерильным.

Поскольку Morant *et al.* (2007) показали, что ген *CYP703A2* или соответствующие нокаут-линии *Arabidopsis* проявляют только частичную мужскую стерильность, и что такой фенотип не пригоден для гибридизации, то согласно варианту осуществления изобретения ген *CYP703A2 Arabidopsis thaliana* и мутанты, описанные Morant *et al.*, и соавт., особенно те, чьи последовательности представлены на Рис. 1, не являются предметом настоящего изобретения. Соответственно, согласно предпочтительному варианту осуществления изобретения культивируемые и/или культурные растения вида *Arabidopsis thaliana* предпочтительно исключены из объема настоящего изобретения.

Растения обладают двумя или более копиями генетической информации в каждой клетке в форме эукариот. Каждый ген обычно представлен двумя аллелями, которые могут быть идентичными в гомозиготном состоянии и различными в гетерозиготном состоянии.

Фенотип растения по данному изобретению обусловливается мутацией в ядерном геноме, которая вызывается экспрессией рецессивного признака. Таким образом, согласно варианту осуществления изобретения растение является мужски фертильным, если мутация проявляется в гетерозиготном состоянии, и мужски стерильным, если мутация проявляется в гомозиготном состоянии.

Образование функциональной пыльцы супрессируется, предпочтительно полностью супрессируется у стерильного растения. В контексте настоящего изобретения термин «супрессированный» означает, что у растения, являющегося гомозиготным по мутации гена CYPgst и мужски стерильным, 95%, предпочтительно 96%, более предпочтительно 97% и наиболее предпочтительно 98% или 99% пыльцы не образуется, термин «полностью супрессированный» означает, что образование пыльцы супрессируется на более 99%, предпочтительно 100%. В данном контексте «супрессированный» предпочтительно означает, что в опытах по скрещиванию такого растения, служащего в качестве мужского родителя, с соответствующим растением дикого типа по существу не образуются семена и/или растения-потомки.

Это четко видно на Рис. 1 в отношении *Beta vulgaris*, подвид *vulgaris*. У закрытых цветков, у которых чашелистики и лепестки удалены вручную, наблюдаются жизнеспособные пыльники фертильного генотипа (A), имеющие светлую (желтую) окраску. Наоборот, пыльники стерильных генотипов имеют темную (коричневую) окраску (B). Во время цветения пыльники фертильных генотипов раскрываются и выделяется пыльца (C), а пыльники стерильных генотипов не созревают и пыльца не выделяется (D).

У *Arabidopsis thaliana* белок CYP703 катализирует преобразование среднцепочечных насыщенных жирных кислот в простые гидроксилированные жирные кислоты предпочтительно в процессе гидроксилирования лауриновой кислоты в позиции С-7. Безотносительно какой-либо теории, представляется возможным, что белок CYPgst не выполняет ту же самую функцию, что и белок CYP703 *Arabidopsis thaliana*, но сходную функцию, потому что инактивация обоих генов влияет на образование пыльцы. Таким образом, можно констатировать, что CYPgst выполняет функцию синтеза спорополленина - основного компонента внешних стенок жизнеспособной пыльцы.

Согласно варианту осуществления изобретения белок CYPgst участвует в синтезе спорополленина и катализирует преобразование среднцепочечных насыщенных жирных кислот с образованием простых гидроксилированных жирных кислот предпочтительно в процессе гидроксилирования лауриновой кислоты в позиции С7.

Транскрипционный анализ (Пример 2) фертильных генотипов *Beta vulgaris*, подвид *vulgaris*, показывает, что ген CYPgst экспрессируется в закрытых цветках и плодах, но не экспрессируется в корнях и листьях (Рис. 5). Таким образом, согласно варианту осуществления изобретения растение по изобретению представляет собой растение, описанное выше, отличающееся тем, что ген CYPgst экспрессируется по меньшей мере в закрытых цветках и плодах, предпочтительно специфически в закрытых цветках и плодах.

Согласно варианту осуществления изобретения мутация предотвращает транскрипцию и/или трансляцию функционального белка у растения по данному изобретению, при этом мутация предпочтительно представляет собой делецию, присоединение, инсерцию или замещение в кодирующей нуклеотидной последовательности

гена CYPgst, последовательности сигнала для сплайсинга или регуляторной последовательности, предпочтительно в последовательности промотора гена CYPgst.

Согласно предпочтительному варианту осуществления изобретения делеция представляет собой удаление по меньшей мере 500-600 бр из кодирующего участка или промоторного участка гена CYPgst. Делеция может также иметь длину по меньшей мере в 20, 30 или 50 последовательных пар оснований, по меньшей мере в 100, 150, 200 или 250 последовательных пар оснований и предпочтительно по меньшей мере в 300, 400 или 500 последовательных пар оснований. Присоединение предпочтительно представляет собой добавление нуклеотида или нескольких нуклеотидов в геномную последовательность, предпочтительно в кодирующую последовательность гена, приводящее к сдвигу рамки считывания. Замещение предпочтительно представляет собой точечную мутацию в геномной последовательности, предпочтительно в кодирующей последовательности, в результате которой образуются стоп-кодоны или происходят ошибки сплайсинга.

Сравнительное секвенирование фрагментов геномной ДНК, содержащих ген CYPgst и предполагаемый участок промотора из мужски стерильных и мужски фертильных растений *Beta vulgaris*, подвид *vulgaris*, показало, что делеция 533 бр ответственна за мужски стерильный фенотип (см. Пример 2 и Рис. 3), и что участок делеции находится между нуклеотидными позициями 1560 и 2095 SEQ ID No. 1.

Согласно еще одному предпочтительному варианту осуществления изобретения делеция представляет собой удаление 533 бр и содержит части 5'-UTR и первый экзон гена CYPgst *Beta vulgaris*, подвид *vulgaris*; см. Рис. 3. Функциональный ген *BvCYPgst* содержит два экзона общей длиной 1554 бр. Модель гена, аннотированная в RefBeet 1.2, показана на Рис. 2, и последовательность геномной ДНК CYPgst с делецией, приводящей к образованию усеченного экзона 1, представлена в SEQ ID No. 8. Возможные точечные мутации, в результате которых может происходить преждевременное прекращение транскрипции гена CYPgst *Beta vulgaris*, подвид *vulgaris*, либо которые могут вызывать нарушение сплайсинга, перечислены в Таблице 1 и предпочтительно находятся между нуклеотидными позициями 1560 и 2095 SEQ ID No. 1.

Как показано на Примере 1 настоящего изобретения, тонкое картирование позволяет идентифицировать тесно сцепленные фланкирующие маркеры гена CYPgst и, таким образом, определить позицию гена CYPgst в геноме *Beta vulgaris*, подвид *vulgaris*. В свою очередь это создает основу для разработки маркеров мужского родителя, с помощью которых можно детектировать делецию гена CYPgst.

Согласно варианту осуществления настоящего изобретения растение отличается тем, что делецию у ратений сахарной свеклы (*Beta vulgaris*, подвид *vulgaris*) можно детектировать посредством определения отсутствия одного или обоих маркерных локусов sle5983d14 (продукт амплификации праймера, имеющего последовательности SEQ ID No. 4 и 5) и sle5983d17 (продукт амплификации праймера, имеющего последовательности SEQ ID No. 6 и 7) и присутствия убиквиторного маркера. Убиквиторный маркер подтверждает удовлетворительную экстракцию ДНК.

Согласно еще одному варианту осуществления изобретения ген *Beta vulgaris*, подвид *vulgaris* (сахарная свекла), локализуется в сегменте хромосомы 1 между маркерными локусами sxn215s01 и sle3305s02. В предпочтительном варианте осуществления изобретения

указанные маркерные локусы находятся на расстоянии 33,42 или 35, 15 сМ в хромосоме 1 (основана на генетической карте ZR INT 1202) и основываются на физической карте генома (Physmapv2). Этот участок имеет физический размер 215,4 kbp и располагается между позициями 3185718 bp и 3401120 bp. Были разработаны KASP-маркеры (KASP™ - технология SNP генотипирования от LGC Limited), при помощи которых можно идентифицировать SNP или соответствующую референсную последовательность, подлежащую детекции. Последовательность маркера sxn215s01, представленная в SEQ ID No. 24, и последовательность маркера sle3305s02, представленная в SEQ ID No. 26, указывают на присутствие локуса *gst*; последовательность маркера sxn215s01, представленная в SEQ ID No. 25, и последовательность маркера sle3305s02, представленная в SEQ ID No. 27, указывают на присутствие референсной последовательности, причем последовательности каждого из маркеров отличаются друг от друга в нуклеотидной позиции 21, и в генотипе, несущем локус *gst*, в этой позиции присутствует «G», а в референсном генотипе KWS2320 в этой позиции присутствует «A».

Согласно предпочтительному варианту осуществления изобретения длина сегмента составляет приблизительно от 50 до 5000 kbp, предпочтительно от 100 до 1000 kbp, более предпочтительно от 100 до 500 kbp и наиболее предпочтительно от 200 до 250 kbp, при этом сегмент содержит другие гены, кодирующие белки, предпочтительно 21 ген.

Согласно варианту осуществления настоящего изобретения не мутировавший ген представляет собой функциональный ген *BvCYPgst Beta vulgaris*, предпочтительно *Beta vulgaris*, подвид *vulgaris*, либо функциональный гомологичный, аналогичный или ортологичный ген другого культурного или культивируемого растения.

Специалист в данной области техники может получать информацию о других белках CYPgst из соответствующих публикаций, а также из баз данных, используя подходящие параметры поиска и компьютерные программы для скрининга гомологичных последовательностей или сравнения последовательностей. Кроме того, специалист в данной области техники может определить другие последовательности ДНК, кодирующие белок CYPgst, при помощи традиционных технологий молекулярной биологии и применять их в рамках настоящего изобретения. Таким образом из последовательности гена CYPgst могут быть получены, например, подходящие зонды для гибридизации и использованы для скрининга баз данных геномных ДНК и/или сДНК для поиска желаемого организма. Специалист в данной области техники может, например, получить сведения о стандартных методах гибридизации, клонирования и секвенирования из публикации Sambrook *et al.*, Molecular Cloning; A Laboratory Manual 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001. Специалист в данной области техники может также синтезировать и использовать олигонуклеотиды (праймеры) для амплификации последовательностей CYPgst на основе известных последовательностей.

В частности, согласно предпочтительному варианту осуществления изобретения гомологичный, аналогичный или ортологичный ген представляет собой ген *Zea mays*, который предпочтительно содержит нуклеотидную последовательность SEQ ID No. 9 или 10 или кодирует аминокислотную последовательность, представленную в SEQ ID No. 11, ген *Solanum tuberosum*, который предпочтительно содержит одну из нуклеотидных последовательностей SEQ ID No. 12 или 13 или кодирует аминокислотную последовательность, представленную в SEQ ID No. 14, ген *Triticum aestivum*, который

предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 15, ген *Helianthus annuus*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 16, ген *Hordeum vulgare*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 17, ген *Brassica napus*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 18, ген *Brassica oleracea*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 19, ген *Brassica rapa*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 20, ген *Glycine max*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 21, ген *Gossypium*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 22, а также ген *Sorghum bicolor*, который предпочтительно кодирует аминокислотную последовательность, представленную в SEQ ID No. 23. Указанные растения можно классифицировать как культурные растения и предпочтительно как культивируемые растения.

Согласно варианту осуществления растение по данному изобретению является растением, описанным выше, которое отличается тем, что не мутировавший ген (ген дикого типа) имеет одну из нуклеотидных последовательностей, выбираемых из группы, состоящей из нуклеотидных последовательностей, представленных в SEQ ID No. 1, 2, 9, 10, 12 и 13.

Согласно варианту осуществления изобретения не мутировавший ген (ген дикого типа) имеет нуклеотидную последовательность, которая кодирует аминокислотную последовательность, представленную в SEQ ID No. 3, 11 или 14.

Нуклеотидную последовательность можно вводить в ген, используя традиционные методы, известные из предшествующего уровня техники, например, посредством сайт-направленного мутагенеза, ПЦР-индуцированного мутагенеза, транспозонного мутагенеза, геномного редактирования и т.п., замещений, делеций, инсерций, присоединений и/или любой иной модификации, как в отдельности, так и в любой их комбинации, которые модифицируют нуклеотидную последовательность с сохранением той же функции стартовой последовательности.

Изобретение также охватывает растение, описанное выше, отличающееся тем, что нуклеотидная последовательность может содержать функциональный фрагмент нуклеотидных последовательностей SEQ ID No. 1, 2, 9, 10, 12 и 13. Термин «фрагмент» включает гены, нуклеотидную последовательность которых в значительной степени сходная с вышеупомянутой нуклеотидной последовательностью. Термин «в значительной степени сходный» означает, что последовательность первого нуклеотида или первой аминокислоты содержит достаточное или минимальное количество нуклеотидов или аминокислотных остатков, идентичных или эквивалентных последовательности второго нуклеотида или второй аминокислоты. Что касается аминокислотных последовательностей, то после модификации с помощью одного из вышеуказанных методов они также имеют общую доменную структуру и/или общую функциональную активность. В значительной степени сходными являются нуклеотидные последовательности или аминокислотные последовательности, которые обладают по меньшей мере около 45%, по меньшей мере около 50%, по меньшей мере около 55%, по меньшей мере около 60%, по меньшей мере около 65%, по меньшей мере около 70%, по меньшей мере около 75%, по меньшей мере около

80%, по меньшей мере около 85%, по меньшей мере около 90%, по меньшей мере около 95%, по меньшей мере около 96%, по меньшей мере около 97%, по меньшей мере около 98%, по меньшей мере около 99% или по меньшей мере около 100% идентичностью. Функциональные фрагменты предпочтительно являются в значительной степени сходными, если они обладают теми же общими свойствами, что и вышеупомянутые нуклеотидные или аминокислотные последовательности по данному изобретению.

Согласно варианту осуществления изобретения не мутировавший ген (ген дикого типа), содержащийся в растении, имеет нуклеотидную последовательность, способную гибридизоваться в жестких условиях с нуклеотидной последовательностью, которая комплементарна одной из нуклеотидных последовательностей SEQ ID No. 1, 2, 9, 10, 12 и 13, либо которая кодирует аминокислотную последовательность, представленную в SEQ ID No. 3, 11 или 14. Согласно другому варианту осуществления изобретения не мутировавший ген (ген дикого типа) содержит нуклеотидную последовательность, которая кодирует аминокислотную последовательность, имеющую различия по сравнению с аминокислотной последовательностью, представленной в SEQ ID No. 3, 11 или 14, в форме аминокислотных делеций, замещений, присоединений и/или инсерций в аминокислотной последовательности, предпочтительно не превышающие 20%, 15%, 10%, 5%, 4%, 3%, 2% или 1% относительно всей длины аминокислотной последовательности.

Согласно еще одному или дополнительному варианту осуществления изобретения нуклеотидная последовательность не мутировавшего гена (гена дикого типа) кодирует белок, обладающий той же ферментативной активностью, что и белок, кодируемый ДНК по предыдущему варианту осуществления.

Согласно еще одному варианту осуществления изобретения не мутировавший ген (ген дикого типа), содержащийся в растении, имеет ДНК, содержащую по меньшей мере 200 или 400, предпочтительно по меньшей мере 600 или 800, наиболее предпочтительно по меньшей мере 1000 последовательных нуклеотидов из промотора нуклеотидной последовательности SEQ ID No. 1 с нуклеотидными позициями от 1 до 1518, предпочтительно от 518 до 1518, наиболее предпочтительно от 1318 до 1518, либо последовательности, являющейся гибридной в данном участке, причем нуклеотидная последовательность способна контролировать экспрессию гена или гетерологичной молекулы нукleinовой кислоты, оперативно связанной с ДНК, особенно в закрытых цветках и/или плодах.

Согласно варианту осуществления изобретения растение может быть инбредным или гибридным растением. Инбредное растение можно использовать в качестве родительского растения для получения гибридов. Преимуществом использования рецессивного ядерно-кодируемого мужски стерильного гетерозиготного инбредного растения в отношении признака является то, что оно расщепляется на фертильные и стерильные индивиды на каждом репродуктивном этапе. Индивид с мужской стерильностью можно использовать для получения гибридов, при этом исключается ручное удаление пыльников и не требуется иметь линию-закрепитель стерильности.

Согласно варианту осуществления изобретения растение по данному изобретению является растением рода *Zea*, *Solanum*, *Triticum*, *Triticale*, *Helianthus*, *Secale*, *Hordeum*, *Brassica*, *Brachypodium*, *Glycine*, *Gossypium*, *Sorghum*, *Saccharum*, *Setaria*, *Aegilops*, *Oryza*, *Daucus*, *Eucalyptus*, *Erythranthe*, *Genlisea*, *Musa*, *Avena*, *Nicotiana*, *Coffea*, *Vitis*, *Cucumis*,

Morus, Crucihimalaya, Cardamine, Lepidium, Capsella, Olimarabidopsis, Arabis, Raphanus, Eruca, Citrus, Jatropha, Populus или *Beta*, предпочтительно растение вида *Zea mays*, *Solanum tuberosum*, *Triticum aestivum*, *Triticum durum*, *Triticum spelta*, *Helianthus annuus*, *Secale cereal*, *Hordeum vulgare*, *Hordeum bulbosum*, *Brassica napus*, *Brassica oleracia*, *Brassica rapa*, *Brassica juncea*, *Brassica nigra*, *Glycine max*, *Gossypium* sp., *Sorghum bicolor*, *Triticale*, *Saccharum officinarum*, *Setaria italica*, *Oryza sativa*, *Oryza minuta*, *Oryza australiensis*, *Oryza alta*, *Brachypodium distachyon*, *Hordeum marinum*, *Aegilops tauschii*, *Daucus glochidiantus*, *Daucus pusillus*, *Daucus muricatus*, *Daucus carota*, *Eucalyptus grandis*, *Erythranthe guttate*, *Genlisea aurea*, *Musa* sp., *Avena* sp., *Nicotiana sylvestris*, *Nicotiana tabacum*, *Nicotiana tomentosiformis*, *Solanum lycopersicum*, *Coffea canephora*, *Vitis vinifera*, *Cucumis sativus*, *Morus notabilis*, *Crucihimalaya himalaica*, *Crucihimalaya wallichii*, *Cardamine flexuosa*, *Lepidium virginicum*, *Capsella bursa-pastoris*, *Olimarabidopsis pumila*, *Arabis hirsute*, *Raphanus sativus*, *Eruca vesicaria sativa*, *Citrus sinensis*, *Jatropha curcas*, *Populus trichocarpa* или *Beta vulgaris*. Эти растения относятся к культурным растениям, в частности культивируемым растениям.

Согласно еще одному варианту осуществления настоящего изобретение касается не только растение по данному изобретению, имеющему мутацию в гене CYPgst, но также молекулы ДНК, содержащей нуклеотидную последовательность, как показано выше, имеющую мутацию в форме делеции, присоединения, инсерции или замещения, при этом мутация не приводит к синтезу функционального белка CYPgst.

Согласно предпочтительному варианту осуществления изобретения мутация в кодирующей нуклеотидной последовательности гена CYPgst представляет собой мутацию в последовательности сигнала для сплайсинга или регуляторной последовательности гена CYPgst, предпочтительно находящуюся в промоторе гена CYPgst. Мутация может быть делецией между нуклеотидными позициями 1560 и 2095 SEQ ID No. 1 или между соответствующими позициями в SEQ ID No. 12 или 9. Делеция может иметь длину по меньшей мере в 20, 30 или 50 последовательных пар оснований, предпочтительно по меньшей мере в 100, 150, 200 или 250 последовательных пар оснований и наиболее предпочтительно по меньшей мере в 300, 400 или 500 последовательных пар оснований. Согласно наиболее предпочтительному варианту осуществления изобретения молекула нукleinовой кислоты содержит нуклеотидную последовательность, соответствующую последовательности, представленной в SEQ ID No. 8. Согласно еще одному предпочтительному варианту осуществления изобретения молекула нукleinовой кислоты содержит точечную мутацию в нуклеотидной последовательности SEQ ID No. 1 согласно Таблице 1, предпочтительно между нуклеотидными позициями 1560 и 2095 SEQ ID No. 1.

Как показано выше, зонды для гибридизации ДНК, получаемые из последовательности гена CYPgst, можно использовать для скрининга баз данных геномных и/или сДНК для поиска других организмов с целью идентификации гомологичных генов. Для осуществления специфической гибридизации такие зонды должны быть специфическими и иметь по меньшей мере 15 нуклеотидов в длину, предпочтительно по меньшей мере 20 нуклеотидов. Зонды можно использовать для амплификации идентифицированных гомологичных генов посредством процесса, известного как полимеразная цепная реакция (ПЦР). Кроме того, данные зонды можно также использовать для детекции мутаций в гене CYPgst. Подробные инструкции по гибридизации нукleinовых кислот можно найти у Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology – Hybridization with Nucleic Acid Probes, part 1, Chapter 2, “Overview of Principles of

Hybridization and the Strategy of Nucleic Acid Probe Assays," Elsevier, New York (1993); и в Protocols in Molecular Biology, Chapter 2, Ausubel *et al.*, eds., Greene Publishing and Wiley Interscience, New York (1995).

Таким образом, предметом настоящего изобретения является молекула нуклеиновой кислоты, состоящая по меньшей мере из 15, 16, 17, 18, 19 или 20, предпочтительно по меньшей мере из 21, 22, 23, 24 или 25, более предпочтительно по меньшей мере из 30, 35, 40, 45 или 50 и наиболее предпочтительно по меньшей мере из 100, 200, 300, 500 или 1000 нуклеотидов, при этом данная молекула нуклеиновой кислоты специфически гибридизуется с нуклеотидной последовательностью, описанной выше, которая содержит не мутировавший ген CYPgst дикого типа, либо с молекулой ДНК, описанной выше, которая содержит мутацию в форме делеции, присоединения, инсерции или замещения, в результате которой не образуется функциональный белок CYP703. Молекула нуклеиновой кислоты предпочтительно соответствует варианту осуществления изобретения, представленному в параграфе [17].

Позицию гена CYPgst в геноме *Beta vulgaris*, подвид *vulgaris*, можно определять при помощи тонкого картирования, описанного выше. В свою очередь это дает основу для разработки генетических маркеров, посредством которых можно детектировать делецию гена CYPgst.

Таким образом, помимо растений, описанных выше, настоящее изобретение также касается маркеров в форме олигонуклеотидов, в частности олигонуклеотидных праймеров. Они содержат молекулу нуклеиновой кислоты размером по меньшей мере 15 нуклеотидов, специфически гибридизированную с нуклеотидной последовательностью, как описано выше, либо с молекулой ДНК, описанной выше, имеющую мутацию в форме делеции, присоединения, инсерции или замещения, в результате которой не образуется функциональный белок CYPgst. Указанные олигонуклеотиды предпочтительно имеют не более 50 нуклеотидов в длину. Более предпочтительны олигонуклеотиды, которые являются даже более короткими и имеют от 15 до 25 нуклеотидов в длину. Как показано в Примере 2 настоящего изобретения, олигонуклеотиды предпочтительно имеют одну из следующих нуклеотидных последовательностей: (i) SEQ ID No. 4, 6 или их комплемент, либо (ii) SEQ ID No. 5, 7 или их комплемент.

Предметом настоящего изобретения также является белок CYPgst, способный кодироваться нуклеотидной последовательностью, описанной выше, и его функциональный и/или иммунологически активный фрагмент, а также антитело, которое специфически связывается с описываемым здесь белком или его фрагментом. Получение рекомбинантных белков и фрагментов известно специалисту в данной области техники; см., например, Sambrook *et al.*, Molecular Cloning; A Laboratory Manual, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001, или Wingfield, P.T. 2008, Production of Recombinant Proteins, Current Protocols in Protein Science, 52:5.0:a5.0.1-5.0.5. Поликлональные или моноклональные антитела к белку по данному изобретению могут быть получены специалистом в данной области техники согласно методам, известным у E. Harlow *et al.*, Antibodies: A Laboratory Manual (1988). Моноклональные антитела и фрагменты Fab и F(ab')₂, которые можно также использовать для детекции белка, можно получать с помощью различных традиционных методов, таких как у Goding, Monoclonal Antibodies: Principles and Practice, pp. 98-118, New York; Academic Press (1983). Антитела можно затем использовать

для иммунологического скрининга экспрессионных библиотек сДНК с целью идентификации идентичных, гомологичных или гетерологичных генов (Sambrook *et al.*, Molecular Cloning; A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001, или Ausubel *et al.*, 1994, "Current Protocols in Molecular Biology", John Wiley & Sons).

Согласно предпочтительному варианту осуществления изобретения белок CYPgst отноится к аминокислотным последовательностям, представленным в SEQ ID No. 3, 11или 14-23, либо аминокислотной последовательности, которая по меньшей мере на 80%, 82%, 84%, 86% или 88%, предпочтительно на 90%, 91%, 92%, 93%, 94% или 95%, наиболее предпочтительно на 96%, 97%, 98%, 99% или 99,5% идентична аминокислотной последовательности, представленной в SEQ ID No. 3, предпочтительно по всей ее длине.

Изобретение также касается рекомбинантной молекулы ДНК, которая содержит не мутировавший ген CYPgst (ген дикого типа) и обладает вышеуказанными свойствами нуклеотидной последовательности, содержащейся в растении по данному изобретению. Рекомбинантная молекула ДНК предпочтительно содержит промотор и/или иной элемент управления транскрипцией или трансляцией, либо связана с ним. Используемые промоторы являются, в основном, промоторами со специфической активностью в клетках, обеспечивающими транскрипцию ДНК только в заранее определенных клетках. Кроме промоторов, имеются, но не ограничиваются ими, другие элементы управления транскрипцией, например, энхансеры, операторы, репрессоры, а также сигналы терминации транскрипции, которые функционально связаны с ДНК для обеспечения направленной клеточноспецифической транскрипции. Промоторы и другие элементы регуляции транскрипции в общем известны и доступны специалисту в данной области техники; см., например, WO 00/75359, строчки 5-17, стр. 24. Данную рекомбинантную молекулу ДНК можно использовать для восстановления фертильности у растений, являющихся рецессивными ядерно-кодируемыми мужской стерильными по фенотипу.

Поскольку, как показано выше, ген CYPgst экспрессируется в закрытых цветках и плодах и не экспрессируется в корнях или листьях, то в предпочтительном варианте осуществления изобретения рекомбинантная молекула ДНК содержит промотор, который имеет нуклеотидную последовательность, описанную выше, и оперативно связана с гетерологичной нуклеотидной последовательностью, или кодирующую нуклеотидную последовательность, как описано выше, которая содержит ген CYPgst дикого типа и оперативно связана с гетерологичным промотором. Данный промотор предпочтительно способен контролировать экспрессию нуклеотидной последовательности, особенно в закрытых цветках и/или плодах. Рекомбинантная молекула ДНК более предпочтительно содержит естественный промотор не мутировавшего гена CYPgst *Beta vulgaris*, подвид *vulgaris* (SEQ ID No. 1).

Соответственно, настоящее изобретение также касается использования описанных здесь молекулы ДНК или промотора для специфической экспрессии гетерологичных молекул нукleinовой кислоты в цветках и/или плодах растения. Для этого требуется, чтобы гетерологичная молекула нукleinовой кислоты была оперативно связана с соответствующим промотором, и чтобы такая рекомбинантная молекула ДНК была введена в целевую клетку, предпочтительно растительную клетку. Методы гетерологичной экспрессии рекомбинантных молекул ДНК будут рассмотрены более подробно ниже.

Еще одним предметом настоящего изобретения является рекомбинантная молекула ДНК, имеющая нуклеотидную последовательность по данному изобретению, которая кодирует shРНК (*малую РНК, образующую шпильки*), siРНК (*малую интерферирующую РНК*), отрицательно-полярную РНК, положительно-полярную РНК или двухцепочечную РНК. Они ингибируют трансляцию мРНК гена CYPgst или разрушают мРНК гена CYPgst в клетках посредством пар оснований. Соответственно, введение и/или экспрессия рекомбинантной молекулы ДНК в растении приводит к ингибированию экспрессии функционального (не мутировавшего) гена CYPgst. В предпочтительном варианте осуществления изобретения нуклеотидная последовательность проявляет свойства, подробно описанные в параграфе [15].

Еще одним предметом настоящего изобретения являются векторы, содержащие рекомбинантные молекулы ДНК, последовательности нуклеиновой кислоты или молекулы нуклеиновой кислоты по данному изобретению. Вектор по данному изобретению может содержать не мутировавший ген CYPgst (ген дикого типа) с вышеупомянутыми свойствами нуклеотидной последовательности и предпочтительно один из описанных выше промоторов. Другой вектор может содержать рекомбинантную молекулу ДНК, которая содержит не мутировавший ген CYPgst дикого типа, связанный с гетерологичной молекулой нуклеиновой кислоты, или рекомбинантную молекулу ДНК, которая содержит опианную выше нуклеотидную последовательность, оперативно связанную с гетерологичным промотором, причем в любом случае промотор предпочтительно способен специфически контролировать экспрессию нуклеотидной последовательности в закрытых цветках и/или плодах. Помимо этого, вектор может содержать рекомбинантную молекулу ДНК, имеющую нуклеотидную последовательность, которая кодирует shРНК, siРНК, отрицательно-полярную РНК, положительно-полярную РНК или двухцепочечную РНК, что приводит к ингибированию экспрессии гена CYPgst после экспрессии в растительной клетке.

Более того, вектор может содержать молекулу ДНК, имеющую одну из описанных выше мутаций или описанную выше молекулу нуклеиновой кислоты, которая специфически связывается с нуклеотидной последовательностью не мутировавшего гена CYPgst (дикого типа) или с нуклеотидной последовательностью мутировавшего гена CYPgst.

Описанный вектор может представлять собой плазмиду, космиду, фаг или экспрессионный вектор, трансформационный вектор, членочный вектор или клонирующий вектор; он может быть одно- или двухцепочечным, линейным или кольцевым, и он способен трансформировать прокариотического или эукариотического хозяина посредством интеграции в его геном или экстрахромосомно. Молекула ДНК или молекула нуклеиновой по данному изобретению оперативно связана с одной или несколькими регуляторными последовательностями в экспрессионном векторе, которые обеспечивают транскрипцию и, как вариант, экспрессию в прокариотической или эукариотической клетке-хозяине; см., например, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 2001 и международную патентную заявку WO 00/75359, стр. 21, строчка 20, стр. 22, строчка 32. Данные регуляторные последовательности предпочтительно представляют собой промоторы или терминаторы, в частности относительно стартовой точки для инициации транскрипции, сайт связывания рибосомы, сигнал процессинга РНК, сайт терминации транскрипции и/или сигнал к полиаденилированию. Как правило, векторы также содержат индикаторные/репортерные гены или гены устойчивости для детекции переноса желаемого вектора, а также молекулы

ДНК/молекулы нуклеиновой кислоты для отбора содержащих их индивидов, поскольку прямая детекция посредством экспрессии гена обычно весьма затруднена. Примерами индикаторных/репортерных генов являются ген люциферазы и ген, кодирующий зеленый флуоресцентный белок (GFP). Они также позволяют тестировать активность и/или регуляцию промотора гена. Примерами генов устойчивости, особенно для трансформации растений, являются ген неомицин-фосфотрансферазы, ген гигромицин-фосфотрансферазы или ген, который кодирует фосфинотрицин ацетилтрансферазу. Это не исключает другие индикаторные/репортерные гены или гены устойчивости, известные специалисту в данной области техники. В предпочтительном варианте осуществления изобретения вектор является растительным вектором.

Еще одним объектом изобретения является описанный выше вектор, отличающийся тем, что молекула ДНК в качестве трансгена способна экспрессировать функциональный ген CYPgst и предпочтительно генетически связана с другим трансгеном, который предотвращает перенос молекулы ДНК через пыльцу. Фертильность можно восстанавливать путем введения данного вектора в мутант, являющийся мужским стерильным вследствие мутации в гене CYPgst. Поскольку перенос трансгенного функционального CYPgst через пыльцу невозможен, то образуются только гемизиготные семена путем самоопыления трансгенной линии.

Вектор или трансген, соответственно, предпочтительно также содержит экспрессионную кассету, маркирующую семена, предпочтительно при помощи флуоресцентных меток. Используя данный метод, можно легко дифференцировать трансгенные и нетрансгенные семена. Эта система использования ядерно-кодируемой мужской стерильности была разработана компанией Pioneer. Система под названием SEED PRODUCTION TECHNOLOGY (SPT) (US 2006288440 A1), разработанная для кукурузы, основывается на том факте, что стерильный мутант, имеющий мутацию в известном ядерно-кодируемом гене, может быть восстановлен путем введения трансгена. Трансген содержит не мутировавший аллель гена стерильности, благодаря чему трансген выполняет функцию аллеля дикого типа. Трансген восстановления фертильности генетически связан с другим трансгеном, который предотвращает перенос трансгена через пыльцу (убийца пыльцы). В результате образуются только гемизиготные семена путем самоопыления трансгенной линии, что играет основную роль в обеспечении эффективности системы.

Трансген также содержит экспрессионную кассету, маркирующую семена с использованием красного флуоресцентного красителя. В результате можно легко дифференцировать трансгенные и нетрансгенные семена. Поскольку трансгенные семена являются фертильными, растения автоматически разделяются на фертильные и стерильные растения. Материнские формы растений, необходимые для создания гибридов, получают путем высеваания нетрансгенных семян, а трансгенные семена можно использовать в качестве отцовской линии для последующей репродукции материнской линии (линии-закрепителя), а также репродуцировать путем простого самовоспроизводства (см. Рис. 5). Теоретически, систему SPT можно использовать в отношении всех видов растений. Для этого требуется наличие линии с генетической мужской стерильностью и сведения о гене, который ответственен за фенотип с генетической мужской стерильностью. В результате, теоретически, систему SPT можно также использовать для разработки системы гибридизации в отношении каждого вида культивируемого растения, например, такого как сахарная свекла или картофель.

Настоящее изобретение также касается клеток-хозяев, содержащих описанные векторы, рекомбинантные молекулы ДНК и/или молекулы нуклеиновой кислоты. Клетка-хозяин по данному изобретению может быть прокариотической (например, бактериальной) или эукариотической клеткой (например, растительной клеткой или дрожжевой клеткой). Предпочтительно клетка-хозяин представляет собой агробактерию или растительную клетку. Настоящее изобретение предпочтительно относится к трансгенной растительной клетке, содержащей молекулу нуклеиновой кислоты по данному изобретению в качестве трансгена или вектор по данному изобретению. Такая трансгенная растительная клетка представляет собой, например, растительную клетку, трансформированную, предпочтительно стабильно трансформированную, молекулой нуклеиновой кислоты по данному изобретению. В предпочтительным варианте трансгенной растительной клетки молекула нуклеиновой кислоты оперативно связана с одной или несколькими регуляторными последовательностями, которые обеспечивают транскрипцию и, как вариант, экспрессию в растительной клетке. Общая структура, включающая молекулу нуклеиновой кислоты и регуляторные последовательности, представляет собой, таким образом, трансген. Как пример, такие регуляторные последовательности могут содержать промотор или терминатор. Специалисту в данной области техники известны многочисленные функциональные промоторы и терминалы, которые могут использоваться в растениях.

Изобретение также относится к идентификации гена CYPgst, ответственного за экспрессию признака рецессивной ядерно-кодируемой мужской стерильности, для получения трансгенных растений с такой экспрессией признака и, таким образом, восстановления фертильности.

Согласно варианту осуществления изобретения обеспечивается набор, содержащий описанные выше рекомбинантные молекулы ДНК или молекулы нуклеиновой кислоты и векторы, соответственно, которые необходимы для получения растений с рецессивной ядерно-кодируемой мужской стерильностью, а также для восстановления фертильности растений, имеющих этот фенотип. Данный набор также содержит рекомбинантную молекулу ДНК, содержащую промотор с заранее определенной нуклеотидной последовательностью или гетерологичный промотор, при этом первый связан с гетерологичной молекулой нуклеиновой кислоты, а второй связан с заранее определенной нуклеотидной последовательностью, кодирующей не мутировавший (дикого типа) ген CYPgst. Кроме того, набор может содержать вектор, причем молекула ДНК в качестве трансгена способна экспрессировать функциональный ген CYPgst и предпочтительно связана с другим трансгеном, который предотвращает перенос молекулы ДНК через пыльцу, а также экспрессионную кассету, маркирующую семена. Набор может содержать заранее определенную молекулу нуклеиновой кислоты для идентификации мутации в гене CYPgst, которая гибридизуется с заранее определенной нуклеотидной последовательностью, содержащей не мутировавший (дикого типа) ген CYPgst, или с соответствующим геном с заранее определенной мутацией, либо описанный выше олигонуклеотид. Набор может также содержать описанный выше ген CYPgst или его фрагмент, а также описанные выше антитела. Предпочтительно, чтобы набор также содержал реагенты предназначенные для процедур, основанных на детекции нуклеиновых кислот, или для иммунологической детекции.

Трансгенное растение является, например, растением, которое содержит растительные клетки, трансформированные молекулой ДНК/молекулой нуклеиновой кислоты по данному

изобретению или вектором по данному изобретению. В предпочтительном варианте трансгенного растения молекула ДНК/молекула нуклеиновой кислоты оперативно связана с одной или несколькими регуляторными последовательностями, обеспечивающими транскрипцию и, как вариант, экспрессию в растении. Общая структура, включающая молекулу нуклеиновой кислоты по данному изобретению и регуляторные последовательности представляет собой, таким образом, трансген. Термин «трансген» также означает рекомбинантную последовательность нуклеиновой кислоты, кодирующую полипептид.

Олигонуклеотиды, нуклеиновые кислоты, молекулы ДНК и векторы, которые были описаны выше, можно также использовать для получения трансгенного растения. Таким образом, настоящее изобретение также касается использования данных олигонуклеотидов, нуклеиновых кислот, молекул ДНК и векторов для получения трансгенного растения, проявляющего гомозиготный рецессивный ядерно-кодируемый мужски стерильный фенотип, отличающегося тем, что во время получения растения с восстановленной fertильностью по данному изобретению или трансгенной клетки-хозяина, предпочтительно растительной клетки, экспрессия гена CYPgst ингибируется. Кроме того, описанные выше олигонуклеотиды, нуклеиновые кислоты, молекулы ДНК и векторы можно также использовать в соответствующих способах получения таких трансгенных растений или растительных клеток. Трансгенное растение предпочтительно представляет собой культурное растение или более предпочтительно культивируемое растение.

Из предшествующего уровня техники известны различные способы получения, идентификации или отбора трансгенных растений, при которых супрессируется транскрипция/трансляция белка или вводится признак восстановления. Специалисту в данной области техники известны способы получения трансгенных культивируемых растений и их идентификации с использованием методов молекулярной биологии; относительно трансгенных растений сахарной свеклы, устойчивых к глифосату, см., например, международные патентные заявки WO 99/023232 и WO2004/074492, или относительно общей трансформации растений – WO2000/018939 и WO2013/138309.

Соответственно, согласно варианту осуществления изобретения обеспечен способ получения растения, проявляющего гомозиготный рецессивный ядерно-кодируемый мужски стерильный фенотип, отличающийся тем, что экспрессия гена CYPgst ингибируется, причем данное растение предпочтительно является культурным растением и более предпочтительно культивируемым растением. Рекомбинантная молекула ДНК, экспрессирующая полинуклеотид, вводится в растительную клетку путем трансформации, например, с использованием вектора, в результате чего экспрессия полинуклеотида приводит к ингибированию белка CYPgst.

Как пример, описанная выше мутация, приводящая к ингибированию экспрессии CYPgst, может вызываться посредством генетической рекомбинации в процессе скрещивания растений между собой, при этом одно из растений несет мутировавший аллель CYPgst.

Помимо использования традиционных селекционных программ для генерирования генетической рекомбинации, современная биотехнология предоставляет специалисту в данной области техники много других инструментов точной генетической инженерии. Как

пример, Т-ДНК мечение можно использовать для разрушения гена CYPgst посредством инсерционного мутагенеза. Кроме того, весь ген CYPgst или его часть можно делетировать посредством генетической мутации с использованием TALE-нуклеаз (TALEN) или нуклеаз «цинковые пальцы» (ZFN) и систем CRISPR/Cas, которые описаны в WO2014/144155 A1 (Engineering Plant Genomes Using CRISPR/Cas Systems) и у Osakabe & Osakabe, Plant Cell Physiol., 56 (2015), 389-400, так, что предотвращается экспрессия гена CYPgst. Это может также достигаться путем использования метода, называемого TILLING (*целенаправленные индуцированные локальные нарушения в геномах*), при котором в гене дикого типа вызывается точечная мутация, как показано в немецкой патентной заявке DE 10 2013 101 617, и затем отбираются растения, проявляющие подходящий, т.е. обеспечивающий устойчивость, ген, например, ячмень, который, проявляет устойчивость к вирусу желтой мозаики; см., например, DE 10 2013 101 617, стр. 4, 8 и 12 в параграфах [0014], [0026] и [0038]. Метод TILLING также подробно рассматривается в публикации Henikoff *et al.* (Henikoff *et al.*, Plant Physiol., 135, 2004, 630-636). Точечные мутации в гене CYPgst Beta vulgaris, подвид vulgaris, способные приводить к стоп-кодонам или ошибкам сплайсинга, приводятся в Таблице 1.

Ингибирование экспрессии также возможно с помощью подходов iРНК и косупрессии. Они заключаются во введении рекомбинантной молекулы ДНК или молекулы нуклеиновой кислоты, описанных выше, или соответствующего вектора в растение, при этом в результате экспрессия кодируемых молекул shРНК, отрицательно-полярной РНК или положительно-полярной РНК приводит к ингибированию экспрессии гена CYPgst. Такие методы, основанные на iРНК и косупрессии, являются стандартными методами для ингибирования экспрессии генов и известны специалисту в данной области техники. Подход положительной полярности, включающий использование молекулы мишень-специфической неполиаденилированной РНК, может приводить к ингибированию экспрессии гена CYPgst. Данный метод описывается, как пример, в международной патентной заявке WO2001/012824.

Согласно еще одному варианту осуществления изобретение касается способа восстановления фертильности растения по данному изобретению, включающего введение в растение функционального гена CYPgst. Ген CYPgst можно вводить путем использования методов генетической инженерии, с помощью рекомбинантной ДНК по данному изобретению, предпочтительно содержащей элементы контроля транскрипции, предпочтительно промотор для специфической экспрессии гена в закрытых цветках и/или плодах, либо посредством векторов по данному изобретению. Ген CYPgst можно также вводить путем скрещивания с растением, несущим ген CYPgst дикого типа или функциональный ген CYPgst, предпочтительно в гомозиготном состоянии. Как вариант, отбор по признаку присутствия гена CYPgst дикого типа или функционального гена CYPgst можно проводить в поколениях после скрещивания.

Объектом настоящего изобретения является растение, которое содержит заранее определенную растительную клетку и/или которое получают описанным выше способом. Это означает растение, имеющее рецессивный ядерно-кодируемый мужски стерильный фенотип, а также растение с восстановленной фертильностью, полученное либо посредством генетической рекомбинации с использованием традиционных селекционных методов, либо являющееся трансгенным растением, в котором экспрессия гена CYPgst ингибируется вследствие различных методов, упомянутых выше, в результате которых получается

растение с рецессивным ядерно-кодируемым мужски стерильным фенотипом или восстанавливается fertильность растения посредством введения рекомбинантных молекул ДНК. Такое растение предпочтительно является культивируемым растением, наиболее предпочтительно культурным растением, предпочтительно имеющим запасающие органы, которые потенциально могут функционировать как репродуктивные органы, такие как корни или клубни растений сахарной свеклы или картофеля, зерно тритикале, овса, проса и кукурузы, плоды, например, томаты и т.п. Согласно варианту осуществления изобретения низшие растения и резушка *Arabidopsis thaliana* непосредственно не являются предметом настоящего изобретения.

Помимо растений, проявляющих рецессивный ядерно-кодируемый мужски стерильный фенотип в результате спонтанной мутации в гене CYPgst, растений, которым данный фенотип придается путем генетической рекомбинации с использованием традиционных селекционных методов, или растений, которым данный фенотип был придан посредством генной инженерии с использованием современной биотехнологии, а также растений, у которых fertильность восстановлена посредством соответствующих методов, изобретение также касается органов, растительных частей, тканей, клеток, а также семян или потомков таких растений. Согласно варианту осуществления изобретения семена или потомки имеют одну или несколько описанных выше мутаций, которые приводят к ингибированию экспрессии гена CYPgst, и/или семена или потомки проявляют рекомбинантную молекулу ДНК, молекулу нуклеиновой кислоты или вектор, описанные выше.

Согласно варианту осуществления изобретения обеспечен способ идентификации растения по данному изобретению. Данный способ идентификации можно использовать в отношении растения, имеющего рецессивный ядерно-кодируемый мужски стерильный фенотип в результате спонтанной мутации в гене CYPgst, а также растения, которому данный фенотип был придан путем генетической рекомбинации с использованием традиционных селекционных методов, или растения, которому данный фенотип был придан путем генной инженерии с использованием современной биотехнологии. Кроме того, при помощи данного способа можно идентифицировать растение, у которого посредством соответствующего метода была восстановлена fertильность. Для идентификации растения по данному изобретению в качестве зонда для гибридизации можно использовать ранее определенную молекулу нуклеиновой кислоты, которая имеет минимальную длину 15 нуклеотидов и специфически связывается с ранее определенной нуклеотидной последовательностью, содержащей не мутировавший (дикого типа) ген CYPgst и ген CYPgst с ранее определенной мутацией, приводящей к ингибированию экспрессии гена. Более того, для идентификации можно использовать вышеописанные олигонуклеотид, белок CYPgst или его фрагмент, а также антитела и компоненты набора, описанные выше.

Настоящее изобретение также касается использования рекомбинантных молекул ДНК или молекул нуклеиновой кислоты, векторов по данному изобретению и компонентов набора по данному изобретению для получения рецессивного ядерно-кодируемого мужски стерильного растения, растения с восстановленной fertильностью, гибридного растения в программах селекции на устойчивость или для производства семян.

Способ получения обратимой мужской стерильности у растения, в котором можно использовать ген CYPgst, описан, как пример, в международной патентной заявке WO96017945 и состоит из этапов:

(а) введения первой рекомбинантной молекулы ДНК в геном пыльцы для получения растения, которое может быть генетически трансформировано, причем первая рекомбинантная молекула ДНК содержит:

(i) нуклеотидную последовательность, кодирующую генетический продукт, который ингибитирует образование или функцию пыльцы, в зависимости от экспрессии в растении, в данном случае гена CYPgst по данному изобретению, или генетического продукта, например, посредством экспрессии последовательности iРНК;

(ii) оператор, контролирующий экспрессию нуклеотидной последовательности; и

(iii) промотор со специфической активностью в клетках, критически важных для образования пыльцы или ее функционирования, при этом промотор функционально связан с нуклеотидной последовательностью, кодирующей генетический продукт;

(б) как вариант, селекции растения, полученного на этапе (а), в условиях, обусловливающих возникновение мужской стерильности в результате экспрессии нуклеотидной последовательности;

(с) скрещивания мужски стерильных растений, полученных на этапах (а) или (б), с пыльцой мужски фертильной лини для получения гибридного растения, являющегося мужски фертильным, при этом пыльца другой рекомбинантной молекулы ДНК содержит: нуклеотидную последовательность, которая кодирует ДНК-связывающий белок и подавляет транскрипцию, и промотор, который контролирует экспрессию нуклеотидной последовательности, причем ДНК-связывающий белок способен связываться с оператором рекомбинантной ДНК мужски стерильного растения и подавлять транскрипцию.

Еще одна система получения растений со стерильной пыльцой, где в ядерный геном вводится чужеродная ДНК, которую можно использовать по данному изобретению, описана в европейской патентной заявке EP 0 344 029.

Настоящее изобретение также касается использования растений по данному изобретению для селекции или получения потомков, отличающегося тем, что для рекуррентной селекции используется ядерно-кодируемый мужски-стерильный фенотип. Кроме того, помимо растения по данному изобретению, предметом настоящего изобретения также являются семена или потомки, органы, растительные части, ткани или их клетки, используемые для производства продуктов, обычно изготавляемых из возобновляемых сырьевых материалов, таких как пищевые продукты и корма, предпочтительно сахар или патока (меласса), причем меласса также применяется в промышленном производстве, например, для ферментации спирта или как питательное вещество для производства продуктов биотехнологии, при производстве сырьевых материалов или субстанций для химической промышленности, например, тонких химикатов, лекарственных средств или их ингредиентов, диагностических продуктов, косметики, биоэтанола или биогаза. Пример использования растений сахарной свеклы в качестве биогенетического сырьевого материала для получения биогаза описан в патентной заявке DE 10 2012 022 178 A1, см. параграф 10.

Настоящее изобретение также относится к продуктам, получаемым из растений, органов, растительных частей, тканей, клеток, семян и потомков по данному изобретению, таким как пищевые продукты, корма и сырьевые материалы, содержащиеся в растении, семенах, потомках, органах, растительных частях, тканях или клетках, а также их компонентах.

Следующие примеры осуществления изобретения приводятся только с целью его иллюстрации и не должны рассматриваться как ограничивающие его объем. Если не установлено иное, то применялись стандартные микробиологические методы, см., например (Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001); Fritsch *et al.*, Cold Spring Harbor Laboratory Press, 1989; Mayer *et al.*, Immunochemical Methods In Cell And Molecular Biology, eds. Academic Press, London, 1987) и Weir *et al.*, Handbook of Experimental Immunology, Volumes I-IV, Blackwell, eds., 1986).

Примеры

1. Идентификация локуса, вызывающего ядерно-кодируемую мужскую стерильность

Для идентификации локуса (рабочее название *gst*), вызывающего ядерно-кодируемую мужскую стерильность у растений сахарной свеклы, в качестве исходного растения, проявляющего ядерно-кодируемый рецессивный мужской стерильный фенотип, использовали донор с внутренней меткой C311 [2043_K5]. Провести заранее (т.е. до цветения) основное определение присутствия и степени зиготности локуса *gst*, связанного с экспрессией признака в данном доноре, однако, оказалось невозможным в тестируемом материале. Вместо этого большое количество растений с предполагаемой стерильностью доводилось до стадии цветения в поле или блоке самоопыления (S-блок). Цветущие растения затем вручную проверялись на фертильность и стерильность, соответственно (Рис. 1). Фертильные индивидуумы удалялись, и собирались семена стерильных индивидуумов.

Для генетического и физического ограничения локуса *gst*, вызывающего ядерно-кодируемую мужскую стерильность, в отношении данного признака создавали раздельные карты популяции сахарной свеклы. Целевой участок на хромосоме 1 был уже известен из предварительной информации широко геномного картирования. Индивидуумы с мужской стерильностью *gst* донора C311[2043_K5] скрещивали с однолетней линией, и полученные в результате индивидуумы F1 размножали путем самоопыления. Потомков поколения S1 затем фенотипировали для целей картирования и помечали маркером класса KASP-DNA (технология KASP™ для генотипирования компании LGC Limited). Были разработаны два маркера класса KASP-DNA, sxn2151s01 и sle3305s02, при помощи которых из референсного генотипа KWS2320 идентификацией SNP выделяли генотип, несущий *gst*. Последовательность маркера sxn2151s01, представленная в SEQ ID No. 24, и последовательность маркера sle3305s02, представленная в SEQ ID No. 26, указывают на присутствие локуса *gst*; последовательность маркера sxn2151s01, представленная в SEQ ID No. 25, и последовательность маркера sle3305s02, представленная в SEQ ID No. 27, указывают на референсную последовательность, причем каждая из маркерных последовательностей отличается друг от друга в позиции 2, и в генотипе, несущем локус *gst*, в этом месте присутствует «G», а в референсном генотипе KWS2320 в этом месте присутствует «A». В результате такого тонкого картирования участок на хромосоме 1

сахарной свеклы, фланкируемый маркерами класса KASP-DNA sxn2151s01 на расстоянии 33,42 сМ и sle3305s02 на расстоянии 35,15 сМ (на основании генетической карты ZR INT 1202) и несущий локус *gst*, четко выделялся. Исходя из физической карты генома (Physmapv2), данный участок имел физический размер 215.4 kbp и находился между позициями 3185718 bp и 3401120 bp. С учетом идентифицированной позиции и общедоступного представления генома RefBeet 1.2 (<http://bvseq.molgene.mpg.de/>) в этом сегменте генома был идентифицирован 21 ген, кодирующий белок.

Гомологичные гены у модельных растений (например, *Arabidopsis thaliana* и *Oryza sativa*) для всех 21 моделей генов изучали с использованием биоинформационических походов. Широкий анализ и оценку проводили на основании идентифицированных гомологичных генов у модельных растений. Ген, кодируемый в локусе *gst*, идентифицировали как член семейства цитохром Р450-зависимых монооксигеназ (CYP) на основании широкого анализа последовательностей и предсказанных структур. Несмотря на значительное разнообразие последовательностей в цитохром Р450-зависимых монооксигеназах, все CYP имеют общие структурные признаки, которые высоко сохранялись на участке активного центра (см., например, Fischer *et al.*, Bioinformatics 23 (2007), 2015-2017), и также обнаруживались в отношении предполагаемого гена *gst*. В связи с этим ген получил название CYP_{gst}.

Путем более специфической характеристики этого гена был идентифицирован ген CYP *Arabidopsis thaliana*, т.е. CYP703A2, имеющий более высокую идентичность последовательности с этим геном из локуса *gst*. Мутант *Arabidopsis thaliana*, в котором этот ген инактивировался посредством инсерции Т-ДНК, проявлял фенотип с частичной мужской стерильностью (Morant *et al.*, Plant Cell 19 (2007), 1473-1487). Возможно это объясняется механическими причинами, как результат функции CYP703A2 в синтезе спорополленина, являющегося основным компонентом внешних стенок жизнеспособной пыльцы. Отсутствие внешней стенки препятствует созреванию пыльцы, либо подвергает ее воздействию окружающей среды. Между фенотипом *gst* фенотипа сахарной свеклы и фенотипом мутанта *Arabidopsis thaliana* имеется, однако, существенное различие:

- (i) в отличие от *Arabidopsis* нокаут гена из растений сахарной свеклы приводит к полной мужской стерильности,
- (ii) в то время как стерильная пыльца в основном образуется у мутантов *Arabidopsis*, и пыльца не образуется у *gst* растений сахарной свеклы, как следует из анализа текущего состояния.

Таким образом, нельзя исключать вероятность того, что они принадлежат к разным членам семейства CYP, и/или что функции обоих белков разные у *Arabidopsis* и растений сахарной свеклы. Известно, что *Arabidopsis* – это дикорастущее травянистое растение семейства крестоцветных, имеющее небольшой компактный геном, в то время как сахарная свекла является культивируемым растением, т.е. растением, высаживаемым, возделываемым и размножаемым человеком и используемым как культурное растение. Таким образом, исследования на основе *Arabidopsis* и их результаты не могут быть непосредственно применены к культивируемым растениям. Более того, для культурных растений характерны важные агропроцессы, которые никогда не имеют места в отношении *Arabidopsis*. Это включает, например, в отличие от *Arabidopsis*, образование корней свеклы, клубней и зерна,

которые функционируют как запасающие органы и как вегетативные и репродуктивные органы, а также их связь с симбиотическими микоризными грибами или патогенами.

2. Характеристика гена CYPgst

После идентификации потенциального гена, приводящего к фенотипу *gst*, проводили сравнительное секвенирование фрагмента геномной ДНК размером приблизительно 5 кбр. Были секвенированы фертильный референсный генотип KWS2320 сахарной свеклы, стерильный *gst* донор C311, три индивидуума вышеупомянутой популяции тонкого картирования, классифицированные согласно фенотипам и маркерным данным как стерильные, и три их индивидуума, классифицированные согласно фенотипам и маркерным данным как фертильные гомозиготные. Секвенированный участок генома содержит модель гена *BvCYPgst* (*GST*, g6845.t1), показанный на Рис. 2, и приблизительно 1,5 кбр участка предполагаемого промотора. Сравнительное секвенирование показало, помимо серии одиночных нуклеотидных полиморфизмов (SNP) между стерильными и фертильными индивидуумами, делецию размером 533 бр у стерильных генотипов, содержащую 5'-UTR и первый экзон модели гена (Рис. 2 и Рис. 3).

Анализ всех идентифицированных полиморфизмов показал, что некоторые из делеций повлияли на кодируемый белок, а все остальные мутации либо находились в нетранслируемых областях, либо генерировали синонимичные кодоны. Наоборот, эти мутации препятствовали конкретной транскрипции мРНК, и трансляция функционального белка была поэтому невозможной. Последующие анализы транскрипции подтвердили эти данные (Рис. 5). *BvCYPgst* (*GST*, g6845.t1) весьма специфически экспрессировался в закрытых цветках и плодах у фертильных генотипов. Его экспрессия не детектировалась в корнях и листьях. В противоположность этому ген *GST* не детектировался в закрытых цветках у стерильных генотипов. Это позволяет сделать вывод о том, что ген стерильных генотипов полностью инактивируется.

Были разработаны ДНК-маркеры, позволяющие проводить различие между стерильными и фертильными генотипами. Для этого разрабатывали KASP-маркеры, имеющие фертильный аллель (инсерция) в качестве доминантного признака (sle5983d14, sle5983d17).

Маркер	Прямой праймер	Обратный праймер
sle5983d1 4	SEQ ID NO: 4: ACCAAAATTATACCAATGGCTCA AG	SEQ ID NO: 5: GGCCGGGAGGGAGTTGTATGTT
sle5983d1 7	SEQ ID NO: 6: AGAAATCATACGTGAGATCTTAGTT CG	SEQ ID NO: 7: GGTATGTGGACGAGACGCAAATAC AT

В результате, что касается данной гомозиготной делеции, можно сделать косвенный вывод, что оба маркера (sle5983d14, sle5983d17) проявляют нулевой аллель, а третий убиквитарный маркер подтверждает экстракцию достаточного количества ДНК. Потенциальные точечные мутации в гене *BvCYPgst*, приводящие к преждевременному прекращению транскрипции гена *CYPgst*, или вызывающие разрушение сплайсинга, которые

можно тестировать, используя традиционные методы для детекции точечных мутаций ДНК (SNP-анализ), приводятся в Таблице 1.

Таблица 1. Потенциальные точечные мутации в гене CYPgst *Beta vulgaris*, подвид *vulgaris*, приводящие к преждевременному прекращению транскрипции гена CYPgst или вызывающие разрушение сплайсинга.

Позиция	Нуклеотид	Мутация	
SEQ ID NO: 1			Действие мутации
1771	G	T	стоп-кодон
1778	T	A или G	стоп-кодон
1788	T	A или G	стоп-кодон
1790	T	A или G	стоп-кодон
1797	T	A	стоп-кодон
1813	A	T	стоп-кодон
1820	T	A или G	стоп-кодон
1824	C	A или G	стоп-кодон
1825	C	T	стоп-кодон
1829	G	A	стоп-кодон
1830	G	A	стоп-кодон
1834	A	T	стоп-кодон
1842	C	A или G	стоп-кодон
1844	T	A или G	стоп-кодон
1848	C	A или G	стоп-кодон
1857	C	A или G	стоп-кодон
1858	A	T	стоп-кодон
1883	G	A	стоп-кодон
1884	G	A	стоп-кодон
1889	T	A или G	стоп-кодон
1894	G	T	стоп-кодон

1906	C	T	стоп-кодон
1940	C	A или G	стоп-кодон
1947	T	A	стоп-кодон
1948	G	T	стоп-кодон
1951	A	T	стоп-кодон
1956	T	A или G	стоп-кодон
1964	T	A или G	стоп-кодон
1971	C	A или G	стоп-кодон
2014	G	T	стоп-кодон
2026	G	T	стоп-кодон
2033	T	A или G	стоп-кодон
2038	C	T	стоп-кодон
2041	C	T	стоп-кодон
2075	T	A или G	стоп-кодон
2090	T	A	стоп-кодон
2097	C	A или G	стоп-кодон
2117	T	A	стоп-кодон
2128	G	T	стоп-кодон
2138	G	A	стоп-кодон
2139	G	A	стоп-кодон
2140	A	T	стоп-кодон
2143	A	T	стоп-кодон
2149	A	T	стоп-кодон
2160	C	A	стоп-кодон
2164	G	T	стоп-кодон
2171	T	A	стоп-кодон
2182	A	T	стоп-кодон
2185	C	T	стоп-кодон
2191	G	T	стоп-кодон

2218	G	T	стоп-кодон
2224	C	T	стоп-кодон
2231	T	A	стоп-кодон
2236	C	T	стоп-кодон
2246	T	A или G	стоп-кодон
2260	A	T	стоп-кодон
2266	A	T	стоп-кодон
2279	T	A	стоп-кодон
2284	G	T	стоп-кодон
2291	T	A или G	стоп-кодон
2320	A	T	стоп-кодон
2327	T	A	стоп-кодон
2335	A	T	стоп-кодон
2338	C	T	стоп-кодон
2343	C	A или G	стоп-кодон
2368	C	T	стоп-кодон
2380	G	T	стоп-кодон
2401	G	T	стоп-кодон
2405	T	A	стоп-кодон
2411	G	A	стоп-кодон
2412	G	A	стоп-кодон
2414	T	A или G	стоп-кодон
2423	T	A	стоп-кодон
2430	C	A или G	стоп-кодон
2432	T	A	стоп-кодон
2442	T	A или G	стоп-кодон
2444	T	A	стоп-кодон
2453	G	A	стоп-кодон
2454	G	A	стоп-кодон

2459	G	A	стоп-кодон
2460	G	A	стоп-кодон
2472	T	А или G	стоп-кодон
2473	G	T	стоп-кодон
2478	T	A	стоп-кодон
2479	G	T	стоп-кодон
2482	A	T	стоп-кодон
2485	A	T	стоп-кодон
2494	G	T	стоп-кодон
2500	G	T	стоп-кодон
2503	A	T	стоп-кодон
2527	A	T	стоп-кодон
2536	G	T	стоп-кодон
2539	G	T	стоп-кодон
2548	A	T	стоп-кодон
2551	G	T	стоп-кодон
2554	A	T	стоп-кодон
2557	A	T	стоп-кодон
2563	A	T	стоп-кодон
2566	G	T	стоп-кодон
2569	G	T	стоп-кодон
2575	G	T	стоп-кодон
2584	G	T	стоп-кодон
2590	G	T	стоп-кодон
2612	T	A	стоп-кодон
2615	T	A	стоп-кодон
2621	T	A	стоп-кодон
2629	G	T	стоп-кодон
2635	G	T	стоп-кодон

2641	G	T	стоп-кодон
2665	A	T	стоп-кодон
2677	C	T	стоп-кодон
2679	G	A	расщепление мутации
2680	G	A	расщепление мутации
2681	T	A	расщепление мутации
3505	A	C или G или T	расщепление мутации
3506	G	A или C или T	расщепление мутации
3535	C	A или G	стоп-кодон
3549	G	T	стоп-кодон
3553	G	A	стоп-кодон
3554	G	A	стоп-кодон
3564	G	T	стоп-кодон
3573	A	T	стоп-кодон
3594	A	T	стоп-кодон
3600	C	T	стоп-кодон
3603	C	T	стоп-кодон
3606	G	T	стоп-кодон
3624	G	T	стоп-кодон
3633	C	T	стоп-кодон
3645	G	T	стоп-кодон
3649	C	A или G	стоп-кодон
3671	C	A или G	стоп-кодон
3680	T	A	стоп-кодон
3690	G	T	стоп-кодон
3699	C	T	стоп-кодон
3724	T	A или G	стоп-кодон
3735	G	T	стоп-кодон
3739	C	A или G	стоп-кодон

3767	T	A или G	стоп-кодон
3811	T	A или G	стоп-кодон
3813	G	T	стоп-кодон
3825	A	T	стоп-кодон
3832	G	A	стоп-кодон
3833	G	A	стоп-кодон
3843	G	T	стоп-кодон
3854	C	A или G	стоп-кодон
3858	G	T	стоп-кодон
3861	A	T	стоп-кодон
3868	G	A	стоп-кодон
3869	G	A	стоп-кодон
3874	T	A	стоп-кодон
3879	G	T	стоп-кодон
3885	A	T	стоп-кодон
3891	G	T	стоп-кодон
3903	G	T	стоп-кодон
3915	A	T	стоп-кодон
3922	T	A или G	стоп-кодон
3939	A	T	стоп-кодон
3942	A	T	стоп-кодон
3945	A	T	стоп-кодон
3950	T	A	стоп-кодон
3982	T	A	стоп-кодон
3987	G	T	стоп-кодон
3991	T	A	стоп-кодон
4018	G	A	стоп-кодон
4019	G	A	стоп-кодон
4021	T	A или G	стоп-кодон

4032	G	T	стоп-кодон
4038	A	T	стоп-кодон
4044	G	T	стоп-кодон
4047	G	T	стоп-кодон
4059	A	T	стоп-кодон
4062	G	T	стоп-кодон
4070	T	А или G	стоп-кодон
4086	A	T	стоп-кодон
4092	C	T	стоп-кодон
4099	T	А или G	стоп-кодон
4108	T	A	стоп-кодон
4113	A	T	стоп-кодон
4132	T	А или G	стоп-кодон
4136	T	А или G	стоп-кодон

3. Использование гена или локуса CYPgst в селекции гибридов

Как было показано ранее, мужски стерильный фенотип, обусловливающий локус *gst*, используется в селекционных программах на устойчивость для простого скрещивания на базе рекуррентного отбора. До начала клонирования гена в рамках настоящего изобретения и связанного с этим создания геномных маркеров требовалось вырастить и убрать в четыре раза больше растений, чем было необходимо, по причине их ожидаемого расщепления 3:1 по фенотипу. Эти растения оценивали на стерильность сразу после цветения, фертильные индивидуумы удаляли, чтобы предотвратить самоопыление. При ежегодном выращивании нескольких тысяч растений такой ручной отбор требует больших трудовых затрат и подвержен ошибкам. Теперь обеспечены геномные маркеры по данному изобретению (см. Пример 2 и Рис. 2), позволяющие тестировать 30000 растений и отбирать 75000 мужски стерильных растений для последующего выращивания.

Ведется постоянная работа по упрощению селекционных программ и производства семян растений сахарной свеклы и, таким образом, снижению затрат. В результате коммерческие сорта сахарной свеклы, которые производятся в настоящее время, являются тройными гибридами, что позволяет получать достаточное количество семян. Производство гибридов по селекционным программам, которые не являются коммерческими, также является трудоемким и дорогостоящим и связано с преодолением многих преград. Теперь возможно, используя фенотип с *gst* по данному изобретению и связанные с этим ДНК-

маркеры, после инсерции локуса *gst* или после мутации/ингибиования гена CYPgst в селекционной программе селекции отбирать мужски стерильные растения до посадки при помощи ДНК-маркеров и, таким образом, упрощать производственный процесс. Параллельно создается большое число многозародышевых форм генотипов-тестеров (MUS-тестеры). В перспективе возможно, что на смену современным ЦМС-технологиям придет альтернативная система, позволяющая делать мужски стерильными родительские компоненты семян, например, при помощи SPT-системы, указанной выше и показанной на Рис. 5. Соответственно, разумно предположить, что системы CYPgst по данному изобретению можно использовать и для других культивируемых растений, в частности культурных растений, как в случае коммерческого производства двойных гибридов, таких как кукуруза. Что касается кукурузы (*Zea mays*), важную роль в разработке альтернативных систем получения ее гибридных семян играют *ms*-гены. Анализ последовательностей свидетельствует о предполагаемом существовании *BvCYPgst*(GRMZM5g830329) гомолога кукурузы. В кукурузе также имеется много *ms*-мутантов, из которых только часть была клонирована. *ms*-мутанты, *ms*-мутация которых обусловлена мутацией или ингибиением *BvCYPgst* гомолога кукурузы, теперь можно отбирать при помощи настоящего изобретения и использовать для целенаправленного производства семян. Настоящее изобретение можно также применять для создания гибридного картофеля, например, для введения направленных мутаций в ген-гомолог *BvCYPgst* картофеля и использовать полученную таким образом мужскую стерильность картофеля для создания диплоидного гибридного картофеля, как показано в SPT-системе.

Специфическая экспрессия гена *BvCYPgst* в цветках и тапетуме способствует биотехнологическому использованию промотора, например, для экспрессии положительно-полярной/отрицательно-полярной РНК или рибозима для ингибиования гена *BvCYPgst* или экспрессии функционального белка CYPgst, либо предполагаемого гомолога, аналога или ортолога гена *BvCYPgst* для улучшения мутации и восстановления мужски фертильного фенотипа. Предполагается, что обеспечение локуса гена и нуклеиновых и аминокислотных последовательностей гена *BvCYPgst*, в добавление к генетическим маркерам и обусловленным этим вариантам осуществления изобретения, сделает селекционные программы более простыми и менее дорогостоящими, так как, помимо прочего, это способствует раннему отбору стерильных индивидуумов. Это также упрощает организацию материального обеспечения и расширение селекционных программ.

SEQUENCE LISTING

<110> KWS SAAT SE
<120> Kernkodierte männliche Sterilität durch Mutation in Cytochrom P450 Oxidase
<130> KWS0231PCT
<150> DE 10 2016 106 656.7
<151> 2016-04-12
<160> 27
<170> PatentIn version 3.5
<210> 1
<211> 4328
<212> DNA
<213> Beta vulgaris

<220>
<221> Promotor
<222> (1)..(1518)
<223> putativer Promotor

<220>
<221> Gen
<222> (1519)..(4275)

<220>
<221> 5'UTR
<222> (1519)..(1761)

<220>
<221> exon
<222> (1762)..(2679)
<223> Exon 1

<220>
<221> Intron
<222> (2680)..(3506)

<220>
<221> exon
<222> (3507)..(4142)
<223> Exon 2

<220>
<221> 3'UTR
<222> (4143)..(4275)

<400> 1
ttattaaacc tgattggaac ttattgaacc ttattagacc tgattggaac ttattgcacc 60
tgattggaac ttattggaac ttattagacc ttattggaac ttattgcact tattagacct 120
tattgcaact tatctgaact tatctgaaca aatctgaact tattggacct gaaacttaat 180
tttttaagtt gaacagaacg cacccttagt atatcggtgc cacatgtgcg ttgaatttt 240
cctttcccta tccttcac tccatattct cctcaaaaagt gtgtaaaaat ccgacacacg 300

agtagaatgg gattgaagtg ggtcaagatc taaaaccaat gggtaatgc cacaataa	360
ggtaaggttt ctcgcagtag caaaaaata aagttaagtt gagagaaaaa ttatgaatag	420
ttgtttctcg tgaagagttg tataaaaaaa aagtctaatt tgatacattt tctttacat	480
ttataaagga ttgaccaatc atccaaatta ccaaataattt aggatataaa tcttcagat	540
tacaaccat atatgataca ctaaatttta catgaggcaa tggaggattt gcatgaatat	600
cgaggagaga aaaaattagt tacaaaactt gcataatttca tccaaaccaa atcaagtcaa	660
gaaacaacga acaatattat cattagtact ataagtataat attataggct tagagcaaag	720
ccctaactac cacactgcac acaaatgata actagtaaga gagaaaaata caaatttaag	780
attcaacata gcaaatttattt catgattcat gattcatgat tcatgattca tgattcacga	840
acatcaagaa tggtagtact gataaaggac aatttaaaca taagtgtaaa gctgcacat	900
catcaattat attcgacatac tactagacca atctttactt agtacatgtg tttagtacatg	960
tgttacttca tatcagatgt attgattgtt gccaatgaca tatcatgttc acttaatctt	1020
agggccattt aattataaca tggagaataa tacaacttaa aattatgtgg tggctatcat	1080
ctcattttct agataattaa acctttattt tgtatacata tatattgtct ttacatagca	1140
aaacaatattt gaaggtataa caaccttcc cttttctttt actacatgtt tatgttagag	1200
tttttcgatt tacgattgtg gtaaatttat ttaatttgat cggttgtctt gtatcaaga	1260
aatgacgtat gaatcaattt agggcatgtt ctcttcggca taaaacagct gaactgaatt	1320
gaactgaact gaaatgaata gtgatatgtg agagtaaaag tattgtcaag agctgaactg	1380
agctgaactg aacggatctg aactgaactg atctaattctg aactaatctg aactgaactg	1440
aactgaattt aactgaaaat aagctaggaa aaacagaccc ttactactat tatataacct	1500
cgtttaataa tttagaaattt aaaaaataa ttatatttct ttatactta ttaacctatt	1560
gaggttttta tattgactcc caaatactat tttatagatc atgcatgtt aatgagcaaa	1620
ctactttctc acatcttat aggagaaaaa gtagatcact cactagcata tcatgaccag	1680
cggaaaccaac caacgctaat agtttattt cttccattttt agataagagt taactaataa	1740
taccatctt gtgaaattttt g atg gat ttt gga act tta gca ata tat tta Met Asp Phe Gly Thr Leu Ala Ile Tyr Leu 1 5 10	1791
ctg tgt gca ctt ttt gct acc aaa att tta tac caa tgg ctc aag tcc Leu Cys Ala Leu Phe Ala Thr Lys Ile Leu Tyr Gln Trp Leu Lys Ser 15 20 25	1839
tac tta tac aca aca tac aaa ctc cct ccc ggc cca cca agg tgg ccc Tyr Leu Tyr Thr Tyr Lys Leu Pro Pro Gly Pro Pro Arg Trp Pro 30 35 40	1887
tta ttt gga aac ctc ctt caa cta ggg cca ctt ccc cac cgc gat ttc Leu Phe Gly Asn Leu Leu Gln Leu Gly Pro Leu Pro His Arg Asp Phe 45 50 55	1935

gcc tca ttt tgt gaa aaa tat ggg cct tta gtc tac ata agg ctt ggt Ala Ser Phe Cys Glu Lys Tyr Gly Pro Leu Val Tyr Ile Arg Leu Gly 60 65 70	1983
aat gtg gat gcc ata acc act aat gat cca gaa atc ata cgt gag atc Asn Val Asp Ala Ile Thr Thr Asn Asp Pro Glu Ile Ile Arg Glu Ile 75 80 85 90	2031
tta gtt cga caa gat gat gta ttt gcg tct cgt cca cat acc tta gcc Leu Val Arg Gln Asp Asp Val Phe Ala Ser Arg Pro His Thr Leu Ala 95 100 105	2079
gca acc cac ttg gct tac aat agt ggt gat gtg gcc ttg gct cca atg Ala Thr His Leu Ala Tyr Asn Ser Gly Asp Val Ala Leu Ala Pro Met 110 115 120	2127
gga cca aaa tgg aaa aga atg aga agg ata tgc atg gag cac ttg ctc Gly Pro Lys Trp Lys Arg Met Arg Arg Ile Cys Met Glu His Leu Leu 125 130 135	2175
aca act aga cga ctt gaa cta ttt gtg agt cat agg gct gat gag gca Thr Thr Arg Arg Leu Glu Leu Phe Val Ser His Arg Ala Asp Glu Ala 140 145 150	2223
cga cat ttg gtc caa gac gta tta act cgt tcc cac aaa gat aaa gtt Arg His Leu Val Gln Asp Val Leu Thr Arg Ser His Lys Asp Lys Val 155 160 165 170	2271
gtt aat ttg agg gaa gtg tta ggt gca ttt tct atg aat aac gtg act Val Asn Leu Arg Glu Val Leu Gly Ala Phe Ser Met Asn Asn Val Thr 175 180 185	2319
aga atg ttg cta ggg aag caa tac ttt ggg gcc ggg acg gcg ggc cca Arg Met Leu Leu Gly Lys Gln Tyr Phe Gly Ala Gly Thr Ala Gly Pro 190 195 200	2367
caa gag gct cta gag ttt atg cat ata aca cat gag ttg ttt tgg tta Gln Glu Ala Leu Glu Phe Met His Ile Thr His Glu Leu Phe Trp Leu 205 210 215	2415
cta ggc ttg att tac ttg ggt gat tat ttg cct ttt tgg agg tgg gtt Leu Gly Leu Ile Tyr Leu Gly Asp Tyr Leu Pro Phe Trp Arg Trp Val 220 225 230	2463
gat cca tat gga tgt gaa aag aaa atg agg gaa gtt gaa aaa agg gta Asp Pro Tyr Gly Cys Glu Lys Lys Met Arg Glu Val Glu Lys Arg Val 235 240 245 250	2511
gat gat ttc cat cgc aaa att ata gag gaa cat agg aag gag aag aaa Asp Asp Phe His Arg Lys Ile Ile Glu Glu His Arg Lys Glu Lys Lys 255 260 265	2559
agg aaa gaa gaa atg gga gtg aat gag ggt gaa atg gat ttt gta gat Arg Lys Glu Glu Met Gly Val Asn Glu Gly Glu Met Asp Phe Val Asp 270 275 280	2607
att ttg ttg gct ttg cct ggt gaa aat gga aat gag cat atg gat gat Ile Leu Leu Ala Leu Pro Gly Glu Asn Gly Asn Glu His Met Asp Asp 285 290 295	2655
gca gat att aaa gct cta att cag gtaattcatg tataatttga atgtgatcga Ala Asp Ile Lys Ala Leu Ile Gln 300 305	2709

tacaaagttt gatagaaaaac atatttgcataaataatatgg ttgccctact agacccaata	2769
aaatacataa ttattgcctt actagttgaa agttgaaaca accttagctac cattttgttg	2829
tgatttatcat tagccaaacca aaattatttc ttgcattccat atattaatgt tgagatcaga	2889
gtcggcatat ttacaattac ttgtAACATT ttaAGCAAAC aaattaaaat atttttggc	2949
aagtccattt tattgaataa tacctatac ttaAAATGAA ttcttggtca tgtacacttg	3009
ccttcaagg taccaatatt tgaccatatg taattactat taacaaattt gataaaatct	3069
aataatatgt aaatatacat tcacgcacat attagaaaca aagatcacaa atgataatgc	3129
aaaataactt atttgaacta atgttgtgaa gttaaatttga aacaaaagg tatatttgtt	3189
ttgcccattt ttaatttata aattactat aacaacaact caatatgtaa aactgttaag	3249
atggagtgtg gatagaatga gatgagtata ctTTTACTAG ttaccactcg aaaatgcatt	3309
tcctccTTTGTG tttatAGTTG ttctaacttc tattatcata aataattttt tggacttatt	3369
tcaatgtata ttTACAACGC taattgttta atTTTTAAA aatacataat gtAAACAAGG	3429
atTTCACTGGT caattataca cattataaaat attatctaa AAAACTTATT aatgctcaat	3489
tagtatccat aatatacgat atgata gca gca aca gac aca tca gct	3539
Asp Met Ile Ala Ala Ala Thr Asp Thr Ser Ala	
310 315	
gta acc aac gaa tgg gcc atg gca gaa gta ata aaa cac cca cgt gtc	3587
Val Thr Asn Glu Trp Ala Met Ala Glu Val Ile Lys His Pro Arg Val	
320 325 330	
ctc cac aag atc caa caa gag ctt aac aca ata gta gga ccc aat cga	3635
Leu His Lys Ile Gln Gln Glu Leu Asn Thr Ile Val Gly Pro Asn Arg	
335 340 345	
atg gta aca gaa tca gat ctt ccc cac ctt aac tac cta cgt tgt gtc	3683
Met Val Thr Glu Ser Asp Leu Pro His Leu Asn Tyr Leu Arg Cys Val	
350 355 360 365	
gta cgt gaa acg ttc cga atg cat cca gca gga ccc ttt tta atc cca	3731
Val Arg Glu Thr Phe Arg Met His Pro Ala Gly Pro Phe Leu Ile Pro	
370 375 380	
cat gaa tca cta cgc cat aca aca atc aac ggc tat gat atc cca tct	3779
His Glu Ser Leu Arg His Thr Thr Ile Asn Gly Tyr Asp Ile Pro Ser	
385 390 395	
ggg aca cgt gtc ttc atc aac aca cat ggg tta gga cgt aac ctt aaa	3827
Gly Thr Arg Val Phe Ile Asn Thr His Gly Leu Gly Arg Asn Leu Lys	
400 405 410	
gtg tgg gac aac ata gag gat ttt tac cct gaa aga cat tgg ccg ttg	3875
Val Trp Asp Asn Ile Glu Asp Phe Tyr Pro Glu Arg His Trp Pro Leu	
415 420 425	
gat gga agt aga gtt gag att agc cat gga tct gat ttt aaa ata tta	3923
Asp Gly Ser Arg Val Glu Ile Ser His Gly Ser Asp Phe Lys Ile Leu	
430 435 440 445	
cca ttt agt gct ggg aag aga aga ttt cct ggg gcc cca ctt ggg gtg	3971
Pro Phe Ser Ala Gly Lys Arg Arg Cys Pro Gly Ala Pro Leu Gly Val	

450	455	460	
gtg ttt gtg ttg atg gga ttg gct aca ctt ttt cat gca ttt gat tgg Val Phe Val Leu Met Gly Leu Ala Thr Leu Phe His Ala Phe Asp Trp			4019
465	470	475	
tta cca cct gat gga atg aag gca gaa gaa att gat act aag gaa gtt Leu Pro Pro Asp Gly Met Lys Ala Glu Glu Ile Asp Thr Lys Glu Val			4067
480	485	490	
tat ggg atg act atg cct aaa gct caa cct tta atg gct ttg gct aaa Tyr Gly Met Thr Met Pro Lys Ala Gln Pro Leu Met Ala Leu Ala Lys			4115
495	500	505	
cct agg ctt gct cat tta tat ctt tga tacatgttca tattgtggtg Pro Arg Leu Ala His Leu Tyr Leu			4162
510	515		
cacttataag cacaatagac aaatacaagt ttgtatcgac tctaacatgt tgtagtat tagtatactg caactctaca agtataaat ttctataaacataaaacaca agtcataacg			4222
cattttgttt tgaaaaaaaaa gaggttacat tgtcttacac cataaa			4282
			4328
 <210> 2			
<211> 1930			
<212> DNA			
<213> Artificial Sequence			
 <220>			
<223> cDNA-CYPgst gene (Beta vulgaris)			
 <220>			
<221> 5'UTR			
<222> (1)..(243)			
 <220>			
<221> exon			
<222> (244)..(1161)			
<223> Exon 1			
 <220>			
<221> exon			
<222> (1162)..(1797)			
<223> Exon 2			
 <220>			
<221> 3'UTR			
<222> (1798)..(1930)			
 <400> 2			
ttaaaaaaat aattatattt ctttatactt tattaacctt ttgaggtttt tatattgact cccaaatact attttataga tcattgccatg ttaatgagca aactactttc tcacatcttt			60
ataggagaaa aagtagatca ctcaactagca tatcatgacc agcgaaaccca accaacgcta atagtttat tgcttccatt agagataaga gttaactaat aataccatct ttgtgaaatt			120
ttg atg gat ttt gga act tta gca ata tat tta ctg tgt gca ctt ttt Met Asp Phe Gly Thr Leu Ala Ile Tyr Leu Leu Cys Ala Leu Phe			180
1	5	10	240
			288
			15

gct acc aaa att tta tac caa tgg ctc aag tcc tac tta tac aca aca Ala Thr Lys Ile Leu Tyr Gln Trp Leu Lys Ser Tyr Leu Tyr Thr Thr	20	25	30	336	
tac aaa ctc cct ccc ggc cca cca agg tgg ccc tta ttt gga aac ctc Tyr Lys Leu Pro Pro Gly Pro Pro Arg Trp Pro Leu Phe Gly Asn Leu	35	40	45	384	
ctt caa cta ggg cca ctt ccc cac cgc gat ttc gcc tca ttt tgt gaa Leu Gln Leu Gly Pro Leu Pro His Arg Asp Phe Ala Ser Phe Cys Glu	50	55	60	432	
aaa tat ggg cct tta gtc tac ata agg ctt ggt aat gtg gat gcc ata Lys Tyr Gly Pro Leu Val Tyr Ile Arg Leu Gly Asn Val Asp Ala Ile	65	70	75	480	
acc act aat gat cca gaa atc ata cgt gag atc tta gtt cga caa gat Thr Thr Asn Asp Pro Glu Ile Ile Arg Glu Ile Leu Val Arg Gln Asp	80	85	90	528	
gat gta ttt gcg tct cgt cca cat acc tta gcc gca acc cac ttg gct Asp Val Phe Ala Ser Arg Pro His Thr Leu Ala Ala Thr His Leu Ala	100	105	110	576	
tac aat agt ggt gat gtg gcc ttg gct cca atg gga cca aaa tgg aaa Tyr Asn Ser Gly Asp Val Ala Leu Ala Pro Met Gly Pro Lys Trp Lys	115	120	125	624	
aga atg aga agg ata tgc atg gag cac ttg ctc aca act aga cga ctt Arg Met Arg Arg Ile Cys Met Glu His Leu Leu Thr Thr Arg Arg Leu	130	135	140	672	
gaa cta ttt gtg agt cat agg gct gat gag gca cga cat ttg gtc caa Glu Leu Phe Val Ser His Arg Ala Asp Glu Ala Arg His Leu Val Gln	145	150	155	720	
gac gta tta act cgt tcc cac aaa gat aaa gtt gtt aat ttg agg gaa Asp Val Leu Thr Arg Ser His Lys Asp Lys Val Val Asn Leu Arg Glu	160	165	170	175	768
gtg tta ggt gca ttt tct atg aat aac gtg act aga atg ttg cta ggg Val Leu Gly Ala Phe Ser Met Asn Asn Val Thr Arg Met Leu Leu Gly	180	185	190	816	
aag caa tac ttt ggg gcc ggg acg gcg ggc cca caa gag gct cta gag Lys Gln Tyr Phe Gly Ala Gly Thr Ala Gly Pro Gln Glu Ala Leu Glu	195	200	205	864	
ttt atg cat ata aca cat gag ttg ttt tgg tta cta ggc ttg att tac Phe Met His Ile Thr His Glu Leu Phe Trp Leu Leu Gly Leu Ile Tyr	210	215	220	912	
ttg ggt gat tat ttg cct ttt tgg agg tgg gtt gat cca tat gga tgt Leu Gly Asp Tyr Leu Pro Phe Trp Arg Trp Val Asp Pro Tyr Gly Cys	225	230	235	960	
gaa aag aaa atg agg gaa gtt gaa aaa agg gta gat gat ttc cat cgc Glu Lys Lys Met Arg Glu Val Glu Lys Arg Val Asp Asp Phe His Arg	240	245	250	1008	
aaa att ata gag gaa cat agg aag gag aag aaa agg aaa gaa gaa atg Lys Ile Ile Glu Glu His Arg Lys Glu Lys Lys Arg Lys Glu Glu Met	260	265	270	1056	

gga gtg aat gag ggt gaa atg gat ttt gta gat att ttg ttg gct ttg Gly Val Asn Glu Gly Glu Met Asp Phe Val Asp Ile Leu Leu Ala Leu 275 280 285	1104
cct ggt gaa aat gga aat gag cat atg gat gat gca gat att aaa gct Pro Gly Glu Asn Gly Asn Glu His Met Asp Asp Ala Asp Ile Lys Ala 290 295 300	1152
cta att cag gat atg ata gca gca gca aca gac aca tca gct gta acc Leu Ile Gln Asp Met Ile Ala Ala Ala Thr Asp Thr Ser Ala Val Thr 305 310 315	1200
aac gaa tgg gcc atg gca gaa gta ata aaa cac cca cgt gtc ctc cac Asn Glu Trp Ala Met Ala Glu Val Ile Lys His Pro Arg Val Leu His 320 325 330 335	1248
aag atc caa caa gag ctt aac aca ata gta gga ccc aat cga atg gta Lys Ile Gln Gln Glu Leu Asn Thr Ile Val Gly Pro Asn Arg Met Val 340 345 350	1296
aca gaa tca gat ctt ccc cac ctt aac tac cta cgt tgt gtc gta cgt Thr Glu Ser Asp Leu Pro His Leu Asn Tyr Leu Arg Cys Val Val Arg 355 360 365	1344
gaa acg ttc cga atg cat cca gca gga ccc ttt tta atc cca cat gaa Glu Thr Phe Arg Met His Pro Ala Gly Pro Phe Leu Ile Pro His Glu 370 375 380	1392
tca cta cgc cat aca aca atc aac ggc tat gat atc cca tct ggg aca Ser Leu Arg His Thr Thr Ile Asn Gly Tyr Asp Ile Pro Ser Gly Thr 385 390 395	1440
cgt gtc ttc atc aac aca cat ggg tta gga cgt aac ctt aaa gtg tgg Arg Val Phe Ile Asn Thr His Gly Leu Gly Arg Asn Leu Lys Val Trp 400 405 410 415	1488
gac aac ata gag gat ttt tac cct gaa aga cat tgg ccg ttg gat gga Asp Asn Ile Glu Asp Phe Tyr Pro Glu Arg His Trp Pro Leu Asp Gly 420 425 430	1536
agt aga gtt gag att agc cat gga tct gat ttt aaa ata tta cca ttt Ser Arg Val Glu Ile Ser His Gly Ser Asp Phe Lys Ile Leu Pro Phe 435 440 445	1584
agt gct ggg aag aga aga tgt cct ggg gcc cca ctt ggg gtg gtg ttt Ser Ala Gly Lys Arg Arg Cys Pro Gly Ala Pro Leu Gly Val Val Phe 450 455 460	1632
gtg ttg atg gga ttg gct aca ctt ttt cat gca ttt gat tgg tta cca Val Leu Met Gly Leu Ala Thr Leu Phe His Ala Phe Asp Trp Leu Pro 465 470 475	1680
cct gat gga atg aag gca gaa gaa att gat act aag gaa gtt tat ggg Pro Asp Gly Met Lys Ala Glu Glu Ile Asp Thr Lys Glu Val Tyr Gly 480 485 490 495	1728
atg act atg cct aaa gct caa cct tta atg gct ttg gct aaa cct agg Met Thr Met Pro Lys Ala Gln Pro Leu Met Ala Leu Ala Lys Pro Arg 500 505 510	1776
ctt gct cat tta tat ctt tga tacatgttca tattgtggtg cacttataag Leu Ala His Leu Tyr Leu 515	1827

cacaatagac aaatacaagt ttgtatcgac tctaacatgt tgtttagtat tagtatactg 1887
 caactctaca agtatgtaat ttctataaac tataaacaca agt 1930

 <210> 3
 <211> 517
 <212> PRT
 <213> Beta vulgaris

 <220>
 <221> PEPTIDE
 <222> (1)..(517)
 <223> protein CYPgst (Beta vulgaris)

 <400> 3

 Met Asp Phe Gly Thr Leu Ala Ile Tyr Leu Leu Cys Ala Leu Phe Ala 1
 1 5 10 15

 Thr Lys Ile Leu Tyr Gln Trp Leu Lys Ser Tyr Leu Tyr Thr Thr Tyr 20
 25 30

 Lys Leu Pro Pro Gly Pro Pro Arg Trp Pro Leu Phe Gly Asn Leu Leu 35
 40 45

 Gln Leu Gly Pro Leu Pro His Arg Asp Phe Ala Ser Phe Cys Glu Lys 50
 55 60

 Tyr Gly Pro Leu Val Tyr Ile Arg Leu Gly Asn Val Asp Ala Ile Thr 65
 70 75 80

 Thr Asn Asp Pro Glu Ile Ile Arg Glu Ile Leu Val Arg Gln Asp Asp 85
 90 95

 Val Phe Ala Ser Arg Pro His Thr Leu Ala Ala Thr His Leu Ala Tyr 100
 105 110

 Asn Ser Gly Asp Val Ala Leu Ala Pro Met Gly Pro Lys Trp Lys Arg 115
 120 125

 Met Arg Arg Ile Cys Met Glu His Leu Leu Thr Thr Arg Arg Leu Glu 130
 135 140

 Leu Phe Val Ser His Arg Ala Asp Glu Ala Arg His Leu Val Gln Asp 145
 150 155 160

 Val Leu Thr Arg Ser His Lys Asp Lys Val Val Asn Leu Arg Glu Val 165
 170 175

 Leu Gly Ala Phe Ser Met Asn Asn Val Thr Arg Met Leu Leu Gly Lys

180 185 190

Gln Tyr Phe Gly Ala Gly Thr Ala Gly Pro Gln Glu Ala Leu Glu Phe
195 200 205

Met His Ile Thr His Glu Leu Phe Trp Leu Leu Gly Leu Ile Tyr Leu
210 215 220

Gly Asp Tyr Leu Pro Phe Trp Arg Trp Val Asp Pro Tyr Gly Cys Glu
225 230 235 240

Lys Lys Met Arg Glu Val Glu Lys Arg Val Asp Asp Phe His Arg Lys
245 250 255

Ile Ile Glu Glu His Arg Lys Glu Lys Lys Arg Lys Glu Glu Met Gly
260 265 270

Val Asn Glu Gly Glu Met Asp Phe Val Asp Ile Leu Leu Ala Leu Pro
275 280 285

Gly Glu Asn Gly Asn Glu His Met Asp Asp Ala Asp Ile Lys Ala Leu
290 295 300

Ile Gln Asp Met Ile Ala Ala Thr Asp Thr Ser Ala Val Thr Asn
305 310 315 320

Glu Trp Ala Met Ala Glu Val Ile Lys His Pro Arg Val Leu His Lys
325 330 335

Ile Gln Gln Glu Leu Asn Thr Ile Val Gly Pro Asn Arg Met Val Thr
340 345 350

Glu Ser Asp Leu Pro His Leu Asn Tyr Leu Arg Cys Val Val Arg Glu
355 360 365

Thr Phe Arg Met His Pro Ala Gly Pro Phe Leu Ile Pro His Glu Ser
370 375 380

Leu Arg His Thr Thr Ile Asn Gly Tyr Asp Ile Pro Ser Gly Thr Arg
385 390 395 400

Val Phe Ile Asn Thr His Gly Leu Gly Arg Asn Leu Lys Val Trp Asp
405 410 415

Asn Ile Glu Asp Phe Tyr Pro Glu Arg His Trp Pro Leu Asp Gly Ser
420 425 430

Arg Val Glu Ile Ser His Gly Ser Asp Phe Lys Ile Leu Pro Phe Ser

435	440	445
Ala Gly Lys Arg Arg Cys Pro Gly Ala Pro Leu Gly Val Val Phe Val		
450	455	460
Leu Met Gly Leu Ala Thr Leu Phe His Ala Phe Asp Trp Leu Pro Pro		
465	470	475
Asp Gly Met Lys Ala Glu Glu Ile Asp Thr Lys Glu Val Tyr Gly Met		
485	490	495
Thr Met Pro Lys Ala Gln Pro Leu Met Ala Leu Ala Lys Pro Arg Leu		
500	505	510
Ala His Leu Tyr Leu		
515		
<210> 4		
<211> 27		
<212> DNA		
<213> Artificial Sequence		
<220>		
<223> s1e5983d14 Forward (Fw) Primer (5' -3')		
<400> 4		
acccaaaattt tataccaatg gctcaag 27		
<210> 5		
<211> 23		
<212> DNA		
<213> Artificial Sequence		
<220>		
<223> s1e5983d14 Reverse (Rv) Primer (5' -3')		
<400> 5		
ggccgggagg gagtttgat gtt 23		
<210> 6		
<211> 27		
<212> DNA		
<213> Artificial Sequence		
<220>		
<223> s1e5983d17 Forward (Fw) Primer (5' -3')		
<400> 6		
agaaatcata cgtgagatct tagttcg 27		
<210> 7		
<211> 26		
<212> DNA		
<213> Artificial Sequence		

<220>
 <223> s1e5983d17 Reverse (Rv) Primer (5' -3')

 <400> 7
 ggtatgtgga cgagacgcaa atacat 26

 <210> 8
 <211> 4220
 <212> DNA
 <213> Beta vulgaris

 <220>
 <221> Promotor
 <222> (1)..(1353)
 <223> putativer Promotor

 <220>
 <221> Gen
 <222> (1354)..(3542)

 <220>
 <221> exon
 <222> (1354)..(1938)
 <223> verkrztes Exon 1

 <220>
 <221> Intron
 <222> (1939)..(2755)

 <220>
 <221> exon
 <222> (2756)..(3394)
 <223> Exon 2

 <220>
 <221> 3'UTR
 <222> (3395)..(3542)

 <400> 8
 gtatatcggtt gccacatgtg cgttgaattt ttccctttcc ttcctttcc actccatatt 60
 ctcctcaaaa gtgtgtaaaa atccgacaca cgagtagaat gggattgaag tgggtcaaga 120
 tctgaaacca atgggtcaat gccacaaaat aaggtaaggt ttctcgcagt agcaaaaaaa 180
 taaagttaag ttgagagaaa aattatgaat agttgttct cgtgaagagt tgtataaaaa 240
 aaaagtctaa tttgatacat ttcttttac atttataaaag gattgaccaa tcataaaaaat 300
 taccaaataat ttaggatata aatctttcag attacaaccc atatatgata cactaaattt 360
 tacatgagggc aatggaggat ttgcatgaat atcgaggaga gaaaaaaatta gttacaaaac 420
 ttgcataatt tatccaaacc aaatcaagtc aagaaacaac gaacaatatt atcattagta 480
 ctataagtat atattatagg cttagagcaa agccctaact accacactgc acacaaatga 540
 taactagtaa gagagggaaa tacaattt aagttcaaca tagcaaatta ttcatgattc 600
 atgattcatg attcatgatt catgattcac gaacatcaag aatggtataag ctgataaaagg 660
 acaatttaaa cataagtgtt aagctcgac atcatcaatt atattcgcat actactagac 720

caatcttac ttagtacatg tgtagtaca tgtgttactt catatcagat gtattgattg	780
ttgccaatga catatcatgt tcacttaatc ttagggccat ttaattataa catggagaat	840
aatacaactt aaaattatgt ggtggctatc atctcatttt ctagataatt aaacctttat	900
tttgtataca tatatatattgt cttacatag caaaaacaata ttgaaggtat aacaaccttt	960
ccctttctt ttactacatg tttatgttag agttttcga tttacgattt tggtaaatta	1020
attgttaattt atcggttgtc ttgttagtcaa gaaatgacgt atgaatcaat tttagggcatg	1080
ttctcttcgg cataaaaacag ctgaactgaa ttgaactgaa ctgaaaatgaa tagtgcata	1140
tgagagtaaa agtattgtca agagctgaac tgagctgaac tgaacggatc tgaactgaac	1200
tgatctaattc tgaactaattc tgaactgaac tgaactgaat tgaactgaaa ataagctagg	1260
gaaaacagac ccttactact attatataac ctcgttaaa tattaggaaa ttaaaaaaaaat	1320
aattatattt ctttatactt tattaaccta tta tac aat agt ggt gat gtg gcc	1374
Tyr Asn Ser Gly Asp Val Ala	
1 5	
ttg gct cca atg gga cca aaa tgg aaa aga atg aga agg ata tgc atg	1422
Leu Ala Pro Met Gly Pro Lys Trp Lys Arg Met Arg Arg Ile Cys Met	
10 15 20	
gag cac ttg ctc aca act aga cga ctt gaa cta ttt gtg agt cat agg	1470
Glu His Leu Leu Thr Thr Arg Arg Leu Glu Leu Phe Val Ser His Arg	
25 30 35	
gct gat gag gca cga cat ttg gtc caa gac gta tta act cgt tcc cac	1518
Ala Asp Glu Ala Arg His Leu Val Gln Asp Val Leu Thr Arg Ser His	
40 45 50 55	
aaa gat aaa gtt gtt aat ttg agg gaa gtg tta ggt gca ttt tct atg	1566
Lys Asp Lys Val Val Asn Leu Arg Glu Val Leu Gly Ala Phe Ser Met	
60 65 70	
aat aac gtg act aga atg ttg cta ggg aag caa tac ttt ggg gcc ggg	1614
Asn Asn Val Thr Arg Met Leu Leu Gly Lys Gln Tyr Phe Gly Ala Gly	
75 80 85	
acg gcg ggc cca caa gag gct cta gag ttt atg cat ata aca cat gag	1662
Thr Ala Gly Pro Gln Glu Ala Leu Glu Phe Met His Ile Thr His Glu	
90 95 100	
ttg ttt tgg tta cta ggc ttg att tac ttg ggt gat tat ttg cct ttt	1710
Leu Phe Trp Leu Leu Gly Leu Ile Tyr Leu Gly Asp Tyr Leu Pro Phe	
105 110 115	
tgg agg tgg gtt gat cca tat gga tgt gaa aag aaa atg agg gaa gtt	1758
Trp Arg Trp Val Asp Pro Tyr Gly Cys Glu Lys Lys Met Arg Glu Val	
120 125 130 135	
gaa aaa agg gta gat gat ttc cat cgc aaa att ata gag gaa cat agg	1806
Glu Lys Arg Val Asp Asp Phe His Arg Lys Ile Ile Glu Glu His Arg	
140 145 150	
aag gag aag aaa agg aaa gaa gaa atg gga gtg aat gag ggt gaa atg	1854
Lys Glu Lys Lys Arg Lys Glu Glu Met Gly Val Asn Glu Gly Glu Met	
155 160 165	

gat ttt gta gat att ttg ttg gct ttg cct ggt gaa aat gga aat gag Asp Phe Val Asp Ile Leu Leu Ala Leu Pro Gly Glu Asn Gly Asn Glu 170 175 180	1902
cat atg gat gat gca gat att aaa gct cta att cag gtaattcatg His Met Asp Asp Ala Asp Ile Lys Ala Leu Ile Gln 185 190 195	1948
tataatttga atgtgataca aagtttgata gaaaacatat ttgcataaat acatggttac cctactagac ccaataaaat acataattat tgccttgcta gttgaaagtt gaaacaacct agctaccatt ttgttgat tatcattagc caacaaaaac tatttcttgc atccatata taatgtttag agtcgccata tttacaatta cttgtaacat tttgagcaaa caaattaaaa tatttttgg caagtccatt ttattgaatg atacctata cttaaaatga atccttggtt atgtacactt gccttcaag gtaccaatat ttgaccatat gtaattacta ttaacaatt tgataaaatc taataatatg taaatatacg ttcacgcaca tattagaaac aaagatcaca aatgataatg caaaataact tatttgaact aatgttgtga agttaaattt ggaacaaaag gtatatttgt attgccgaat ttaatttat aaattactta taacaacaac tcaatatgta aaactgttaa gatggagtgt ggatagaatg agatgagtat acttttacta gttaccactc gaaattgcat ttcctccccc gtttatagtt gttctactt ctattatcat aaataatttt tggacttatt tcaatgtata tttacaacgt taattgtta atttttaaa aatacataat gtaaacaagg atttcatggt caattataca cattatagat attatcttaa aaaacttact aatgctcaat tagtgtccat aatatacgat atg ata gca gca gca aca gac aca Asp Met Ile Ala Ala Ala Thr Asp Thr 200	2008 2068 2128 2188 2248 2308 2368 2428 2488 2548 2608 2668 2728 2782 2830 2878 2926 2974 3022 3070 3118
tca gct gta acc aac gaa tgg gcc atg gca gaa gta ata aaa cac cca Ser Ala Val Thr Asn Glu Trp Ala Met Ala Glu Val Ile Lys His Pro 205 210 215 220	
cgt gtc ctc cac aag atc caa caa gag ctt aac aca ata gta gga ccc Arg Val Leu His Lys Ile Gln Gln Glu Leu Asn Thr Ile Val Gly Pro 225 230 235	
aat cga atg gta aca gaa tca gat ctt ccc cac ctt aac tac cta cgt Asn Arg Met Val Thr Glu Ser Asp Leu Pro His Leu Asn Tyr Leu Arg 240 245 250	
tgt gtc gta cgc gaa acg ttt cggt atg cat cca gca gga ccc ttt tta Cys Val Val Arg Glu Thr Phe Arg Met His Pro Ala Gly Pro Phe Leu 255 260 265	
atc cca cat gaa tca cta cgc cat aca aca atc aac ggc tat gat atc Ile Pro His Glu Ser Leu Arg His Thr Thr Ile Asn Gly Tyr Asp Ile 270 275 280	
cca tct ggg aca cgt gtc ttc atc aac aca cat ggg tta gga cgt aac Pro Ser Gly Thr Arg Val Phe Ile Asn Thr His Gly Leu Gly Arg Asn 285 290 295 300	
ctt aaa gtg tgg gac aac ata gag gat ttt tac cct gaa aga cat tgg	

Leu Lys Val Trp Asp Asn Ile Glu Asp Phe Tyr Pro Glu Arg His Trp		
305	310	315
ccg ttg gat gga agt aga gtt gag att agc cat gga tct gat ttt aaa		3166
Pro Leu Asp Gly Ser Arg Val Glu Ile Ser His Gly Ser Asp Phe Lys		
320	325	330
ata tta cca ttt agt gct ggg aag aga aga tgt cct ggg gcc cca ctt		3214
Ile Leu Pro Phe Ser Ala Gly Lys Arg Arg Cys Pro Gly Ala Pro Leu		
335	340	345
ggg gtg gtg ttt gtg ttg atg gga ttg gct aca ctt ttt cat gca ttt		3262
Gly Val Val Phe Val Leu Met Gly Leu Ala Thr Leu Phe His Ala Phe		
350	355	360
gat tgg tta cca cct gat gga atg aag gca gaa gaa att gat act aag		3310
Asp Trp Leu Pro Pro Asp Gly Met Lys Ala Glu Glu Ile Asp Thr Lys		
365	370	375
380		
gaa gtt tat ggg atg act atg cct aaa gct caa cct tta atg gct ttg		3358
Glu Val Tyr Gly Met Thr Met Pro Lys Ala Gln Pro Leu Met Ala Leu		
385	390	395
gct aaa cct agg ctt gct cct cat tta tat ctt tga tacatgttca		3404
Ala Lys Pro Arg Leu Ala Pro His Leu Tyr Leu		
400	405	
tatttgtggtg cacttataag cacactagct caaatagaca aatacaagct tgtattgact		3464
ctaacatgtt attaatatt agcatactgc aactctacaa ctacaagtat gtaatctcta		3524
taaattataa acataagtca taacgcacatct tgttttgaaa aaaaaagcta cactttctta		3584
caccataaac tggccgtta ctaaattctt cttggcttat tttagaatta atttttcaat		3644
tttcctattt acttataaca caagataaga aaatcgacta acaatatttc acaacattat		3704
atcttgatag ttcggtgaaa ttaataattt tcaataattt caaaacgcaa aaagcttctt		3764
tgttcggata aattcacgatc atttcaata aaatcgcaaa atttccctt gccatcaagg		3824
aacaaatttt tcacatcaa actcttcttc caaaaagtat acttctccat tgattatgta		3884
aaagagttt catgttctaa tccattagta tgtattttc ggataagaag ttgtattaca		3944
agcttaaata aagatataaa cttagttaaa tatcgaccaa tccacaatat gtatctcat		4004
atgtgtccat cagcgcccaa ctttgagatg tagatcatct aaagcacatt aacatggcgt		4064
actcctcctg ttcaaattcag ggtgtttga aaccaaactt ctccaaaata cttcctcagt		4124
aatgaagaga aaaactcaca ttctgaaaaa catcaacatc atggttctat ggtctagtg		4184
ttatgacact ggactctgaa tccagtaacc cgagtt		4220

<210> 9
<211> 4068
<212> DNA
<213> Zea mays

<220>
<221> exon

<222> (2001)..(2927)
 <223> Exon 1

 <220>
 <221> Intron
 <222> (2928)..(3017)
 <223> Intron

 <220>
 <221> exon
 <222> (3018)..(3683)
 <223> Exon 2

 <400> 9
 gagtgtggtg cgttggttt gtcagatgtc acaggaagga tgcagaccta ataccgctgc 60
 ttttagttgt ttgctgagtg ctggcttag cctggctatg ctgaatcagg gattgcaagc
 tcatacgcttat gcagtcaaca tggatggat atttgactca gctgtccatg cttcttgg 120
 gacaatgtat gcaaaatgtg gtaggttggc tgaggctcat cgtgtttct catgcattag 180
 aaatccaagc cttgttgcca ttaattctat gattcagca tttgcacaac atggcttggc 240
 tgaagatgca ctcaaacttt ttaacagaat gcaatatgtat ggccaaaggc ctaatcatgt
 gacattcttg ggaatactga ctgcatgtgc tcgagctgg tgggtcaac aaggctataa 300
 ctactttgaa tctatgagat cagtcataatgg cattcaacca aaccctgacc actacacatg 360
 tatggtaat cttaggcc atgcaggctt cttgtatgaa gcattggaaa tgattaattc
 gatgccccag aaagattatc ctgatgcattt ggcagcttg cttagctcta gtagcctcca 420
 ttctaatctt gathtagcaa aactagcagc acagaggctt cttagatgatg atccttatga
 cacaacagct tacagggtcc tgacaaacat gttctctca gcagggttga agggagatga 480
 agagatggta aaagttgcac aattgtccaa catggctagt aagaggcctg ggtatagcct
 catcatacag gataagacta cagaaaataa ctagccaaa atgagcactt ccatttacag 540
 ataactggca tgtcataact cataagacag ggatattggg ttcacagaaa acaattctag
 cgccaaagatg agatataat ttgtttagt atgcttagtgc agagtcagtt ttagtactta 600
 agaactgaga atgaggaata tctgataagg tgactagtga tttgaagaat aaagagccag
 gtggcttggaa tcaaattttt gtttgttgc cttttat cactgtttcc agatagaat 660
 tttcagttact gcaactgtatg ttgtttagtataa ataatggaaa atatgttttc tccctgtaaa
 caatagttt gtcagatgtaa aatcttatcc acagtagaaat atgatgccat ataatgctgt
 tcagattcat tacaatcgatg aagtttattt cgccttttc agagactggaa gagtttagt 720
 cacttctgca aataattcat ctgctactgt tttgttgcacaa caagaaaatc attgggttgc
 gcatgcttcc aagaatattt tgctgcgtga tctatcttattt ctcctcatca attgtatgag
 aagagtcgct gagagaggta atctttatgc ttattcatttcc atttcgtatgtaa gacaaccata
 ccaaggtgaa ttgggtgatc cctaggtcta ggctttctt ttagcaatag aagcttccaa 780
 960
 1020
 1080
 1140
 1200
 1260
 1320
 1380
 1440
 1500

gtaattcatg gaggtctaa gatgttctcc tcaagaaaaa tagggattgt ctgtgtcttt	1560
gttatagatg tggcagggtgg agtaggacta tggtaaacaa gggtttctc ttatffffgg	1620
ctgagaaatt gtagattagt gggaaacta tcctaattat ggcaacataa tgtcagcttg	1680
cttttgatt aatagtcatg tttcatgacc ttttttggtt ctagtcctct tcttttttt	1740
tccttggcat tttcccagtg atctacagtt tacattgcgc aagtcatgtc aagatcatga	1800
cagttttat atacatggtg gtagttgaga tgcttaacca tacacatcaa cacaaaaacc	1860
atgaccaatc taatactgta accaatttgc ttcaacctca gtggctcaat ttacttagta	1920
cttaagcgaa gaagctttag acgtccccata aataggccag catttgcaat ccaacttttag	1980
ctgccatctg tttctgaacc atg gat cca ttt gtt ctc tcc atc ctc tta tgc Met Asp Pro Phe Val Leu Ser Ile Leu Leu Cys	2033
1 5 10	
tca tcg atc ttt gtt gta gtg tac tgg aga agg ctg aac agc atg agg Ser Ser Ile Phe Val Val Val Tyr Trp Arg Arg Leu Asn Ser Met Arg	2081
15 20 25	
cta aga ctt cca ccg gga cct cca aca tgg cca att ttc ggc aat ctt Leu Arg Leu Pro Pro Gly Pro Pro Thr Trp Pro Ile Phe Gly Asn Leu	2129
30 35 40	
ctc cag ttg agc cct ctt ccc cac aaa gac ttt gcc caa ttt tgc acc Leu Gln Leu Ser Pro Leu Pro His Lys Asp Phe Ala Gln Phe Cys Thr	2177
45 50 55	
aaa tat ggc cct ctc gtc tat ctt cgc ctg gga acc atc gat gcc atc Lys Tyr Gly Pro Leu Val Tyr Leu Arg Leu Gly Thr Ile Asp Ala Ile	2225
60 65 70 75	
acc act gat gac ccc gaa gtg atc cgt gaa ata ctc atc cgg caa gat Thr Thr Asp Asp Pro Glu Val Ile Arg Glu Ile Leu Ile Arg Gln Asp	2273
80 85 90	
gag gtc ttt gct tcg cgg cct cgg aca ctg gct gcc gtc cat ctc gcc Glu Val Phe Ala Ser Arg Pro Arg Thr Leu Ala Ala Val His Leu Ala	2321
95 100 105	
tat ggg tgt ggt gat gtg gct cta gct cca ctg gga ccc aac tgg aaa Tyr Gly Cys Gly Asp Val Ala Leu Ala Pro Leu Gly Pro Asn Trp Lys	2369
110 115 120	
agg atg agg aga gtt tgc atg gag cac ttg ctg acg acc agg cgg ctc Arg Met Arg Arg Val Cys Met Glu His Leu Leu Thr Thr Arg Arg Leu	2417
125 130 135	
gag tct ttc gct gct cac cga gct cag gag gcc gag cac ctc tgc cag Glu Ser Phe Ala Ala His Arg Ala Gln Glu Ala Glu His Leu Cys Gln	2465
140 145 150 155	
ttt gtg tgg gct aaa tct cag tcc ggg aag ccc gtg aac ctc aga gag Phe Val Trp Ala Lys Ser Gln Ser Gly Lys Pro Val Asn Leu Arg Glu	2513
160 165 170	
gtt ctc ggt gcc ttc tcg atg aac aac gtc acg cgg atg ctg ctg ggg Val Leu Gly Ala Phe Ser Met Asn Asn Val Thr Arg Met Leu Leu Gly	2561
175 180 185	

aag cag tac ttt ggg atc cag tcg gca ggc ccc ggc gag gca atg gag Lys Gln Tyr Phe Gly Ile Gln Ser Ala Gly Pro Gly Glu Ala Met Glu 190 195 200	2609
ttc atg cac atc acc cac gag ctg ttc ttc ctg ctg ggc ctg atc tat Phe Met His Ile Thr His Glu Leu Phe Leu Leu Gly Leu Ile Tyr 205 210 215	2657
ctc ggg gac tac ttg ccg gct tgg agg tgg gtc gac ccg tac ggg tgt Leu Gly Asp Tyr Leu Pro Ala Trp Arg Trp Val Asp Pro Tyr Gly Cys 220 225 230 235	2705
gag aag agg atg agg gag gtg gag aag aag gtg gac gac ttc cac cag Glu Lys Arg Met Arg Glu Val Lys Lys Val Asp Asp Phe His Gln 240 245 250	2753
aag atc att gat gag cac agg aga gct agg gag gcc agg aag agt cgt Lys Ile Ile Asp Glu His Arg Arg Ala Arg Glu Ala Arg Lys Ser Arg 255 260 265	2801
tcc tcc gtt gag gaa gat ggc ggc aac ggc aaa gat gag atg gac ttc Ser Ser Val Glu Glu Asp Gly Gly Asn Gly Lys Asp Glu Met Asp Phe 270 275 280	2849
gtc gat gtg ctg tta tct ttg cct ggt gag aac ggg aag gag cac atg Val Asp Val Leu Leu Ser Leu Pro Gly Glu Asn Gly Lys Glu His Met 285 290 295	2897
gac gac atg gag atc aaa gcg ttg atg cag gtgtgtgtat gtgtatgctt Asp Asp Met Glu Ile Lys Ala Leu Met Gln 300 305	2947
tgctattgca taccctggaa ttcaattcat tttgttgaag ctagttatta tgagtaatac	3007
ttcttggcag gac atg atc gct gct act gat act tca tcg gtg acg Asp Met Ile Ala Ala Ala Thr Asp Thr Ser Ser Val Thr 310 315 320	3056
aac gag tgg gtg atg gcg gag gta atc aag aac ccg cgc gtg ctc cgg Asn Glu Trp Val Met Ala Glu Val Ile Lys Asn Pro Arg Val Leu Arg 325 330 335	3104
cgc gtc cag gag gag ctg gac gcg gtg gtg ggg cgc gac ccg atg gtg Arg Val Gln Glu Glu Leu Asp Ala Val Val Gly Arg Asp Arg Met Val 340 345 350	3152
gcg gag tcg gac ctg gcc cac ctc ccc tac ctc cgg tgc gtg gtg cgc Ala Glu Ser Asp Leu Ala His Leu Pro Tyr Leu Arg Cys Val Val Arg 355 360 365 370	3200
gag tca ttc cgg atg cac ccg gcg ggg ccg ttc ctt atc ccg cac gag Glu Ser Phe Arg Met His Pro Ala Gly Pro Phe Leu Ile Pro His Glu 375 380 385	3248
tcg ctg aag gcg acg acc atc atg ggg tac cac gtg ccg gcg cgc acg Ser Leu Lys Ala Thr Thr Ile Met Gly Tyr His Val Pro Ala Arg Thr 390 395 400	3296
cgc gtg ttc atc aac acg cac gcg ctg ggg ccg aac ccg cgc gtg tgg Arg Val Phe Ile Asn Thr His Ala Leu Gly Arg Asn Pro Arg Val Trp 405 410 415	3344
gac tcc gtg ggc gag ttc ccg ccg gag ccg cac ctg ccg gcg gag gag Asp Ser Val Gly Glu Phe Arg Pro Glu Arg His Leu Pro Ala Glu Glu	3392

420	425	430	
ggg gcg cgg gtg gag atc agc cac ctg ccg gac ttc aag atc ctg ccg Gly Ala Arg Val Glu Ile Ser His Leu Pro Asp Phe Lys Ile Leu Pro			3440
435 440 445 450			
ttc agc gcc ggg aag cgcc aag tgc ccc ggc gcg ccg ctg ggc gtg gcg Phe Ser Ala Gly Lys Arg Lys Cys Pro Gly Ala Pro Leu Gly Val Ala			3488
455 460 465			
ctg gtg ctc atg gcg ctc gcc agg ctc ttc cac tgc ttc gac tgg tcc Leu Val Leu Met Ala Leu Ala Arg Leu Phe His Cys Phe Asp Trp Ser			3536
470 475 480			
ccg ccc gac ggc ctc cgcc ccc gag gac gtg gac acc ccg gag gtg tac Pro Pro Asp Gly Leu Arg Pro Glu Asp Val Asp Thr Arg Glu Val Tyr			3584
485 490 495			
ggc atg acc atg ccc aag gcc acg ccc ctc gtc gcc gtc gcc act ccg Gly Met Thr Met Pro Lys Ala Thr Pro Leu Val Ala Val Ala Thr Pro			3632
500 505 510			
cgc ctg ccg ccg cac ttg tac ggc ggc ggc ggc agc tcg gct cct Arg Leu Pro Pro His Leu Tyr Gly Gly Gly Ser Ser Ala Pro			3680
515 520 525 530			
tag ttgcgtgaca ctttgacgca cgtgcgtgc actgccagtc tcattagtc			3733
ttatacgcat gatgtacttc cctccatata caatacacccg cataaaaccaa acgatgaatg			3793
aaatgttagtc gttctatggt ttcaatatga gaagatataa ttgatggta acatcatttc			3853
cctccgttta gtgataagcc atatactcta ggtatgttagt agtgtaacaa tgcgatcccc			3913
gcccgaggtaa cgaggttagca gagaggttag ctagctgttag ctggcgacga agcgaagcat			3973
taaacggttt caacggcagc tagccgtggc gttcatgcac ggtatggcca aaaaaaaagg			4033
accctatcca tctgccagtt ttggcgtacc aatgc			4068
<210> 10			
<211> 1593			
<212> DNA			
<213> Artificial Sequence			
<220>			
<223> c-DNA CYPgst gene (Zea mays)			
<400> 10			
atggatccat ttgttctctc catcctctta tgctcatcga tctttgttgt agtgtactgg			60
agaaggctga acagcatgag gctaaagactt ccaccggac ctccaaacatg gccaattttc			120
ggcaatcttc tccagtttag ccctttccc cacaaagact ttgccccatt ttgcacccaaa			180
tatggccctc tcgtctatct tcgcctggga accatcgatg ccatcaccac tgcgtacccc			240
gaagtgtatcc gtgaaatact catccggcaa gatgaggatct ttgcttcgac gcctcgacca			300
ctggctgccc tccatctcgc ctatgggtgt ggtgtatgtgg ctctagctcc actgggaccc			360
aactggaaaa ggatgaggag agttgcgtg gggacttgc tgacgaccag gcggctcgag			420

tcttcgctg	ctcaccgagc	tcaggaggcc	gagcacctct	gccagttgt	gtgggctaaa	480
tctcagtccg	ggaagcccg	taacctcaga	gaggttctcg	gtgccttctc	gatgaacaac	540
gtcacgcgga	tgctgctggg	gaagcagtac	tttggatcc	agtcggcagg	ccccggcag	600
gcaatggagt	tcatgcacat	caccacgag	ctgttcttc	tgctggcct	gatctatctc	660
ggggactact	tgccggctt	gaggtgggtc	gaccgtacg	ggtgtgagaa	gaggatgagg	720
gaggtggaga	agaagggtgga	cgacttccac	cagaagatca	ttgatgagca	caggagagct	780
agggaggcca	ggaagagtcg	ttcctccgtt	gaggaagatg	gcggcaacgg	caaagatgag	840
atggacttcg	tcgatgtgct	gttatcttg	cctggtgaga	acgggaagga	gcacatggac	900
gacatggaga	tcaaagcg	tttgcaggac	atgatcgctg	ctgctactga	tacttcatcg	960
gtgacgaacg	agtgggtgat	ggcggaggt	atcaagaacc	cgcgcgtgct	ccggcgcgtc	1020
caggaggagc	tggacgcgt	ggtggggcgc	gaccggatgg	tggcggagtc	ggacctggcc	1080
cacccccc	acctccgg	cgtggtg	gagtcattcc	ggatgcaccc	ggcggggccg	1140
ttccatcc	cgcacgagtc	gctgaaggcg	acgaccatca	tgggtacca	cgtgcggcg	1200
cgcacgcgc	tgttcatcaa	cacgcacgc	ctggggcgg	acccgcgcgt	gtgggactcc	1260
gtggcgcagt	tccggccgg	gcggcacctg	ccggcggagg	agggggcgcg	ggtggagatc	1320
agccacctgc	cggacttcaa	gatcctgccc	ttcagcgcc	ggaagcgcaa	gtgccccggc	1380
gcgcgcgtgg	gcgtggcgct	ggtgctcatg	gcgcgc	ggctttcca	ctgcttcgac	1440
tggtccccgc	ccgacggcct	ccgcggag	gacgtggaca	cccgagg	gtacggcatg	1500
accatgccc	aggccacg	cctcg	gtcgccactc	cgcgc	gccgcacttg	1560
tacggcggcg	gcggcggcag	ctcg	ttct	tag		1593

<210> 11
<211> 530
<212> PRT
<213> Zea mays

<220>
<221> PEPTIDE
<222> (1)..(530)
<223> protein CYPgst (Zea mays)

<400> 11

Met	Asp	Pro	Phe	Val	Leu	Ser	Ile	Leu	Leu	Cys	Ser	Ser	Ile	Phe	Val
1				5				10					15		

Val	Val	Tyr	Trp	Arg	Arg	Leu	Asn	Ser	Met	Arg	Leu	Arg	Leu	Pro	Pro
								20					25		30

Gly	Pro	Pro	Thr	Trp	Pro	Ile	Phe	Gly	Asn	Leu	Leu	Gln	Leu	Ser	Pro
								35					40		45

Leu Pro His Lys Asp Phe Ala Gln Phe Cys Thr Lys Tyr Gly Pro Leu
50 55 60

Val Tyr Leu Arg Leu Gly Thr Ile Asp Ala Ile Thr Thr Asp Asp Pro
65 70 75 80

Glu Val Ile Arg Glu Ile Leu Ile Arg Gln Asp Glu Val Phe Ala Ser
85 90 95

Arg Pro Arg Thr Leu Ala Ala Val His Leu Ala Tyr Gly Cys Gly Asp
100 105 110

Val Ala Leu Ala Pro Leu Gly Pro Asn Trp Lys Arg Met Arg Arg Val
115 120 125

Cys Met Glu His Leu Leu Thr Thr Arg Arg Leu Glu Ser Phe Ala Ala
130 135 140

His Arg Ala Gln Glu Ala Glu His Leu Cys Gln Phe Val Trp Ala Lys
145 150 155 160

Ser Gln Ser Gly Lys Pro Val Asn Leu Arg Glu Val Leu Gly Ala Phe
165 170 175

Ser Met Asn Asn Val Thr Arg Met Leu Leu Gly Lys Gln Tyr Phe Gly
180 185 190

Ile Gln Ser Ala Gly Pro Gly Glu Ala Met Glu Phe Met His Ile Thr
195 200 205

His Glu Leu Phe Phe Leu Leu Gly Leu Ile Tyr Leu Gly Asp Tyr Leu
210 215 220

Pro Ala Trp Arg Trp Val Asp Pro Tyr Gly Cys Glu Lys Arg Met Arg
225 230 235 240

Glu Val Glu Lys Lys Val Asp Asp Phe His Gln Lys Ile Ile Asp Glu
245 250 255

His Arg Arg Ala Arg Glu Ala Arg Lys Ser Arg Ser Ser Val Glu Glu
260 265 270

Asp Gly Gly Asn Gly Lys Asp Glu Met Asp Phe Val Asp Val Leu Leu
275 280 285

Ser Leu Pro Gly Glu Asn Gly Lys Glu His Met Asp Asp Met Glu Ile
290 295 300

Lys Ala Leu Met Gln Asp Met Ile Ala Ala Ala Thr Asp Thr Ser Ser
305 310 315 320

Val Thr Asn Glu Trp Val Met Ala Glu Val Ile Lys Asn Pro Arg Val
325 330 335

Leu Arg Arg Val Gln Glu Glu Leu Asp Ala Val Val Gly Arg Asp Arg
340 345 350

Met Val Ala Glu Ser Asp Leu Ala His Leu Pro Tyr Leu Arg Cys Val
355 360 365

Val Arg Glu Ser Phe Arg Met His Pro Ala Gly Pro Phe Leu Ile Pro
370 375 380

His Glu Ser Leu Lys Ala Thr Thr Ile Met Gly Tyr His Val Pro Ala
385 390 395 400

Arg Thr Arg Val Phe Ile Asn Thr His Ala Leu Gly Arg Asn Pro Arg
405 410 415

Val Trp Asp Ser Val Gly Glu Phe Arg Pro Glu Arg His Leu Pro Ala
420 425 430

Glu Glu Gly Ala Arg Val Glu Ile Ser His Leu Pro Asp Phe Lys Ile
435 440 445

Leu Pro Phe Ser Ala Gly Lys Arg Lys Cys Pro Gly Ala Pro Leu Gly
450 455 460

Val Ala Leu Val Leu Met Ala Leu Ala Arg Leu Phe His Cys Phe Asp
465 470 475 480

Trp Ser Pro Pro Asp Gly Leu Arg Pro Glu Asp Val Asp Thr Arg Glu
485 490 495

Val Tyr Gly Met Thr Met Pro Lys Ala Thr Pro Leu Val Ala Val Ala
500 505 510

Thr Pro Arg Leu Pro Pro His Leu Tyr Gly Gly Gly Gly Ser Ser
515 520 525

Ala Pro
530

<210> 12
<211> 4880

<212> DNA
 <213> Solanum tuberosum

<220>
 <221> exon
 <222> (1762)..(2032)
 <223> Exon 1

<220>
 <221> Intron
 <222> (2033)..(2448)
 <223> Intron 1

<220>
 <221> exon
 <222> (2449)..(3161)
 <223> Exon 2

<220>
 <221> Intron
 <222> (3162)..(4031)
 <223> Intron 2

<220>
 <221> exon
 <222> (4032)..(4694)
 <223> Exon 3

<400> 12	
ataataaaata tttttttaa aaaaagagag ggttgttagt tgcaaggta taagtgagca	60
aaaaggtgga tggaggata attttagacc aaaaagatga gtagaagggt attttagac	120
caaaagatgg atgaaggata ttttagacc atttcctgta tttcagaggt attttgcc	180
cttttccgta tgaaaattt aggttgtatt taagttcaat tggtcaaatt aataagtatt	240
tttaagactg tcaaaaattt agaaataaaaa ctaataattc acatcaattc tacaaatatt	300
tcttttaat atatagacat tcgggtcaag taaaatattt taagttttta tttaatttct	360
tgttgtttg attatggatt gccttggatt ctatgcgt aatggaaatg gcagctctt	420
tcagtgtagg attggacaag ttagcttggaa tttttttca cctactctac aaaaagtgtat	480
ggggcgcca ccaaccatac gtcaacaact ttgtccattc ttgtacaca ctaccatcac	540
aatctatcaa aagatttgtt cttagtctc gtttaacata ttcttcggtt ttcatttaa	600
tttattttgtc ttatttttt ttataaaata aaataaaatat tttttgatat aatacataaa	660
tatattcttt aatttcattt catttaacat ctctatactt taacttggga gtcgagtgtat	720
tgggttggaa gtcaggtcgt gattagaggt taggagttag gtcccaactc ggaattgaga	780
ttagaggtcc agattagatc tcaggttaga gtcgggagtt gggcttaag ttgtggcag	840
gatttgggtc ctatgttggg tcacaagtca ggtccccgt cgagatcgaa aattaggtcc	900
cgagttgaga tagaaagttt tttgtatcaag attgaggtcg tagtcaagag tcgagttctg	960
agtccggatt gtaatcgaaa gtgggtcattt acaggttagg gccccgggtcg aaagtcgggt	1020

ctcaaggtag gactcgttt agggtggtaa ggtttaagta ggggtgtgca tcgatcggtt	1080
cggtttatcg atttcagtt tttaaatatg ctaaattaaat aaccaaatac aaaggaaaa	1140
tttatcgatt ttggtcatta atgactcgat tttctattt accaattaga aaatgctcat	1200
aaaacaataa gatgacttct ctaacaaaat ttgaccgaca agacaagaca ataatgtgat	1260
tattatttttta ggtttgacg ttttgtataa tgtgaagtgt gaattgaagg tttaaaggca	1320
aagacagtaa ctagtaatat attgagatta atatttatgt aataagtaaa agaagtaaat	1380
tactggagta taatcttatt gagttattgt tttacccat aacacaatat taaaaatcaa	1440
tatcgaaatcg ataacccgat atttttttta aaaaaaaaaatt gttaacccaa ttgatgcatt	1500
tttttatata ttcgtaaaat gcgtaaattt cattaaattt ttcccataaa taaaaaagaa	1560
acttgttcat aaattcaaac ttatttgca aattgaggaa gcaaaaaagt gtgatttcaa	1620
gattaatact tatagttgaa acatgtcaag atgatggct tcctaattgc tccaaaaaaa	1680
gattataataat tagaattaaa catatttcct tattaagcct ttgtaaaacta cttctttctt	1740
tttttgaca aataattaaa g atg att gac ttg act agt ttt gtt att gtc Met Ile Asp Leu Thr Ser Phe Val Ile Val 1 5 10	1791
ctt ctt tgc acg tat ctt ctt aat ttg atc aat tat agt ata gtc ctt Leu Leu Cys Thr Tyr Leu Leu Asn Leu Ile Asn Tyr Ser Ile Val Leu 15 20 25	1839
ttt ggc gca tat ctt att tcc aag cta ctt cat ttt tca ttc gtc gat Phe Gly Ala Tyr Leu Ile Ser Lys Leu Leu His Phe Ser Phe Val Asp 30 35 40	1887
aag tcg aat cga gaa atc aat caa ctc cct cct ggt cca aaa caa tgg Lys Ser Asn Arg Glu Ile Asn Gln Leu Pro Pro Gly Pro Lys Gln Trp 45 50 55	1935
cct att gta ggc aac ctt ttt cag tta ggg caa tta cct cat cga gac Pro Ile Val Gly Asn Leu Phe Gln Leu Gly Gln Leu Pro His Arg Asp 60 65 70	1983
atg gcg tct ttc tgc gaa aaa tat ggc cca ttg gtc tat ctc cga cta g Met Ala Ser Phe Cys Glu Lys Tyr Gly Pro Leu Val Tyr Leu Arg Leu 75 80 85 90	2032
gtaatgttga tacaataataat tcacaaatga ccacaatttta gtccttgtaa ttaataataa gtcattgttt tcgagtttca aatttataag ggaaacttca agtacacttg acttggagat	2092
tcacttatgg atgttgcac tcgatgaaca attactattt tacaaaaaaa tgatcgaaaa	2152
gtgataagtg ataagtgata agtgaacggtt atttctcata taaagttcaa tgctatacag	2212
attcggtaaa actcagtaaa ttttacctaa gtactttatc catgttaaga aaaaatcatt	2272
atataatgtac acacattaaa ttttaaaccg agttattagc actagaattt atcgttctaa	2332
cattttgaac ccataaaagtt gatatcgtga ctccacctt tgggttctct taaaag gt Gly	2392

aat gtt gat gct atc acc acc aat gat cca gaa atc ata aga gaa ata Asn Val Asp Ala Ile Thr Thr Asn Asp Pro Glu Ile Ile Arg Glu Ile 95 100 105	2498
ctt gta caa caa gat gat gtt ttt gca tct agg cca aga act ctt gct Leu Val Gln Gln Asp Asp Val Phe Ala Ser Arg Pro Arg Thr Leu Ala 110 115 120	2546
gcc att cat cta gct tat ggt tgt ggg gat gtg gca ttg gct cct tta Ala Ile His Leu Ala Tyr Gly Cys Gly Asp Val Ala Leu Ala Pro Leu 125 130 135	2594
ggt cca aaa tgg aaa aga atg aga aga ata tgt atg gaa cat tta ttg Gly Pro Lys Trp Lys Arg Met Arg Arg Ile Cys Met Glu His Leu Leu 140 145 150 155	2642
aca act aaa aga ctt gaa tca ttt gca aaa cat agg gca gat gaa gcc Thr Thr Lys Arg Leu Glu Ser Phe Ala Lys His Arg Ala Asp Glu Ala 160 165 170	2690
caa agt cta gtt aaa gat att tgg acc aaa gcc caa aaa gga caa ata Gln Ser Leu Val Lys Asp Ile Trp Thr Lys Ala Gln Lys Gly Gln Ile 175 180 185	2738
gtg aat ttg agg gaa gtt ttg ggt gga ttt tca atg aat aat gtg act Val Asn Leu Arg Glu Val Leu Gly Gly Phe Ser Met Asn Asn Val Thr 190 195 200	2786
aga atg ttg tta ggt aaa caa tat ttt ggg gca gaa tca gca ggt cca Arg Met Leu Leu Gly Lys Gln Tyr Phe Gly Ala Glu Ser Ala Gly Pro 205 210 215	2834
caa gaa gca atg gaa ttt atg cat gta aca cat gag tta ttt tgg tta Gln Glu Ala Met Glu Phe Met His Val Thr His Glu Leu Phe Trp Leu 220 225 230 235	2882
ctt gga gtg ata tat tta ggt gat tat tta cct ttt tgg agg tgg att Leu Gly Val Ile Tyr Leu Gly Asp Tyr Leu Pro Phe Trp Arg Trp Ile 240 245 250	2930
gat cct tat ggt tgt gag aaa aaa atg agg gat gtt gaa aaa agg att Asp Pro Tyr Gly Cys Glu Lys Lys Met Arg Asp Val Glu Lys Arg Ile 255 260 265	2978
gat gat ttt cat atg aga ata att gaa gaa cat aga aag aag aaa ggt Asp Asp Phe His Met Arg Ile Ile Glu Glu His Arg Lys Lys Lys Gly 270 275 280	3026
aat aaa aat aat aat att gat gat gat gaa atg gac ttt gtg gat Asn Lys Asn Asn Asn Ile Asp Asp Asp Glu Met Asp Phe Val Asp 285 290 295	3074
gtt tta ttg tct cta cca gga gaa gat gaa gga gat ggt aat gga aaa Val Leu Leu Ser Leu Pro Gly Glu Asp Glu Gly Asp Gly Asn Gly Lys 300 305 310 315	3122
caa aat atg gat gta gag att aaa gct cta att cag gtctaatttt Gln Asn Met Asp Asp Val Glu Ile Lys Ala Leu Ile Gln 320 325	3171
tttttatatat accaactttc tacctattta atatgtcaat gcatttatct caaactactt	3231
atatatatat ccaatatata tgcatttggt gagcaacaac taccacccat gggatttcta	3291

agcttgatta acagagggtt gggtaagt ataagtactt cttcatccct aactagaggt	3351
cttggggtcg agtcttgctg gatacaaagt cgtcttgaa aaagagtgtt accccataat	3411
gtgagacttt ccggcacgaa tcaaaattt gttggaatct aatgggtat cgaacaccag	3471
ttgagaaaag aaagagaaga agaaattcct acaatgtggg actttcggga gcgaacccaa	3531
attttgttgg actctaata gggatggga cacccgtaga aaaaaaaaga aataaattca	3591
caagtttgc tatatactaa aacataaaat ttaagttata tatactgata gagtaaaaca	3651
aattacatta ttagtccaat ttaatcggtt ctaccatgtt attactctat taaatttata	3711
tatgagactt aatatacgaga gttacatgtc aacttgatag tgttagacatt tttagactat	3771
tcatgcacaa aaacttaact ctaatacaat caatttcctt gcaactttt tattagagaa	3831
tatatatata tatatatata tatagtacag tatattttaa catgtataaa	3891
caagatgcct attttatattt cagtttata ctaatcacaa tatatatttg ctacatcaat	3951
agtataaaaa aaagctaaat atggttcttt tttctgctgt gacttatttc ttttctttaa	4011
ttaaattata ggatatgata gct gca gcc aca gac act tct gct gtg acc aac	4064
Ala Ala Ala Thr Asp Thr Ser Ala Val Thr Asn	
330 335	
gaa tgg gca atg gct gag gta atc aga cat cca cat gtc ctc aaa aag	4112
Glu Trp Ala Met Ala Glu Val Ile Arg His Pro His Val Leu Lys Lys	
340 345 350 355	
atc caa gaa ctc gat ata gtt gtc ggg tcg ggt cggt atg gta acc	4160
Ile Gln Glu Glu Leu Asp Ile Val Val Gly Ser Gly Arg Met Val Thr	
360 365 370	
gaa tcc gac ttg atc cat ctc aag tac ctc cgt tgt gta gta cgt gaa	4208
Glu Ser Asp Leu Ile His Leu Lys Tyr Leu Arg Cys Val Val Arg Glu	
375 380 385	
aca ttc cga atg cac cct gcg ggt cca ttc cta atc cca cat gaa tca	4256
Thr Phe Arg Met His Pro Ala Gly Pro Phe Leu Ile Pro His Glu Ser	
390 395 400	
att cgc gat act atg atc aac ggc tat tac atc ccg gcc aag aca cgc	4304
Ile Arg Asp Thr Met Ile Asn Gly Tyr Tyr Ile Pro Ala Lys Thr Arg	
405 410 415	
gtg ttc atc aac aca cat ggt ctt ggc cgg aac aca aag att tgg gac	4352
Val Phe Ile Asn Thr His Gly Leu Gly Arg Asn Thr Lys Ile Trp Asp	
420 425 430 435	
aac ata gat gag ttt agg cca gag aga cat tta cca cca aat gat gat	4400
Asn Ile Asp Glu Phe Arg Pro Glu Arg His Leu Pro Pro Asn Asp Asp	
440 445 450	
gaa aaa aac atg atc atg act act agt agt aga gtt gag att agt	4448
Glu Lys Asn Met Ile Met Thr Thr Ser Ser Arg Val Glu Ile Ser	
455 460 465	
cat ggt cca gat ttc aag att ttg cca ttt agt gct gga aaa agg aag	4496
His Gly Pro Asp Phe Lys Ile Leu Pro Phe Ser Ala Gly Lys Arg Lys	
470 475 480	

tgt cct ggt gca cca ttg ggt gtg aaa ttg gtg ctt atg gca ttg gct		4544
Cys Pro Gly Ala Pro Leu Gly Val Lys Leu Val Leu Met Ala Leu Ala		
485 490 495		
agg ttg ttt cat tgc tat gat tgg agt cca cca aat gga gta aag cat		4592
Arg Leu Phe His Cys Tyr Asp Trp Ser Pro Pro Asn Gly Val Lys His		
500 505 510 515		
caa gat att gac aca aat gaa gtt tat gga atg act atg cct aaa gct		4640
Gln Asp Ile Asp Thr Asn Glu Val Tyr Gly Met Thr Met Pro Lys Ala		
520 525 530		
aag cca ttg atg gct att gct aaa cct aga ctg cct gct cac ttg tac		4688
Lys Pro Leu Met Ala Ile Ala Lys Pro Arg Leu Pro Ala His Leu Tyr		
535 540 545		
caa taa ttagttacta gtactcaa at caaaggaa tggtcttaa tggcttcgt		4744
Gln		
aacgaatcaa ataaaagagaa atttctattt gtacctcgta aaaaaaaagct ctgcgcctcct		4804
gtaatagtat acttcaacat ttttctcacc tgtggcaatc cacacagcag acttaagtct		4864
gtctgattac atagta		4880
<210> 13		
<211> 1656		
<212> DNA		
<213> Artificial Sequence		
<220>		
<223> cDNA CYPgst gene (Solanum tuberosum)		
<400> 13		
atgattgact tgacttagttt tgttattgtc cttcttgca cgtatcttct taatttgatc		60
aattataatgt tagtcctttt tggcgcatat cttatttcca agctacttca ttttcattc		120
gtcgataagt cgaatcgaga aatcaatcaa ctccctcctg gtccaaaaca atggcctatt		180
gtaggcaacc ttttcagttt aggcaatta cctcatcgag acatggcgctc tttctgcgaa		240
aaatatggcc cattggctca tctccgacta ggtaatgtt atgctatcac caccaatgt		300
ccagaaaatca taagagaaat acttgtacaa caagatgatg ttttgcatc taggccaaga		360
actcttgctg ccattcatct agcttatgggt tggggatg tggcattggc tccttaggt		420
ccaaaaatgga aaagaatgag aagaatatgt atgaaacatt tattgacaac taaaagactt		480
aatcatttg caaaacatag ggcagatgaa gcccaaagtc tagttaaaga tatttggacc		540
aaagccccaaa aaggacaaat agtgaatttg agggaaagttt tgggtggatt ttcaatgaat		600
aatgtgacta gaatgttgtt aggtaaacaa tattttgggg cagaatcagc aggtccacaa		660
gaagcaatgg aatttatgca tgtaacacat gagttatccc ggttacttgg agtgatataat		720
tttaggtgatt atttaccttt ttggaggtgg attgatcctt atggttgtga gaaaaaaaaatg		780
aggatgttgg aaaaaaggat tggatgtttt catatgagaa taattgaaga acatagaaag		840

aagaaaaggta ataaaaataa taataatatt gatgatgatg aaatggactt tgtggatgtt	900
ttattgtctc taccaggaga agatgaagga gatggtaatg gaaaacaaaa tatggatgt	960
gtagagatta aagctctaattcaggatatg atagctgcag ccacagacac ttctgctgt	1020
accaacgaat gggcaatggc tgaggtaatc agacatccac atgtcctcaa aaagatccaa	1080
gaagaactcg atatagttgt cgggtcgggt cgatggtaa ccgaatccga cttgatccat	1140
ctcaagtacc tccgttgtt agtacgtgaa acattccgaa tgacccctgc gggccattc	1200
ctaattccac atgaatcaat tcgcgatact atgatcaacg gctattacat cccggccaag	1260
acacgcgtgt tcatcaacac acatggtctt ggccggaaca caaagatttg ggacaacata	1320
gatgagtttggccagagag acatttacca ccaaattatg atgaaaaaaaaa catgatcatg	1380
actactagta gtagtagagt tgagattagt catggccag atttcaagat tttgccattt	1440
agtgctggaa aaaggaagtg tcctggtgca ccattgggtg tggaaattgggt gcttatggca	1500
ttggctaggt tgtttcattt ctatgattgg agtccaccaa atggagtaaa gcatcaagat	1560
attgacacaa atgaagtttggaaatgact atgcctaaag ctaagccatt gatggctatt	1620
gcttaaaccta gactgcctgc tcacttgtac caataa	1656

<210> 14
 <211> 551
 <212> PRT
 <213> Solanum tuberosum

<220>
 <221> PEPTID
 <222> (1)..(551)
 <223> protein CYPgst (Solanum tuberosum)

<400> 14

Met Ile Asp Leu Thr Ser Phe Val Ile Val Leu Leu Cys Thr Tyr Leu			
1	5	10	15

Leu Asn Leu Ile Asn Tyr Ser Ile Val Leu Phe Gly Ala Tyr Leu Ile			
20	25	30	

Ser Lys Leu Leu His Phe Ser Phe Val Asp Lys Ser Asn Arg Glu Ile			
35	40	45	

Asn Gln Leu Pro Pro Gly Pro Lys Gln Trp Pro Ile Val Gly Asn Leu			
50	55	60	

Phe Gln Leu Gly Gln Leu Pro His Arg Asp Met Ala Ser Phe Cys Glu			
65	70	75	80

Lys Tyr Gly Pro Leu Val Tyr Leu Arg Leu Gly Asn Val Asp Ala Ile			
85	90	95	

Thr Thr Asn Asp Pro Glu Ile Ile Arg Glu Ile Leu Val Gln Gln Asp
100 105 110

Asp Val Phe Ala Ser Arg Pro Arg Thr Leu Ala Ala Ile His Leu Ala
115 120 125

Tyr Gly Cys Gly Asp Val Ala Leu Ala Pro Leu Gly Pro Lys Trp Lys
130 135 140

Arg Met Arg Arg Ile Cys Met Glu His Leu Leu Thr Thr Lys Arg Leu
145 150 155 160

Glu Ser Phe Ala Lys His Arg Ala Asp Glu Ala Gln Ser Leu Val Lys
165 170 175

Asp Ile Trp Thr Lys Ala Gln Lys Gly Gln Ile Val Asn Leu Arg Glu
180 185 190

Val Leu Gly Gly Phe Ser Met Asn Asn Val Thr Arg Met Leu Leu Gly
195 200 205

Lys Gln Tyr Phe Gly Ala Glu Ser Ala Gly Pro Gln Glu Ala Met Glu
210 215 220

Phe Met His Val Thr His Glu Leu Phe Trp Leu Leu Gly Val Ile Tyr
225 230 235 240

Leu Gly Asp Tyr Leu Pro Phe Trp Arg Trp Ile Asp Pro Tyr Gly Cys
245 250 255

Glu Lys Lys Met Arg Asp Val Glu Lys Arg Ile Asp Asp Phe His Met
260 265 270

Arg Ile Ile Glu Glu His Arg Lys Lys Lys Gly Asn Lys Asn Asn Asn
275 280 285

Asn Ile Asp Asp Asp Glu Met Asp Phe Val Asp Val Leu Leu Ser Leu
290 295 300

Pro Gly Glu Asp Glu Gly Asp Gly Asn Gly Lys Gln Asn Met Asp Asp
305 310 315 320

Val Glu Ile Lys Ala Leu Ile Gln Asp Met Ile Ala Ala Ala Thr Asp
325 330 335

Thr Ser Ala Val Thr Asn Glu Trp Ala Met Ala Glu Val Ile Arg His
340 345 350

Pro His Val Leu Lys Lys Ile Gln Glu Glu Leu Asp Ile Val Val Gly
355 360 365

Ser Gly Arg Met Val Thr Glu Ser Asp Leu Ile His Leu Lys Tyr Leu
370 375 380

Arg Cys Val Val Arg Glu Thr Phe Arg Met His Pro Ala Gly Pro Phe
385 390 395 400

Leu Ile Pro His Glu Ser Ile Arg Asp Thr Met Ile Asn Gly Tyr Tyr
405 410 415

Ile Pro Ala Lys Thr Arg Val Phe Ile Asn Thr His Gly Leu Gly Arg
420 425 430

Asn Thr Lys Ile Trp Asp Asn Ile Asp Glu Phe Arg Pro Glu Arg His
435 440 445

Leu Pro Pro Asn Asp Asp Glu Lys Asn Met Ile Met Thr Thr Ser Ser
450 455 460

Ser Arg Val Glu Ile Ser His Gly Pro Asp Phe Lys Ile Leu Pro Phe
465 470 475 480

Ser Ala Gly Lys Arg Lys Cys Pro Gly Ala Pro Leu Gly Val Lys Leu
485 490 495

Val Leu Met Ala Leu Ala Arg Leu Phe His Cys Tyr Asp Trp Ser Pro
500 505 510

Pro Asn Gly Val Lys His Gln Asp Ile Asp Thr Asn Glu Val Tyr Gly
515 520 525

Met Thr Met Pro Lys Ala Lys Pro Leu Met Ala Ile Ala Lys Pro Arg
530 535 540

Leu Pro Ala His Leu Tyr Gln
545 550

<210> 15
<211> 525
<212> PRT
<213> Triticum aestivum

<220>
<221> PEPTID
<222> (1)..(525)
<223> protein CYPgst (Triticum aestivum)

<400> 15

Met Asp Pro Phe Leu Leu Ser Ile Ile Leu Cys Ser Cys Ile Phe Ala
1 5 10 15

Ala Val Ser Trp Lys Lys Leu Asn Gly Met Arg Leu Arg Leu Pro Pro
20 25 30

Gly Pro Pro Arg Trp Pro Ile Phe Gly Asn Leu Leu Gln Leu Ser Pro
35 40 45

Leu Pro His Lys Asp Phe Ala Arg Phe Cys Thr Lys Tyr Gly Pro Leu
50 55 60

Val Tyr Leu Arg Leu Gly Thr Ile Asp Ala Ile Thr Thr Asp Asp Pro
65 70 75 80

Glu Val Ile Arg Glu Ile Leu Ile Arg Gln Asp Glu Val Phe Ala Ser
85 90 95

Arg Pro Arg Thr Leu Ala Ala Val His Leu Ala Tyr Gly Cys Gly Asp
100 105 110

Val Ala Leu Ala Pro Leu Gly Pro Asn Trp Lys Arg Met Arg Arg Val
115 120 125

Cys Met Glu His Leu Leu Thr Thr Arg Arg Leu Glu Ser Phe Ala Ala
130 135 140

His Arg Ala Glu Glu Ala Glu His Leu Cys Glu Phe Val Trp Ala Lys
145 150 155 160

Ser Gln Ser Gly Lys Pro Val Asn Leu Arg Glu Val Leu Gly Ala Phe
165 170 175

Ser Met Asn Asn Val Thr Arg Met Leu Leu Gly Lys Gln Tyr Phe Gly
180 185 190

Leu Gln Ser Ala Gly Pro Gly Glu Ala Met Glu Phe Met His Ile Thr
195 200 205

His Glu Leu Phe Phe Leu Leu Gly Leu Ile Tyr Leu Gly Asp Tyr Leu
210 215 220

Pro Ala Trp Arg Trp Leu Asp Pro Tyr Gly Cys Glu Lys Lys Met Arg
225 230 235 240

Glu Val Glu Lys Lys Val Asp Asp Phe His Gln Lys Ile Ile Asp Glu

245 250 255

His Arg Lys Ala Arg Asp Val Arg Lys Ser Gly Ala Ser Leu Asp Asp
260 265 270

Asp Gly Asp Asp Ser Lys Glu Gly Met Asp Phe Val Asp Val Leu Leu
275 280 285

Ser Leu Pro Gly Glu Asn Gly Asn Glu His Met Asp Asp Val Glu Ile
290 295 300

Lys Ala Leu Met Gln Asp Met Ile Ala Ala Ala Thr Asp Thr Ser Ser
305 310 315 320

Val Thr Asn Glu Trp Val Met Ala Glu Val Ile Lys Asn Pro Arg Val
325 330 335

Leu Arg Lys Ile Gln Glu Glu Leu Asp Ala Val Val Gly Thr Ser Arg
340 345 350

Pro His Gly Gly Gly Pro Pro Pro Pro Asp Val Pro Pro Leu
355 360 365

Arg Arg Pro Gly Val Leu Pro Asp Ala Pro Gly Gly Ala Ile Pro Asp
370 375 380

Pro Ala Arg Val Thr Gln Gly Asp Asp His His Gly Leu Arg His Pro
385 390 395 400

Gly Ala Asp Glu Asp Leu His Gln His Pro Arg Ala Gly Pro Glu Pro
405 410 415

Ala His Leu Gly Arg Arg Arg Val Pro Pro Arg Glu Ala Pro Pro
420 425 430

Gly Gly Arg Arg Ala Arg Gly Asp Gln Pro Pro Ala Gly Leu Gln Asp
435 440 445

Pro Ala Leu Gln Arg Arg Gln Ala Gln Val Pro Arg Gly Ala Ala Gly
450 455 460

Arg Asp Pro Gly Ala His Gly Ala Arg Gln Ala Leu Pro Leu Leu Arg
465 470 475 480

Leu Val Pro Ala Arg Arg Pro Pro Pro Arg Gly His Arg His Arg Arg
485 490 495

Gly Leu Arg Asp Asp His Ala Gln Gly Gln Ala Ala His Arg Arg Arg

500

505

510

Ser Thr Ala Pro Ala Ala Asp Val Arg Leu Leu Ser
515 520 525

<210> 16
<211> 518
<212> PRT
<213> Helianthus annuus

<220>
<221> PEPTID
<222> (1)..(518)
<223> protein CYPgst (Helianthus annuus)

<400> 16

Met Asp Leu Gln Ser Phe Ile Ser Val Leu Ala Phe Ile Val Ala Ser
1 5 10 15

Arg Ile Ile Leu Leu Trp Tyr Val Lys Gln Arg Val Thr Gln Asn Thr
20 25 30

Thr His Arg Leu Pro Pro Gly Pro Pro Arg Trp Pro Ile Val Gly Asn
35 40 45

Leu Leu Gln Leu Gly Pro Leu Pro His Arg Asp Leu Ala Ser Phe Cys
50 55 60

Glu Arg Tyr Gly Pro Leu Val Tyr Leu Arg Leu Gly Lys Val Asp Ala
65 70 75 80

Ile Thr Thr Asn Asp Pro Asn Ile Ile Arg Glu Ile Leu Val Lys Gln
85 90 95

Asp Asp Val Phe Ala Ser Arg Pro Gln Thr Leu Ala Ala Val His Leu
100 105 110

Ala Tyr Asn Cys Gly Asp Val Ala Leu Ala Pro Phe Gly Pro Arg Trp
115 120 125

Lys Trp Met Arg Arg Ile Cys Met Glu Gln Leu Leu Thr Thr Lys Arg
130 135 140

Leu Glu Ser Phe Ala Lys Gln Arg Ala Ser Glu Ala Gln His Leu Val
145 150 155 160

Gln Asp Val Trp Ala Leu Ser Gln Ala Asn Gly Pro Ile Asn Leu Arg
165 170 175

Glu Val Leu Gly Gly Phe Ser Met Asn Asn Val Thr Arg Met Leu Leu
180 185 190

Gly Lys Gln Tyr Phe Gly Ser Gly Ser Ala Gly Pro Lys Glu Ala Thr
195 200 205

Glu Phe Met His Ile Thr His Glu Leu Phe Trp Leu Leu Gly Leu Ile
210 215 220

Tyr Leu Gly Asp Tyr Leu Pro Phe Trp Arg Trp Ile Asp Pro Tyr Gly
225 230 235 240

Cys Glu Lys Lys Met Arg Glu Val Glu Lys Arg Val Asp Asp Phe His
245 250 255

Met Lys Ile Ile Glu Glu His Arg Gln Arg Arg Lys Asn Gly Glu Gln
260 265 270

Lys Asp Glu Gly Ile Met Asp Phe Val Asp Val Leu Leu Ser Leu Pro
275 280 285

Gly Glu Asp Gly Lys Asp His Met Asp Asp Arg Gln Ile Lys Ala Leu
290 295 300

Val Gln Asp Met Ile Ala Ala Ala Thr Asp Thr Ser Ala Val Thr Asn
305 310 315 320

Glu Trp Ala Met Ala Glu Val Ile Lys His Pro His Val Leu Arg Lys
325 330 335

Ile Gln Glu Glu Leu Asp Asn Val Val Gly Pro Asp Arg Met Val Ser
340 345 350

Glu Ser Asp Leu Ser Asn Leu Asn Tyr Leu Arg Cys Val Val Arg Glu
355 360 365

Thr Phe Arg Met His Pro Ala Gly Pro Phe Leu Ile Pro His Glu Ser
370 375 380

Leu Arg Ala Thr Glu Ile Asn Gly Tyr Tyr Ile Pro Ala Lys Thr Arg
385 390 395 400

Val Phe Ile Asn Thr His Gly Leu Gly Arg Asn Thr Val Leu Trp Asp
405 410 415

Asp Ile Asn Val Phe Arg Pro Glu Arg His Leu Thr Ser Asp Gly Ser
420 425 430

Arg Val Glu Ile Ser His Gly Asp Asp Phe Lys Ile Leu Pro Phe Ser
435 440 445

Ala Gly Lys Arg Lys Cys Pro Gly Ala Pro Leu Gly Val Thr Leu Val
450 455 460

Leu Met Ala Leu Ala Arg Leu Phe His Cys Phe Asp Trp Ser Pro Pro
465 470 475 480

Asp Gly Leu Lys Cys Glu Asp Ile Asp Thr Gln Glu Ile Tyr Gly Met
485 490 495

Thr Met Pro Lys Ala Lys Pro Leu Met Ala Val Ala Lys Pro Arg Leu
500 505 510

Ala Ser Tyr Met Tyr Gln
515

<210> 17
<211> 530
<212> PRT
<213> Hordeum vulgare

<220>
<221> PEPTID
<222> (1)..(530)
<223> protein CYPgst (Hordeum vulgare)

<400> 17

Met Asp Leu Phe Leu Phe Ser Ile Ile Leu Cys Ser Cys Ile Phe Ala
1 5 10 15

Ala Val Ser Trp Arg Lys Leu Ser Arg Leu Arg Leu Arg Leu Pro Pro
20 25 30

Gly Pro Pro Arg Trp Pro Ile Phe Gly Asn Leu Leu Gln Leu Ser Pro
35 40 45

Leu Pro His Lys Asp Phe Ala Arg Phe Cys Thr Lys Tyr Gly Pro Leu
50 55 60

Val Tyr Leu Arg Leu Gly Thr Ile Asp Ala Ile Thr Thr Asp Asp Pro
65 70 75 80

Glu Val Ile Arg Glu Ile Leu Val Arg Gln Asp Glu Val Phe Ala Ser
85 90 95

Arg Pro Arg Thr Leu Ala Ala Val His Leu Ala Tyr Gly Cys Gly Asp
100 105 110

Val Ala Leu Ala Pro Leu Gly Pro Asn Trp Lys Arg Met Arg Arg Val
115 120 125

Cys Met Glu His Leu Leu Thr Thr Lys Arg Leu Glu Ser Phe Ala Ala
130 135 140

His Arg Ala Gln Glu Ala Glu His Leu Cys Glu Phe Val Trp Ala Lys
145 150 155 160

Ser Gln Ser Gly Lys Pro Val Asn Leu Arg Glu Val Leu Gly Ala Phe
165 170 175

Ser Met Asn Asn Val Thr Arg Met Leu Leu Gly Lys Gln Tyr Phe Gly
180 185 190

Leu Gln Ser Ala Gly Pro Gly Glu Ala Met Glu Phe Met His Ile Thr
195 200 205

His Glu Leu Phe Phe Leu Leu Gly Leu Ile Tyr Leu Gly Asp Tyr Leu
210 215 220

Pro Ala Trp Arg Trp Val Asp Pro Tyr Gly Cys Glu Lys Lys Met Arg
225 230 235 240

Glu Val Glu Lys Lys Val Asp Asp Phe His Gln Lys Ile Ile Asp Glu
245 250 255

His Arg Lys Ala Arg Asp Ala Arg Lys Ser Ala Ala Ser Leu Asp Asp
260 265 270

Gly Asp Asp Ser Lys Glu Asp Met Asp Phe Val Asp Val Leu Leu Ser
275 280 285

Leu Pro Gly Glu Asn Gly Asn Glu His Met Asp Asp Val Glu Ile Lys
290 295 300

Ala Leu Met Gln Asp Met Ile Ala Ala Ala Thr Asp Thr Ser Ser Val
305 310 315 320

Thr Asn Glu Trp Val Met Ala Glu Val Ile Lys Asn Pro Cys Val Leu
325 330 335

Arg Lys Ile Gln Glu Glu Leu Asp Ala Val Val Gly Arg Ser Arg Met
340 345 350

Val Val Glu Ser Asp Leu Pro His Leu Thr Tyr Leu Arg Cys Val Val
355 360 365

Arg Glu Ser Phe Arg Met His Pro Ala Gly Pro Phe Leu Ile Pro His
370 375 380

Glu Ser Leu Lys Ala Thr Thr Ile Met Gly Tyr Asp Ile Pro Ala Gln
385 390 395 400

Thr Arg Ile Phe Ile Asn Thr His Ala Leu Gly Arg Asn Pro Arg Ile
405 410 415

Trp Asp Asp Val Gly Glu Phe His Pro Glu Arg His Leu Pro Ala Asp
420 425 430

Gly Gly Arg Val Glu Ile Ser His Leu Pro Asp Phe Lys Ile Leu Pro
435 440 445

Phe Ser Ala Gly Lys Arg Lys Cys Pro Gly Ala Pro Leu Gly Val Ile
450 455 460

Leu Val Leu Met Ala Leu Ala Arg Leu Phe His Cys Phe Asp Trp Ser
465 470 475 480

Pro Pro Asp Gly Leu Arg Pro Glu Asp Ile Asp Thr Asp Glu Val Tyr
485 490 495

Gly Met Thr Met Pro Lys Ala Lys Pro Leu Ile Ala Ala Val Gln Pro
500 505 510

Arg Leu Pro Pro Gln Met Tyr Gly Ser Cys Pro Ser His Gly Met Gln
515 520 525

Met Gln
530

<210> 18
<211> 466
<212> PRT
<213> Brassica napus

<220>
<221> PEPTID
<222> (1)..(466)
<223> protein CYPgst (Brassica napus)

<400> 18

Met Ala Ala Leu Cys Ser Lys Tyr Gly Pro Leu Val Tyr Leu Arg Leu
1 5 10 15

Gly Asn Ile Asp Ala Ile Thr Thr Asn Asp Pro Glu Thr Ile Arg Glu
20 25 30

Ile Leu Phe Arg Gln Asp Asp Val Phe Ala Ser Arg Pro Lys Thr Leu
35 40 45

Ala Ala Val His Leu Ala Tyr Gly Cys Gly Asp Val Ala Leu Ala Pro
50 55 60

Met Gly Pro His Trp Lys Arg Met Arg Arg Ile Cys Met Glu His Leu
65 70 75 80

Leu Thr Thr Lys Arg Leu Glu Ser Phe Thr Ser Gln Arg Ala Glu Glu
85 90 95

Ala Gln Tyr Leu Ile Gln Asp Val Cys Lys Arg Ala Glu Cys Gly Lys
100 105 110

Pro Ile Asn Leu Arg Glu Val Leu Gly Ala Phe Ser Met Asn Asn Val
115 120 125

Thr Arg Met Leu Leu Gly Lys Gln Phe Phe Gly Pro Gly Ser Val Val
130 135 140

Gly Ala Lys Glu Ala Gln Glu Phe Met His Ile Thr His Lys Leu Phe
145 150 155 160

Arg Leu Leu Gly Val Ile Tyr Leu Gly Asp Tyr Leu Pro Phe Trp Arg
165 170 175

Trp Val Asp Pro Tyr Gly Cys Glu Lys Glu Met Arg Asp Val Glu Lys
180 185 190

Arg Val Asp Lys Phe His Thr Lys Ile Ile Glu Glu His Arg Arg Ala
195 200 205

Lys Arg Glu Lys Glu Asp Lys Asn Ile Glu Gly Asp Met Asp Phe Val
210 215 220

Asp Val Leu Leu Ser Leu Pro Gly Glu Asn Gly Lys Glu His Met Asp
225 230 235 240

Asp Val Glu Ile Lys Ala Leu Ile Gln Asp Met Ile Ala Ala Thr
245 250 255

Asp Thr Ser Ala Val Thr Asn Glu Trp Ala Met Ala Glu Val Ile Lys
260 265 270

Gln Pro Arg Val Met Arg Lys Ile Gln Glu Glu Leu Asp Asn Val Val
275 280 285

Gly Ser Asn Arg Met Val Asn Glu Thr Asp Leu Val His Leu Asn Tyr
290 295 300

Leu Arg Cys Val Val Arg Glu Thr Phe Arg Met His Pro Ala Gly Pro
305 310 315 320

Phe Leu Ile Pro His Glu Ser Val Arg Pro Thr Thr Ile Asn Gly Tyr
325 330 335

Tyr Ile Pro Ala Lys Thr Arg Val Phe Ile Asn Thr His Gly Leu Gly
340 345 350

Arg Asn Thr Ser Val Trp Thr Thr Asp Ile Glu Glu Phe Arg Pro Glu
355 360 365

Arg His Trp Pro Val Asp Gly Ser Gly Arg Val Glu Ile Ser His Gly
370 375 380

Pro Asp Tyr Lys Ile Leu Pro Phe Ser Ala Gly Lys Arg Lys Cys Pro
385 390 395 400

Gly Ala Pro Leu Gly Val Thr Met Val Leu Met Ala Leu Ala Arg Leu
405 410 415

Phe His Cys Phe Asp Trp Thr Thr Pro Glu Asp Ile Asp Thr Val Glu
420 425 430

Val Tyr Gly Met Thr Met Pro Lys Ala Lys Pro Leu Trp Ala Leu Ala
435 440 445

Lys Pro Arg Leu Ala Ala His Leu Tyr Thr Ile Thr His Asp Thr Ile
450 455 460

Gly His
465

<210> 19
<211> 530
<212> PRT
<213> Brassica oleracea

<220>
<221> PEPTID
<222> (1)..(530)
<223> protein CYPgst (Brassica oleracea)

<400> 19

Met Asp Leu Phe Leu Leu Ser Ile Ile Leu Cys Ser Trp Ile Phe Val

1

5

10

15

Ala Val Tyr Trp Lys Lys Leu Asn Arg Thr Lys Leu Arg Leu Pro Pro
20 25 30

Gly Pro Pro Arg Trp Pro Ile Phe Gly Asn Leu Leu Gln Leu Ser Pro
35 40 45

Leu Pro His Lys Asp Phe Ala Arg Phe Cys Thr Lys Tyr Gly Pro Leu
50 55 60

Val Tyr Leu Arg Leu Gly Thr Ile Asp Ala Ile Thr Thr Asp Asp Pro
65 70 75 80

Glu Val Ile Arg Glu Ile Leu Ile Arg Gln Asp Glu Val Phe Ala Ser
85 90 95

Arg Pro Arg Thr Leu Ala Ala Val His Leu Ala Tyr Gly Cys Gly Asp
100 105 110

Val Ala Leu Ala Pro Leu Gly Pro Asn Trp Lys Arg Met Arg Arg Val
115 120 125

Cys Met Glu His Leu Leu Thr Thr Lys Arg Leu Glu Ser Phe Ala Ala
130 135 140

His Arg Ala Gln Glu Ala Glu His Leu Cys Gln Phe Val Trp Ala Lys
145 150 155 160

Ser Gln Ser Glu Lys Pro Val Asn Leu Arg Glu Val Leu Gly Ala Phe
165 170 175

Ser Met Asn Asn Val Thr Arg Met Leu Leu Gly Lys Gln Tyr Phe Gly
180 185 190

Leu Gln Ser Ala Gly Pro Gly Glu Ala Met Glu Phe Met His Ile Thr
195 200 205

His Glu Leu Phe Tyr Leu Leu Gly Leu Ile Tyr Leu Gly Asp Tyr Leu
210 215 220

Pro Ala Trp Arg Trp Val Asp Pro Tyr Gly Cys Glu Lys Lys Met Arg
225 230 235 240

Glu Val Glu Lys Lys Val Asp Asp Phe His Gln Lys Ile Ile Asp Glu
245 250 255

His Arg Lys Ala Arg Glu Ala Arg Lys Ser Ala Ser Ser Leu Asp Asp

260

265

270

Gly Asp Asp Ser Lys Glu Glu Met Asp Phe Val Asp Val Leu Leu Ser
 275 280 285

Leu Pro Gly Glu Asn Gly Lys Glu His Met Asp Asp Val Glu Ile Lys
 290 295 300

Ala Leu Met Gln Asp Met Ile Ala Ala Ala Thr Asp Thr Ser Ser Val
 305 310 315 320

Thr Asn Glu Trp Val Met Ala Glu Val Ile Lys His Pro Arg Val Leu
 325 330 335

Arg Lys Ile Gln Glu Glu Leu Asp Ala Val Val Gly Arg Ala Arg Met
 340 345 350

Val Ser Glu Ser Asp Leu Pro His Leu Pro Tyr Leu Arg Cys Val Val
 355 360 365

Arg Glu Ser Phe Arg Met His Pro Ala Gly Pro Phe Leu Ile Pro His
 370 375 380

Glu Ser Leu Lys Pro Thr Thr Ile Met Gly Tyr Asp Ile Pro Ala Arg
 385 390 395 400

Thr Arg Ile Phe Ile Asn Thr His Ala Leu Gly Arg Asn Pro Arg Val
 405 410 415

Trp Asp Asp Val Gly Gln Phe Arg Pro Glu Arg His Met Pro Ala Asp
 420 425 430

Gly Gly Ala Arg Val Glu Ile Ser His Leu Pro Asp Phe Lys Ile Leu
 435 440 445

Pro Phe Ser Ala Gly Lys Arg Lys Cys Pro Gly Ala Pro Leu Gly Val
 450 455 460

Ile Leu Val Leu Met Ala Leu Ala Arg Leu Phe His Cys Phe Asp Trp
 465 470 475 480

Ser Pro Pro Asp Gly Glu Glu Ile Asp Thr Asp Glu Val Tyr Gly Met
 485 490 495

Thr Met Pro Lys Ala Leu Pro Leu Phe Ala Ala Ala Arg Pro Arg Leu
 500 505 510

Pro Pro Glu Met Tyr His Gly Ser Ser Cys Pro Ser His Gly Lys Gln

515

520

525

Thr Met
530

<210> 20
<211> 466
<212> PRT
<213> Brassica rapa

<220>
<221> PEPTID
<222> (1)..(466)
<223> protein CYPgst (Brassica rapa)

<400> 20

Met Ala Ala Leu Cys Ser Lys Tyr Gly Pro Leu Val Tyr Leu Arg Leu
1 5 10 15

Gly Asn Ile Asp Ala Ile Thr Thr Asn Asp Pro Glu Thr Ile Arg Glu
20 25 30

Ile Leu Phe Arg Gln Asp Asp Val Phe Ala Ser Arg Pro Lys Thr Leu
35 40 45

Ala Ala Val His Leu Ala Tyr Gly Cys Gly Asp Val Ala Leu Ala Pro
50 55 60

Met Gly Pro His Trp Lys Arg Met Arg Arg Ile Cys Met Glu His Leu
65 70 75 80

Leu Thr Thr Lys Arg Leu Glu Ser Phe Thr Ser Gln Arg Ala Glu Glu
85 90 95

Ala Gln Tyr Leu Ile Gln Asp Val Cys Lys Arg Ala Glu Cys Gly Lys
100 105 110

Pro Ile Asn Leu Arg Glu Val Leu Gly Ala Phe Ser Met Asn Asn Val
115 120 125

Thr Arg Met Leu Leu Gly Lys Gln Phe Phe Gly Pro Gly Ser Val Val
130 135 140

Gly Ala Lys Glu Ala Gln Glu Phe Met His Ile Thr His Lys Leu Phe
145 150 155 160

Arg Leu Leu Gly Val Ile Tyr Leu Gly Asp Tyr Leu Pro Phe Trp Arg
165 170 175

Trp Val Asp Pro Tyr Gly Cys Glu Lys Glu Met Arg Asp Val Glu Lys
180 185 190

Arg Val Asp Lys Phe His Thr Lys Ile Ile Glu Glu His Arg Arg Ala
195 200 205

Lys Arg Glu Lys Glu Asp Lys Asn Ile Glu Gly Asp Met Asp Phe Val
210 215 220

Asp Val Leu Leu Ser Leu Pro Gly Glu Asn Gly Lys Glu His Met Asp
225 230 235 240

Asp Val Glu Ile Lys Ala Leu Ile Gln Asp Met Ile Ala Ala Ala Thr
245 250 255

Asp Thr Ser Ala Val Thr Asn Glu Trp Ala Met Ala Glu Val Ile Lys
260 265 270

Gln Pro Arg Val Met Arg Lys Ile Gln Glu Glu Leu Asp Asn Val Val
275 280 285

Gly Ser Asn Arg Met Val Asn Glu Thr Asp Leu Val His Leu Asn Tyr
290 295 300

Leu Arg Cys Val Val Arg Glu Thr Phe Arg Met His Pro Ala Gly Pro
305 310 315 320

Phe Leu Ile Pro His Glu Ser Val Arg Pro Thr Thr Ile Asn Gly Tyr
325 330 335

Tyr Ile Pro Ala Lys Thr Arg Val Phe Ile Asn Thr His Gly Leu Gly
340 345 350

Arg Asn Thr Ser Val Trp Thr Thr Asp Ile Glu Glu Phe Arg Pro Glu
355 360 365

Arg His Trp Pro Val Asp Gly Ser Gly Arg Val Glu Ile Ser His Gly
370 375 380

Pro Asp Tyr Lys Ile Leu Pro Phe Ser Ala Gly Lys Arg Lys Cys Pro
385 390 395 400

Gly Ala Pro Leu Gly Val Thr Met Val Leu Met Ala Leu Ala Arg Leu
405 410 415

Phe His Cys Phe Asp Trp Thr Thr Pro Glu Asp Ile Asp Thr Val Glu
420 425 430

Val Tyr Gly Met Thr Met Pro Lys Ala Lys Pro Leu Trp Ala Leu Ala
435 440 445

Lys Pro Arg Leu Ala Ala His Leu Tyr Thr Ile Thr His Asp Thr Ile
450 455 460

Gly His
465

<210> 21
<211> 523
<212> PRT
<213> Glycine max

<220>
<221> PEPTID
<222> (1)..(523)
<223> protein CYPgst (Glycine max)

<400> 21

Met Asp Leu Thr Thr Phe Ile Ser Thr Leu Phe Leu Gly Thr Leu Ala
1 5 10 15

Ser Arg Ile Ile Arg His Trp Leu Ile Gly Arg Ser Leu Ser Ser His
20 25 30

Lys Asn Lys Leu Pro Pro Gly Pro Pro Arg Trp Pro Ile Val Gly Asn
35 40 45

Leu Leu Gln Leu Gly Gln Leu Pro His Arg Asp Leu Ala Ser Leu Cys
50 55 60

Asp Lys Tyr Gly Pro Leu Val Tyr Leu Lys Leu Gly Lys Ile Asp Ala
65 70 75 80

Ile Thr Thr Asn Asp Pro Asp Ile Ile Arg Glu Ile Leu Leu Ser Gln
85 90 95

Asp Asp Val Phe Ala Ser Arg Pro His Thr Phe Ala Ala Val His Leu
100 105 110

Ala Tyr Gly Cys Gly Asp Val Ala Leu Ala Pro Leu Gly Pro His Trp
115 120 125

Lys Arg Met Arg Arg Ile Cys Met Glu His Leu Leu Thr Thr Lys Arg
130 135 140

Leu Glu Ser Phe Ser Asn His Arg Leu Asp Glu Ala Gln His Leu Val
145 150 155 160

Lys Asp Val Met Ala Trp Ala Gln Asp Lys Lys Pro Ile Asn Leu Arg
165 170 175

Glu Val Leu Gly Ala Phe Ser Met Asn Asn Val Thr Arg Met Leu Leu
180 185 190

Gly Lys Gln Tyr Phe Gly Ser Glu Ser Ser Gly Pro Gln Glu Ala Met
195 200 205

Glu Phe Met His Ile Thr His Glu Leu Phe Trp Leu Leu Gly Val Ile
210 215 220

Tyr Leu Gly Asp Tyr Leu Pro Ile Trp Arg Trp Val Asp Pro Tyr Gly
225 230 235 240

Cys Glu Lys Lys Met Arg Glu Val Glu Lys Arg Val Asp Asp Phe His
245 250 255

Ser Asn Ile Ile Glu Glu His Arg Lys Ala Arg Lys Asp Arg Lys Gly
260 265 270

Lys Arg Lys Glu Gly Asp Gly Asp Met Asp Phe Val Asp Val Leu Leu
275 280 285

Ser Leu Pro Gly Glu Asp Gly Lys Glu His Met Asp Asp Val Glu Ile
290 295 300

Lys Ala Leu Ile Gln Asp Met Ile Ala Ala Ala Thr Asp Thr Ser Ala
305 310 315 320

Val Thr Asn Glu Trp Ala Met Ala Glu Val Met Lys His Pro His Val
325 330 335

Leu His Lys Ile Gln Glu Glu Leu Asp Thr Ile Val Gly Pro Asn Arg
340 345 350

Met Val Leu Glu Ser Asp Leu Pro His Leu Asn Tyr Leu Arg Cys Val
355 360 365

Val Arg Glu Thr Phe Arg Met His Pro Ala Gly Pro Phe Leu Ile Pro
370 375 380

His Glu Ser Leu Arg Ala Thr Thr Ile Asn Gly Tyr His Ile Pro Ala
385 390 395 400

Lys Thr Arg Val Phe Ile Asn Thr His Gly Leu Gly Arg Asn Thr Lys
405 410 415

Ile Trp Asp Asn Val Asp Glu Phe Arg Pro Glu Arg His Trp Pro Ser
420 425 430

Asn Gly Asn Gly Thr Arg Val Glu Ile Ser His Gly Val Asp Phe Lys
435 440 445

Ile Leu Pro Phe Ser Ala Gly Lys Arg Lys Cys Pro Gly Ala Pro Leu
450 455 460

Gly Val Thr Leu Val Leu Met Ala Leu Ala Arg Leu Phe His Cys Phe
465 470 475 480

Asp Trp Glu Pro Pro Lys Gly Leu Ser Cys Gly Asp Val Asp Thr Arg
485 490 495

Glu Val Tyr Gly Met Thr Met Pro Lys Ala Glu Pro Leu Ile Ala Ile
500 505 510

Ala Lys Pro Arg Leu Ala Lys His Leu Tyr Asp
515 520

<210> 22
<211> 537
<212> PRT
<213> Gossypium raimondii

<220>
<221> PEPTID
<222> (1)..(537)
<223> protein CYPgst (Gossypium raimondii)

<400> 22

Met Ile Pro Tyr Lys Pro Ile Phe Phe Pro Leu Asp Asn Met Glu Leu
1 5 10 15

Phe Thr Phe Ala Leu Ala Leu Val Gly Ala Leu Val Val Asn Ala
20 25 30

Leu Trp Arg Trp Arg Leu Asp Trp Lys Ser Leu Phe Lys Thr Arg Lys
35 40 45

Leu Pro Pro Gly Pro Pro Arg Trp Pro Ile Val Gly Asn Leu Leu Gln
50 55 60

Leu Ser Ser Leu Pro His Arg Asp Leu Ala Ser Leu Cys Asp Lys Tyr
65 70 75 80

Gly Pro Leu Val Tyr Leu Arg Leu Gly Lys Val Asp Ala Ile Thr Thr
85 90 95

Asn Asp Pro Asp Ile Ile Arg Glu Ile Leu Leu Arg Gln Asp Glu Val
100 105 110

Phe Ala Ser Arg Pro Arg Thr Leu Ala Ala Val His Leu Ala Tyr Gly
115 120 125

Cys Gly Asp Val Ala Leu Ala Pro Leu Gly Pro His Trp Lys Arg Met
130 135 140

Arg Arg Ile Cys Met Glu His Leu Leu Thr Thr Lys Arg Leu Glu Ser
145 150 155 160

Phe Ala Lys His Arg Ala Asp Glu Ala Gln His Leu Val Arg Asp Val
165 170 175

Ser Ala Arg Ala Glu Asn Gly Gln Leu Val Asn Leu Arg Glu Val Leu
180 185 190

Gly Ala Phe Ser Met Asn Asn Val Thr Arg Met Leu Leu Gly Arg Gln
195 200 205

Tyr Phe Gly Ala Val Ser Ala Gly Pro Ser Glu Ala Met Glu Phe Met
210 215 220

His Ile Thr His Glu Leu Phe Trp Leu Leu Gly Val Ile Tyr Leu Gly
225 230 235 240

Asp Tyr Leu Pro Ile Trp Arg Trp Val Asp Pro Tyr Gly Cys Glu Lys
245 250 255

Arg Met Arg Glu Val Glu Lys Arg Val Asp Asp Phe His Glu Arg Ile
260 265 270

Ile Glu Glu His Arg Arg Ala Arg Glu Leu Lys Asn Lys Gly Tyr Gly
275 280 285

Lys Asp Asp Asp Tyr Gly Glu Glu Met Asp Phe Val Asp Val Leu Leu
290 295 300

Ser Leu Pro Gly Glu Asp Gly Asn Pro His Met Asp Asp Thr Asp Ile
305 310 315 320

Lys Ala Leu Ile Gln Asp Met Ile Ala Ala Ala Thr Asp Thr Ser Ala
325 330 335

Val Thr Asn Glu Trp Thr Met Ala Glu Val Ile Lys His Pro Arg Val
340 345 350

Leu Arg Lys Ile Gln Asp Glu Leu Asp Ser Val Val Gly Pro Asn Arg
355 360 365

Met Val Asn Glu Ser Asp Leu Pro His Leu Asn Tyr Leu Arg Cys Val
370 375 380

Val Arg Glu Thr Phe Arg Met His Pro Ala Gly Pro Phe Leu Ile Pro
385 390 395 400

His Glu Ser Leu Arg Ala Thr Thr Ile Asn Gly Phe Tyr Ile Pro Ala
405 410 415

Lys Thr Arg Val Phe Ile Asn Thr His Gly Leu Gly Arg Asn Thr Lys
420 425 430

Leu Trp Asp Asp Val Glu Ser Phe Arg Pro Glu Arg His Trp Leu Ala
435 440 445

Asp Gly Ala Arg Val Glu Ile Ser His Gly Ala Asp Phe Lys Ile Leu
450 455 460

Pro Phe Ser Ala Gly Lys Arg Lys Cys Pro Gly Ala Pro Leu Gly Val
465 470 475 480

Thr Leu Val Leu Met Ala Leu Ala Arg Leu Phe His Cys Phe Asp Trp
485 490 495

Ala Pro Gln Asn Gly Met Arg Pro Glu Asp Ile Asn Thr Met Glu Val
500 505 510

Tyr Gly Met Thr Met Pro Lys Ala Glu Pro Leu Met Ala Met Ala Lys
515 520 525

Pro Arg Leu Ala Asp His Val Met Phe
530 535

<210> 23
<211> 526
<212> PRT
<213> Sorghum bicolor

<220>
<221> PEPTID
<222> (1)..(526)
<223> protein CYPgst (Sorghum bicolor)

<400> 23

Met Asp Pro Phe Val Leu Ser Ile Leu Ile Cys Ser Trp Ile Phe Val

1

5

10

15

Val Val Tyr Trp Arg Arg Leu Asn Ser Met Arg Leu Arg Leu Pro Pro
20 25 30

Gly Pro Pro Thr Trp Pro Ile Phe Gly Asn Leu Leu Gln Leu Ser Pro
35 40 45

Leu Pro His Lys Asp Phe Ala Arg Phe Cys Thr Lys Tyr Gly Pro Leu
50 55 60

Val Tyr Leu Arg Leu Gly Thr Ile Asp Ala Ile Thr Thr Asp Asp Pro
65 70 75 80

Glu Val Ile Arg Glu Ile Leu Ile Arg Gln Asp Glu Val Phe Ala Ser
85 90 95

Arg Pro Arg Thr Leu Ala Ala Val His Leu Ala Tyr Gly Cys Gly Asp
100 105 110

Val Ala Leu Ala Pro Leu Gly Pro Asn Trp Lys Arg Met Arg Arg Val
115 120 125

Cys Met Glu His Leu Leu Thr Thr Lys Arg Leu Glu Ser Phe Ala Ala
130 135 140

His Arg Ala Gln Glu Ala Glu His Leu Cys Gln Phe Val Trp Ala Lys
145 150 155 160

Ser His Ser Gly Lys Pro Val Asn Leu Arg Glu Val Leu Gly Ala Phe
165 170 175

Ser Met Asn Asn Val Thr Arg Met Leu Leu Gly Lys Gln Tyr Phe Gly
180 185 190

Ile Gln Ser Ala Gly Pro Gly Glu Ala Met Glu Phe Met His Ile Thr
195 200 205

His Glu Leu Phe Phe Leu Leu Gly Leu Ile Tyr Leu Gly Asp Tyr Leu
210 215 220

Pro Ala Trp Arg Trp Val Asp Pro Tyr Gly Cys Glu Lys Lys Met Arg
225 230 235 240

Asp Val Glu Lys Lys Val Asp Asp Phe His Gln Lys Ile Ile Asp Glu
245 250 255

His Arg Arg Ala Arg Glu Ala Lys Lys Thr Arg Arg Ser Ser Leu Asp

260

265

270

Asp Asp Asp Gly Lys Glu Asp Met Asp Phe Val Asp Val Leu Leu Ser
 275 280 285

Leu Pro Gly Glu Asn Gly Lys Glu His Met Asp Asp Met Glu Ile Lys
 290 295 300

Ala Leu Met Gln Asp Met Ile Ala Ala Ala Thr Asp Thr Ser Ser Val
 305 310 315 320

Thr Asn Glu Trp Val Met Ala Glu Val Ile Lys Asn Pro Arg Val Leu
 325 330 335

Arg Arg Val Gln Glu Glu Leu Asp Ala Val Ile Gly Arg Asp Arg Met
 340 345 350

Val Ala Glu Ser Asp Leu Thr His Leu Pro Tyr Leu Arg Cys Val Val
 355 360 365

Arg Glu Ser Phe Arg Met His Pro Ala Gly Pro Phe Leu Ile Pro His
 370 375 380

Glu Ser Leu Lys Pro Thr Thr Ile Met Gly Tyr His Val Pro Ala Arg
 385 390 395 400

Thr Arg Val Phe Ile Asn Thr His Ala Leu Gly Arg Asn Pro Arg Val
 405 410 415

Trp Asp Asp Val Asp Ala Phe Arg Pro Glu Arg His Leu Pro Ala Glu
 420 425 430

Glu Gly Ala Arg Val Glu Ile Ser His Leu Pro Asp Phe Lys Ile Leu
 435 440 445

Pro Phe Ser Ala Gly Lys Arg Lys Cys Pro Gly Ala Pro Leu Gly Val
 450 455 460

Ala Leu Val Leu Met Ala Leu Ala Arg Leu Phe His Cys Phe Asp Trp
 465 470 475 480

Ser Pro Pro Asp Gly Leu Arg Pro Glu Asp Val Asp Thr Gln Glu Val
 485 490 495

Tyr Gly Met Thr Met Pro Lys Ala Thr Pro Leu Val Ala Val Ala Thr
 500 505 510

Pro Arg Leu Pro Pro His Leu Tyr Gly Gly Ser Ala Ser

515

520

525

<210> 24
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> sxn2151s01-Markersequenz zum Anzeigen des gst-Locus in Beta vulgaris

<400> 24
aacactcatg taataagatg ggtgagttag tagcaaaaaa a 41

<210> 25
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> sxn2151s01-Markersequenz zum Anzeigen des KWS2320-Referenzgenotyps

<400> 25
aacactcatg taataagatg agtgagttag tagcaaaaaa a 41

<210> 26
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> sle3305s02-Markersequenz zum Anzeigen des gst-Locus in Beta vulgaris

<400> 26
aaagtctaga gtaaatttag gttgcagtgg agtgggaagt c 41

<210> 27
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> sle3305s02-Markersequenz zum Anzeigen des KWS2320-Referenzgenotyps

<400> 27
aaagtctaga gtaaatttag attgcagtgg agtgggaagt c 41

Формула изобретения

1. Растение, в частности культурное растение, проявляющее рецессивный ядерно-кодируемый мужски стерильный фенотип, отличающееся тем, что по сравнению с соответствующим (мужски фертильным) растением дикого типа данный фенотип обусловливается мутацией гена эндогенной цитохром Р450 оксидазы (*CYPgst*) или отсутствием, низким содержанием или низкой активностью функционального белка *CYPgst*, кодируемого геном *CYPgst* дикого типа, при этом ген *CYPgst* представляет собой ген, выбираемый из:

(а) *BvCYPgst Beta vulgaris*, подвид *vulgaris*, который предпочтительно содержит одну из нуклеотидных последовательностей SEQ ID No. 1 или 2 или кодирует аминокислотную последовательность, представленную в SEQ ID No. 3, либо его гомолога, аналога или ортолога;

(б) *StCYPgst Solanum tuberosum*, который предпочтительно содержит одну из нуклеотидных последовательностей SEQ ID No. 12 или 13 или кодирует аминокислотную последовательность, представленную в SEQ ID No. 14, либо его гомолога, аналога или ортолога;

(с) *ZmCYPgst Zea mays*, который предпочтительно содержит одну из нуклеотидных последовательностей SEQ ID No. 9 или 10 или кодирует аминокислотную последовательность, представленную в SEQ ID No. 11, либо его гомолога, аналога или ортолога.

2. Растение по п. 1, отличающееся тем, что оно является гетерозиготным по мутации и мужски фертильным или гомозиготным по мутации и мужски стерильным, при этом образование пыльцы супрессируется у стерильного растения.

3. Растение по пп. 1 или 2, отличающееся тем, что мутация предотвращает транскрипцию и/или трансляцию функционального белка, при этом мутация предпочтительно представляет собой делецию, присоединение, инсерцию или замещение в кодирующей последовательности нуклеиновой кислоты гена *CYPgst*, последовательности сигнала для сплайсинга или регуляторной последовательности, предпочтительно последовательности промотора гена *CYPgst*, делеция у *Beta vulgaris* предпочтительно детектируется путем определения отсутствия одного или обоих маркерных локусов sle5983d14 (продукт амплификации праймера, имеющего последовательности SEQ ID No. 4 и 5) и sle5983d17 (продукт амплификации праймера, имеющего последовательности SEQ ID No. 6 и 7) и присутствия убиквитарного маркера, ген локализуется в сегменте хромосомы 1 между маркерными локусами sxn215s01 и sle3305s02, последовательность маркера sxn215s01, представленная в SEQ ID No. 24, и последовательность маркера sle3305s02, представленная в SEQ ID No. 26, указывают на присутствие локуса *gst*, а последовательность маркера sxn215s01, представленная в SEQ ID No. 25, и последовательность маркера sle3305s02, представленная в SEQ ID No. 27, указывают на присутствие референсной последовательности.

4. Растение по любому из пп. 1-3, отличающееся тем, что не мутировавший ген (ген дикого типа) имеет нуклеотидную последовательность, выбиаемую из группы, состоящей из:

- (a) нуклеотидной последовательности, представленной в SEQ ID NO: 2, или ее функционального фрагмента;
- (b) нуклеотидной последовательности, которая кодирует аминокислотную последовательность, представленную в SEQ ID NO: 3;
- (c) нуклеотидной последовательности, способной гибридизоваться в жестких условиях с нуклеотидной последовательностью, комплементарной нуклеотидной последовательности по пп. (a) или (b);
- (d) нуклеотидной последовательности, которая кодирует аминокислотную последовательность, имеющую различия с аминокислотной последовательностью SEQ ID NO: 3, проявляющиеся в форме аминокислотных делеций, замещений, присоединений и/или инсерций в аминокислотной последовательности, и которая предпочтительно идентична по меньшей мере на 60% аминокислотной последовательностью по всей ее длине;
- (e) нуклеотидной последовательности, которая кодирует белок, обладающий одной и той же ферментативной активностью, что и белок, кодируемый нуклеотидной последовательностью согласно любой из нуклеотидных последовательностей (a) – (d);
- (f) нуклеотидной последовательности, содержащей по меньшей мере 200 или 400, предпочтительно по меньшей мере 600 или 800, наиболее предпочтительно по меньшей мере 1000 последовательных нуклеотидов промотора последовательности нуклеиновой кислоты SEQ ID No. 1 в позициях от 1 до 1518, предпочтительно от 518 до 1518, наиболее предпочтительно от 1318 до 1518, либо последовательности, которая гибридизуется в этом участке, причем нуклеотидная последовательность способна специфически контролировать экспрессию гена или гетерологичной молекулы нуклеиновой кислоты, оперативно связанной с нуклеотидной последовательностью, в цветках и/или плодах.

5. Растение по любому из пп. 1-4, отличающееся тем, что растение является инбредным или гибридным растением, предпочтительно растением рода *Zea*, *Solanum*, *Triticum*, *Triticale*, *Helianthus*, *Secale*, *Hordeum*, *Brassica*, *Brachypodium*, *Glycine*, *Gossypium*, *Sorghum*, *Saccharum*, *Setaria*, *Aegilops*, *Oryza*, *Daucus*, *Eucalyptus*, *Erythranthe*, *Genlisea*, *Musa*, *Avena*, *Nicotiana*, *Coffea*, *Vitis*, *Cucumis*, *Morus*, *Crucihimalaya*, *Cardamine*, *Lepidium*, *Capsella*, *Olimarabidopsis*, *Arabis*, *Raphanus*, *Eruca*, *Citrus*, *Jatropha*, *Populus* или *Beta*, предпочтительно растение вида *Zea mays*, *Solanum tuberosum*, *Triticum aestivum*, *Triticum durum*, *Triticum spelta*, *Helianthus annuus*, *Secale cereal*, *Hordeum vulgare*, *Hordeum bulbosum*, *Brassica napus*, *Brassica oleracea*, *Brassica rapa*, *Brassica juncea*, *Brassica nigra*, *Glycine max*, *Gossypium sp.*, *Sorghum bicolor*, *Triticale*, *Saccharum officinarum*, *Setaria italica*, *Oryza sativa*, *Oryza minuta*, *Oryza australiensis*, *Oryza alta*, *Brachypodium distachyon*, *Hordeum marinum*, *Aegilops tauschii*, *Daucus glochidiantus*, *Daucus pusillus*, *Daucus muricatus*, *Daucus carota*, *Eucalyptus grandis*, *Erythranthe guttata*, *Genlisea aurea*, *Musa sp.*, *Avena sp.*, *Nicotiana sylvestris*, *Nicotiana tabacum*, *Nicotiana tomentosiformis*, *Solanum lycopersicum*, *Coffea canephora*, *Vitis vinifera*, *Cucumis sativus*, *Morus nobilis*, *Crucihimalaya himalaica*, *Crucihimalaya wallichii*, *Cardamine flexuosa*, *Lepidium virginicum*, *Capsella bursa-pastoris*, *Olimarabidopsis pumila*, *Arabis hirsute*, *Raphanus sativus*, *Eruca vesicaria sativa*, *Citrus sinensis*, *Jatropha curcas*, *Populus trichocarpa* или *Beta vulgaris*.

6. Молекула нуклеиновой кислоты или рекомбинантная молекула ДНК, имеющая нуклеотидную последовательность по п. 4, где рекомбинантная молекула ДНК предпочтительно содержит (i) промотор с нуклеотидной последовательностью по п. 4 (f), оперативно связанный с гетерологичной молекулой нукleinовой кислоты или (ii) кодирующую нуклеотидную последовательность по п. 4 (а) – (е), оперативно связанную с гетерологичным промотором, и предпочтительно способная контролировать экспрессию нуклеотидной последовательности в закрытых цветках и/или плодах.

7. Рекомбинантная молекула ДНК, содержащая нуклеотидную последовательность, кодирующую shРНК (*малую РНК, образующую шпильки*), siРНК (*малую интерферирующую РНК*), отрицательно-полярную РНК, положительно-полярную РНК или двухцепочечную РНК, которая после ее экспрессии или введения в растительную клетку приводит к ингибированию экспрессии функционального (не мутировавшего) гена CYPgst.

8. Молекула нукleinовой кислоты, содержащая нуклеотидную последовательность по п. 4, с мутацией в форме делеции, присоединения, инсерции, замещения, при этом данная мутация не приводит к синтезу функционального белка CYPgst.

9. Молекула нукleinовой кислоты, состоящая по меньшей мере из 15, 16, 17, 18, 19 или 20, предпочтительно по меньшей мере из 21, 22, 23, 24 или 25, более предпочтительно по меньшей мере из 30, 35, 40, 45 или 50, наиболее предпочтительно по меньшей мере из 100, 200, 300, 500 или 1000 нуклеотидов, которая специфически гибридизуется с нуклеотидной последовательностью по пп. 4 или 8, предпочтительно в форме олигонуклеотида, предпочтительно длиной не более 50 нуклеотидов, предпочтительно имеющего одну из следующих нуклеотидных последовательностей:

- i) SEQ ID No. 4, 6 или их комплемент, либо
- ii) SEQ ID No. 5, 7 или их комплемент.

10. Вектор, предпочтительно растительный вектор, содержащий молекулу ДНК или молекулу нукleinовой кислоты по любому из пп. 6-8 или молекулу нукleinовой кислоты по п. 9, в котором молекула ДНК или молекула нукleinовой кислоты предпочтительно способна экспрессировать функциональный ген CYPgst в качестве трансгена и предпочтительно генетически связана с другим трансгеном, предотвращающим перенос молекулы ДНК или молекулы нукleinовой кислоты через пыльцу, при этом предпочтительно вектор или трансген также содержит экспрессионную кассету, которая маркирует семена, предпочтительно флуоресцентными метками.

11. Клетка-хозяин, предпочтительно растительная клетка, содержащая рекомбинантную молекулу ДНК или молекулу нукleinовой кислоты по любому из пп. 6-8 или вектор по п. 10.

12. Набор, содержащий молекулу ДНК или молекулу нукleinовой кислоты по любому из пп. 6-8, молекулу нукleinовой кислоты по п. 9, олигонуклеотид по п. 9, вектор по п. 10, белок CYPgst или его функциональный и/или иммунологически активный фрагмент, кодируемый последовательностью ДНК по п. 4, и/или антитело, которое специфически связывается с белком CYPgst или его фрагментом, а также потенциально реагенты,

предназначенные для процедур, основанных на детекции нуклеиновых кислот, или для иммунологической детекции.

13. Способ получения растения, в частности культурного растения, проявляющего гомозиготный рецессивный ядерно-кодируемый мужски стерильный фенотип, в котором экспрессия гена CYPgst ингибируется, отличающийся тем, что он предпочтительно включает этап введения рекомбинантной молекулы ДНК по п. 7, молекулы нуклеиновой кислоты по пп. 8 или 9 или вектора по п. 10, например, посредством агробактериальной трансформации, Т-ДНК мечения, гомологичной рекомбинации, мутагенеза, такого как TILLING, и направленный мутагенез, например, с использованием нуклеаз «цинковые пальцы», TALE-нуклеаз (эффекторные нуклеазы, подобные активаторам транскрипции), а также системы CRISPR/Cas, который приводит к ингибированию экспрессии гена, например, посредством iPРНК или косупрессии, либо вследствие введенной мутации.

14. Способ восстановления фертильности растения по любому из пп. 1-5 или растения, которое можно получать посредством способа по п. 13, включающий введение функционального гена CYPgst в растение, причем предпочтительно ген CYPgst вводится с помощью рекомбинантной ДНК по п. 6 или вектора по п. 10 или путем скрещивания с растением, несущим ген CYPgst дикого типа или функциональный ген CYPgst, предпочтительно в гомозиготном состоянии.

15. Растение, которое содержит растительные клетки по п. 11, и/или которое можно получать согласно способу по пп. 13 или 14.

16. Орган, растительная часть, ткань или клетка растения по любому из пп. 1-5 или 15, либо семена или потомки растения по любому из пп. 1-5 или 15, причем семена или потомки содержат мутацию по любому из пп. 1-5 и/или рекомбинантную молекулу ДНК или молекулу нуклеиновой кислоты по любому из пп. 6-8 или вектор по п. 10,

17. Способ идентификации растения по любому из пп. 1-5 или 15 посредством детекции мутации гена CYPgst или маркера, связанного с мутацией.

18. Использование молекулы ДНК или молекулы нуклеиновой кислоты по любому из пп. 6-8, молекулы нуклеиновой кислоты по п. 9, олигонуклеотида по п. 9, векторов по п. 10, белков CYPgst или их фрагментов, антител по п. 12 и/или наборов по п. 12, предназначенных для идентификации растения по любому из пп. 1-5 или 15, получения рецессивного ядерно-кодируемого мужски стерильного растения, растения с восстановленной фертильностью, гибридного растения, в селекционных программах на устойчивость или для производства семян, а также использование молекулы ДНК по п. 14 или промотора по п. 14 для специфической экспрессии гетерологичных молекул нуклеиновой кислоты в цветках и/или плодах растений.

Фертильный



Стерильный

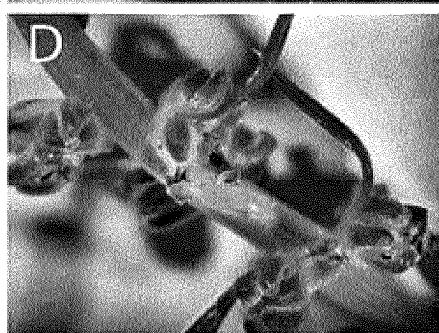


Рис. 1

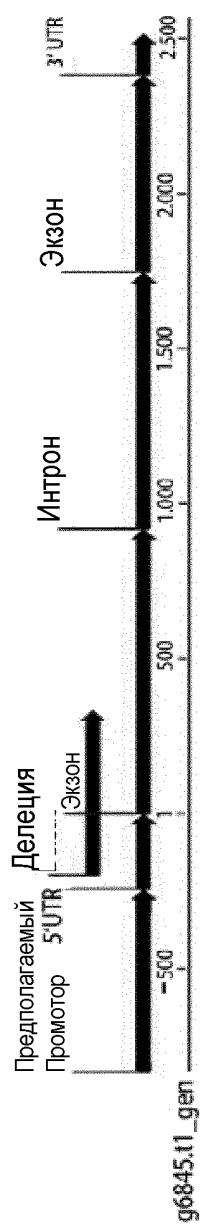


Рис. 2

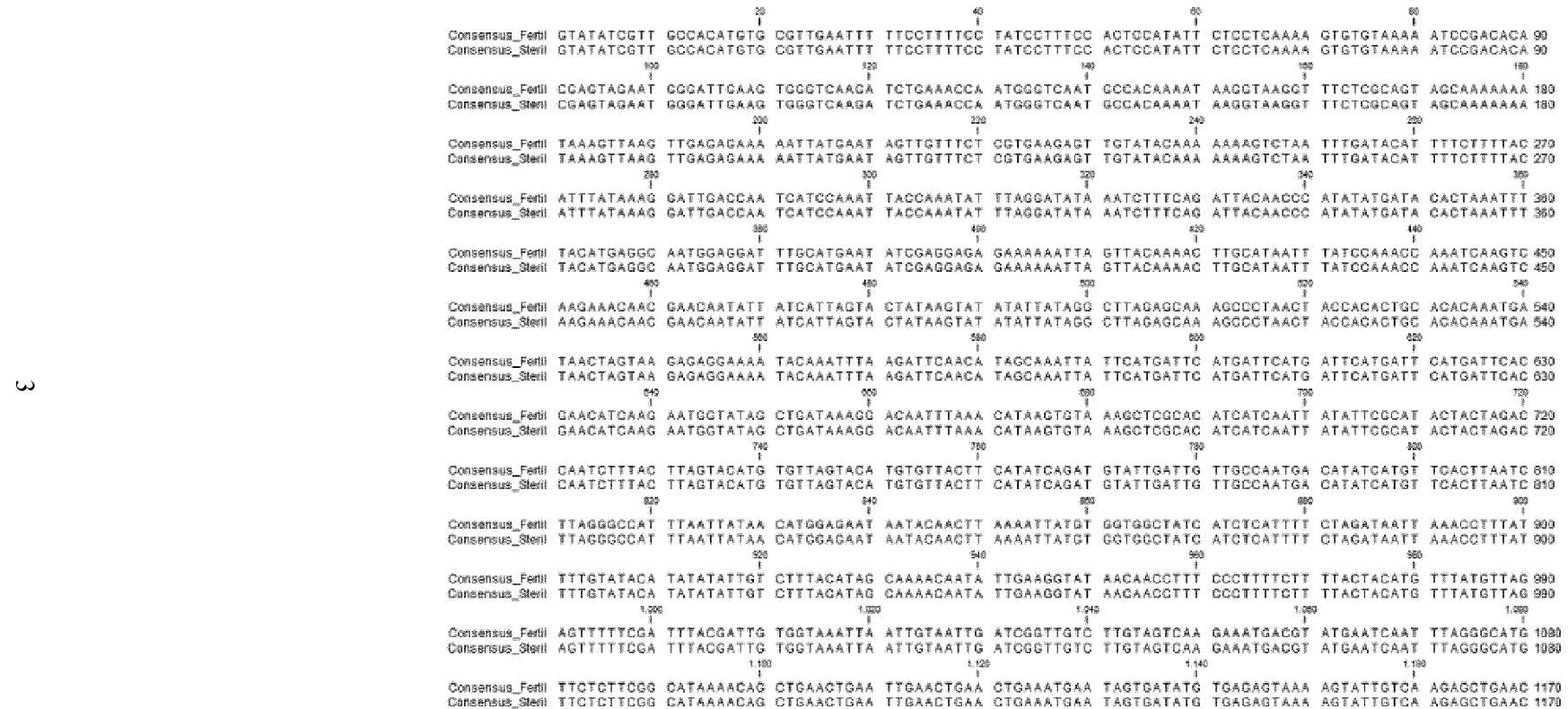


Рис. 3

Рис. 3 (продолжение)

2300 2300 2400 2400
 Consensus_Fertil ACCACAACAA AACCAAAACAA CAAATGCCAC TGAATCAGCC TCAATGGAT TTTGTAGATA TTTCTTCCC TTFCCTGCT GAAAATCGAA 2430
 Consensus_Steril AGGAGAAGAA AAGCAAAAGAA GAAATGGGAG TGAATGAGGG TGAATGGAT TTTGTAGATA TTTCTTCCC TTFCCTGCT GAAAATGGAA 1897
 2400 2400 2400 2400
 Consensus_Fertil ATGAGCATAT GGATGATGCA GATATTAAAG CTCTAATTCA GGTAAATTCA GTATAATTG AATGTGATG ATACAAAGTT TGATAGAAAA 2520
 Consensus_Steril ATGAGCATAT GGATGATGCA GATATTAAAG CTCTAATTCA GGTAAATTCA GTATAATTG AATGTGATG ATACAAAGTT TGATAGAAAA 1983
 2500 2500 2500 2500
 Consensus_Fertil CATATTTCCA TAAATAATG GTTACCTAC TAGACCAAAT AAAATACATA ATTATTGCT TACTAGTTGA AAGTTGAAC AACCTAGCTA 2610
 Consensus_Steril CATATTTCCA TAAATAATG GTTACCTAC TAGACCAAAT AAAATACATA ATTATTGCT TACTAGTTGA AAGTTGAAC AACCTAGCTA 2073
 2600 2600 2600 2600
 Consensus_Fertil CCATTTTGT GTGATTATCA TTAGGCAACC AAAATATT TTTGCGATCCA TATATTAAATG TTGAGAEGCG ATGCGCATA TTTACAATTA 2790
 Consensus_Steril CCATTTTGT GTGATTATCA TTAGGCAACC AAAATATT TTTGCGATCCA TATATTAAATG TTGAGAEGCG ATGCGCATA TTTACAATTA 2158
 2700 2700 2700 2700
 Consensus_Fertil CTTGTAACAT TTTACACAA CAATTAAAAA TATTTTTGCG CAACTCCATT TTATTGATG ATACCTATAT CTTAAATGAA ATCTTCGTC 2790
 Consensus_Steril CTTGTAACAT TTTACACAA CAATTAAAAA TATTTTTGCG CAACTCCATT TTATTGATG ATACCTATAT CTTAAATGAA ATCTTCGTC 2248
 2800 2800 2800 2800
 Consensus_Fertil ATGTACACTT GCCTTCAG GTACCAATAT TTGACCATAT GAAATTACTA TAAACAAATT TGATAAAATC TAATAATATG TAAATATAC 2890
 Consensus_Steril ATGTACACTT GCCTTCAG GTACCAATAT TTGACCATAT GAAATTACTA TAAACAAATT TGATAAAATC TAATAATATG TAAATATAC 2338
 2900 2900 2900 2900
 Consensus_Fertil TTGACGCCAA TATTAGAAC AATGATAATG CAAATAACT TATTGAACT ATGTGTTGAA AGTTAAATTG GGAACAAAG 2870
 Consensus_Steril TTGACGCCAA TATTAGAAC AATGATAATG CAAATAACT TATTGAACT ATGTGTTGAA AGTTAAATTG GGAACAAAG 2428
 3000 3000 3000 3000
 Consensus_Fertil GSTATTTGT ATTGGCGAAT TTAATTATTA AAATTACTTA TAAACAAAC TCAATATGTA AAACGTGTTA GATGGAGCTGT GGATGAAATG 3000
 Consensus_Steril GSTATTTGT ATTGGCGAAT TTAATTATTA AAATTACTTA TAAACAAAC TCAATATGTA AAACGTGTTA GATGGAGCTGT GGATGAAATG 2518
 3100 3100 3100 3100
 Consensus_Fertil AGATGAGTAT ACTTTTACTA GTTACCACTC GAAAATGCA TTCTCTCTT GTTATAGTT GTTCAACTT CTATTATCAT AAATATTCTT 3150
 Consensus_Steril AGATGAGTAT ACTTTTACTA GTTACCACTC GAAAATGCA TTCTCTCTT GTTATAGTT GTTCAACTT CTATTATCAT AAATATTCTT 2890
 3200 3200 3200 3200
 Consensus_Fertil TGGACTTAT TTCAATGAT ATTTCACAAAG GAAATTGTTT AATTTTTTAA AAATACATAA TGAAACAAAG GATTCATGG TCAATTATAC 3240
 Consensus_Steril TGGACTTAT TTCAATGAT ATTTCACAAAG GAAATTGTTT AATTTTTTAA AAATACATAA TGAAACAAAG GATTCATGG TCAATTATAC 2697
 3300 3300 3300 3300
 Consensus_Fertil ACATTATAA TATTATCTTA AAAAACCTTA TAATGCTCAA TTAGTCTCAA TAATATAGGA TATGATGCA GCAGCAACAG ACACATCAGC 3330
 Consensus_Steril ACATTATAA TATTATCTTA AAAAACCTTA TAATGCTCAA TTAGTCTCAA TAATATAGGA TATGATGCA GCAGCAACAG ACACATCAGC 2787
 3400 3400 3400 3400
 Consensus_Fertil TGTAACCAAC GAATGGGCCA TGGCAGAAGT AATAAACAC CCACGTGTCG TCCACAAGAT CCAACAAGAG CTTAACACAA TAGTAGGACD 3420
 Consensus_Steril TGTAACCAAC GAATGGGCCA TGGCAGAAGT AATAAACAC CCACGTGTCG TCCACAAGAT CCAACAAGAG CTTAACACAA TAGTAGGACD 2877
 3500 3500 3500 3500
 Consensus_Fertil CAATCGAATG GTAACGAAAT CAGATGTTCC CCACGTTAAC TACCTACGTT GTGCGTACG TGAACAGTT CGATGCATC GAGCAGGACD 3510
 Consensus_Steril CAATCGAATG GTAACGAAAT CAGATGTTCC CCACGTTAAC TACCTACGTT GTGCGTACG TGAACAGTT CGATGCATC GAGCAGGACD 2967
 3600 3600 3600 3600
 Consensus_Fertil CTTTTTAATC CCACATGAA CACTACGCCA TACAACAAATC AACGGCTATG ATATCCGATC TGGGACACGT GTCTTCATCA ACACACATGG 3600
 Consensus_Steril CTTTTTAATC CCACATGAA CACTACGCCA TACAACAAATC AACGGCTATG ATATCCGATC TGGGACACGT GTCTTCATCA ACACACATGG 3057

Рис. 3 (продолжение)

Consensus_Ferli GTTAGGACGT AACCTTAAAG TGTGGACAA CATAGAGGAT TTTTACCGTG AAAGACATTG GCGCTGGAT GGAAGTAGAG TTGAGATTAG 3890
 Consensus_Stern GTTAGGACGT AACCTTAAAG TGTGGACAA CATAGAGGAT TTTTACCGTG AAAGACATTG GCGCTGGAT GGAAGTAGAG TTGAGATTAG 3147
 Consensus_Ferli CCATGGATCT GATTTTAAAAA TATTACCAATT TAGTCTGGG AAGAGAAGAT GTCCTGGGGC CCCACTTGGG GTGGGTGTTG TGTTGATGGG 3780
 Consensus_Stern CCATGGATCT GATTTTAAAAA TATTACCAATT TAGTCTGGG AAGAGAAGAT GTCCTGGGGC CCCACTTGGG GTGGGTGTTG TGTTGATGGG 3237
 Consensus_Ferli ATGGCTACA CTTTTCATG CATTGATTG CTTACCCACT GATGGAATGA AGCGAGAAGA AATTGATACT AGGAAGTTT ATGGGATGAC 3870
 Consensus_Stern ATGGCTACA CTTTTCATG CATTGATTG CTTACCCACT GATGGAATGA AGCGAGAAGA AATTGATACT AGGAAGTTT ATGGGATGAC 3327
 Consensus_Ferli TATGCCTAA GCTCAACCTT TAATGGCTTT GGCTAACCT AGGCTTGCTC ~~C~~ ATTATACT TCTTGTACAT ATGTTCATAT TGTTGTCAC 3957
 Consensus_Stern TATGCCTAA GCTCAACCTT TAATGGCTTT GGCTAACCT AGGCTTGCTC ~~C~~ ATTATACT TCTTGTACAT ATGTTCATAT TGTTGTCAC 3417
 Consensus_Ferli TTATAACGAC A~~T~~~~T~~~~T~~~~T~~~~T~~ ATAGACAAAT ACAACTTGT ATGACTCTA ACATGTTTT TAATTAAGT ATACTGCAAC TCTACAA~~T~~~~T~~ 4035
 Consensus_Stern TTATAACGAC ~~A~~~~T~~~~T~~~~T~~~~T~~~~T~~ ATAGACAAAT ACAACTTGT ATGACTCTA ACATGTTTT TAATTAAGT ATACTGCAAC TCTACAA~~T~~~~T~~ 3507
 Consensus_Ferli ~~C~~~~G~~CTATGTA AT~~T~~CTATAA AT~~T~~ATATAACA ~~G~~AGCTATAA CGCAT~~T~~TTGT TTGAAAAAAA AGCGGTAC ~~A~~~~T~~~~T~~CTTAC ACCATAAA~~T~~ 4122
 Consensus_Stern ~~C~~~~G~~GTATGTA AT~~T~~CTATAA AT~~T~~ATATAACA ~~G~~AGCTATAA CGCAT~~T~~TTGT TTGAAAAAAA AGCGGTAC ~~A~~~~T~~~~T~~CTTAC ACCATAAA~~T~~ 3595
 Consensus_Ferli GTT~~C~~~~T~~~~T~~~~C~~ TA~~A~~~~T~~~~T~~~~T~~ TTGG~~T~~TATT T~~I~~AGAA~~T~~AA TTTTCAATT TTCTATTT CTATAAAC ~~A~~~~G~~ATAAGAA AGCGACTAA 4211
 Consensus_Stern GTT~~C~~~~T~~~~T~~~~C~~ TA~~A~~~~T~~~~T~~~~T~~ TTGG~~T~~TATT T~~I~~AGAA~~T~~AA TTTTCAATT TTCTATTT CTATAAAC ~~A~~~~G~~ATAAGAA AGCGACTAA 3685
 Consensus_Ferli CAATATTCA CAACATTATA TCTTGAGATC TCGTTGAAT TAATAATTTC CAATAATTTC AAAACGCCAA AGCTTCTTT GTTOGGATAA 4301
 Consensus_Stern CAATATTCA CAACATTATA TCTTGAGATC TCGTTGAAT TAATAATTTC CAATAATTTC AAAACGCCAA AGCTTCTTT GTTCGGATAA 3775
 Consensus_Ferli ATTCACGCTA TTTTCAATAA AATGCCAAA TCTTCCCTTG CCATCAAGGA ACAAA~~T~~TTT CACATCAAA CTCTTCTTCC AAAAGATA 4391
 Consensus_Stern ATTCACGCTA TTTTCAATAA AATGCCAAA TCTTCCCTTG CCATCAAGGA ACAAA~~T~~TTT CACATCAAA CTCTTCTTCC AAAAGATA 3695
 Consensus_Ferli CTTCTCCAA~~T~~ GATTATGTA AAGAGTTGC ATGTTCTAAT ~~C~~ATTAGTAT GTATTTTCG GATAAGAAGT TGTTTACAA GCTTAAATAA 4481
 Consensus_Stern CTTCTCCAA~~T~~ GATTATGTA AAGAGTTGC ATGTTCTAAT ~~C~~ATTAGTAT GTATTTTCG GATAAGAAGT TGTTTACAA GCTTAAATAA 3955
 Consensus_Ferli AGATATAAAC TT~~A~~~~T~~~~T~~~~A~~AT ATCGACCAAT CCACAAATATG TATCTTCATA TGCTGCCATC ~~A~~~~C~~GGCTAAC CTTGAGATGT AGATCATCTA 4571
 Consensus_Stern AGATATAAAC TT~~A~~~~T~~~~T~~~~A~~AT ATCGACCAAT CCACAAATATG TATCTTCATA TGCTGCCATC ~~A~~~~C~~GGCTAAC CTTGAGATGT AGATCATCTA 4045
 Consensus_Ferli AAGCACATTA ACAT~~C~~CGTA CTCCCT~~T~~GT TCA~~T~~~~T~~~~T~~GG GTTGGTTGAA ACC~~A~~ACTTC TCCAAAATAC TTCCCTCAGTA AT~~A~~AGAA~~A~~ 4556
 Consensus_Stern AAGCACATTA ACAT~~C~~CGTA CTCCCT~~T~~GT TCA~~T~~~~T~~~~T~~GG GTTGGTTGAA ACC~~A~~ACTTC TCCAAAATAC TTCCCTCAGTA AT~~A~~AGAA~~A~~ 4135
 Consensus_Ferli AAAC~~T~~ACAT TCTGAAAAAC ATCAACAT~~A~~ TGGTCTATG GTCTAGTGGT TATGACACTG GACTCTGAAT CGAGTAACCC GAGTT 4741
 Consensus_Stern AAAC~~T~~ACAT TCTGAAAAAC ATCAACAT~~A~~ TGGTCTATG GTCTAGTGGT TATGACACTG GACTCTGAAT CGAGTAACCC GAGTT 4220

Рис. 3 (продолжение)

A

TTATTAACCTGATTGGAACCTTATTGAACTTATTAGACCTGATTGGAACCTTATTGCACCTGATTGGAACCTTATTGGAACCTTATTGAACTTATTGCACTTATTAGACCTTATTGCAACTTATCTGAACCTTATCTGAACA
AATCTGAACCTTATTGGACCTGAAACTTAATTTTAAGTTGAACAGAACGCACCCTTAGTATATCGTTGCCACAT
GTGCGTTGAATTTCCTTCCACTCCATATTCTCTCAAAGTGTAAAATCCGACACAG
AGTAGAATGGATTGAAGTGGTCAGATCTGAAACCAATGGGTCAGGACAAAATAAGGTAAAGGTTCTCGC
AGTAGCAAAAAAATAAGTTAAGTGTGAGAGAAAAATTATGAATAGTTGTTCTCGTGAAGAGTTGTATACAAAAA
AAGTCTAATTGATACATTCTTACATTATAAAGGATTGACCAATCATCCAAATTACCAAAATTAGGAT
ATAATCTTCAGATTACAACCCATATATGATACTACTATAATTGAGGCAATTGGAGGATTGCAATGAATAT
CGAGGAGAGAAAAATTAGTTACAAAACCTGATAATTATCCAAACCAATCAAGTCAGAACAAACGAACAT
ATTATCATTACTATAAGTATATATTAGGCTTAGACCAAAGGCCFAACTACCAACTGCAACACAATGATA
ACTAGTAAGAGAGGAAATACAATTTAAGATTCAACATAGCATAATTATTCATGATTGATGATTGATG
ATTCAATTGATTCAAGAACATCAAGAATGGTATAGCTGATAAGGACAATTAAACATAAGTGTAAAGCTCGCACAT
CATCAATTATATTGCAACTACTAGACCAATTCTTACTTAGTACATGTTAGTACATGTTACTTCATATCA
GATGTTATTGATTGTTGCCAATGACATATCATGTTACTTAATCTTAGGCCATTAAATTATAACATGGAGAATAA
TACAACCTTAAATTATGTTGGCTATCATCTCATTTCTAGATAATTAAACCTTATTGTTACATGTTATGTTAG
TTGTTGATTACGATTGTTGGATAATTGTAATTGATCGGTTGTTGAGTCAGAACATGACGTTGAAATC
AATTAGGGCATGTTCTCGGCATAAAACAGCTGAACTGAACTGAACTGAACTGATCTGAAATCTGAACTA
AGAGTTAAAGTATTGTCAGAGCTGAACTGAGCTGAACTGAACTGAACTGATCTGAAATCTGAACTA
ATCTGAACTGAACTGAACTGAAATTGAACTGAAATAAGCTAGGGAAAACACCCCTACTACTATTATAACCT
CGTTTAAATTAGGAAATTAAAAAAATAATTATTTCTTATACCTTATTAAACCTATTGAGGTTTTTATATTG
ACTOCCAACTACTATTATAGATCATGCCAATGAGCAAACACTTCTCACATCTTATAGGAGAAAA
GIAAGATCACICACTAGCATATCATGACCAGOGAAACCAACCGTAAATAGTTTATGCTTCCATTAGAGATA
AGAGTTAACATAATACCATTTGTAATTGATGGATTGGAACCTTACGAAATATTTACTGTTGCA
CTTTTGCTACAAAATTTATACCAATGGCTCAAGTCTACTTACACAAACATACAAACTCCCTCCGGCCA
CCAAGGTGCCCTTATTGAAACCTCCCTCAACTAGGCCACTTCCCACCGCGATTGCGCTCATTTGAA
AAATATGGCCCTTACTGCTACATAAGGCTGGTAATGTTGATGCCATAACCAACTAATGATCCAGAAATCATACGT
GAGATCTTAGTCGACAAGATGATGTTGCGTCTCGTCCACATACCTTAGGCCAACCCACTTGGCTTACAT
AGTGGTGTGGCTTGGCTCAATGGGACCAAAATGGAAAAGAATGAGGATATGCACTGGGACTTGTGTC
ACAACTAGAGCAGTGAACATTGAGCTGAGGCTGATAGGGCTGAGGGCAGCAGACATTGGCCAAGACGTATTAACT
GTTCCACAAAGATAATTGTTAATTGAGGGAGGTTAGGTGCAATTGTTATGATAACTGACTAGAATG
TTGCTAGGGAGCAACTTTGGGCCCCGACGGGGCCCCACAAGAGCTCTAGAGTTATGCAATAACACAT
GAGTTGTTGGTTACTAGGCTTGTATTACTTGGGTTGATTATTGCTTTTGGAGGTGGGTTGATCCATATGGA
TGTAAAAGRAAAATGAGGGAGGTTGAAAAMAGGTAGATGATTCCATCGAAAATTATAAGAGGAACATAGGAAG
GAGAAGAAAAGGAAAGAAGAAATGGGAGTGAATGAGGGTAAATGATTGTTAGATATTGTTGGCTTGCCT
GGTAAAATGAAATGAGCATATGGATGATGCAAGATATTAAAGCTCAATTCAAGTAAATTGAA
TGTGATGATACAAAGTTGATGAAACATATTGCAAAATATGTTGCCCCTACTAGACCCAAATAAAATAC
ATAATTATTGCTTACTAGTTGAAAGTTGAAACAAACCTGACTIACCAATTGTTGATTATCATGCAACCAA
AAITATTCTGATCCATATAATTGTTGAGATCAGAGTOGGCATATTACAAATTACTGTAACATTTAAGC
AAACAAAATAAAATTTTGGCAAGTCCATTGAAATAACCTTATCTTAAATGAAATTCTGTTGCAATTAACT
GTACACTTGCCCTTCAAGGCTACCATATTGACCATATGTAATTACTTAAACAAATTGATAAAATCTAATTAAT
ATGTAATAATACATTGACCATATTAGAAACAAAGGATCACAAATGATAATGCAAAAATAACTTATTGAACTAA
TGTGTAAGCTAAATTGGAACAAAAGGTATATTGTTGCAATTGCGAATTTTAATTATAAAATTACTTATAACAC
AACTCAATAATGTTAAAGATGGAGTGTGGATAGATGAGATGAGTATACTTTACTAGTTACCACTCGA
AAATGCAATTCCCTCCTTATTAGTTGCTTAACCTCTATTATCATAAAATAATTGTTGGACTTATTGCAATG
TATATTACACGCTAATTGTTAATTGTTAAATTGTTAAACATAATGTAACACAGGATTTCATGGTCAATTATACAC
ATTATAAAATTATCTTAAACAAACTTATTATGCTCAATTGATCCATAATAATAGGATAATGATAGCAGCAGCAA
CAGACACATCAGCTGTAACCAACGAATGGGCCATGGCAGAAGTAATAAAACACCCACCGTGTCCCTCACAAGATCC
ACAAGAGCTAACACAAATGAGGACCCATCGAATGGTAACAGAACGATCTTCCCCACCTTAACCTAC
GTTGTTGTCGTACGTGAAACGCTTCCGAATGCACTCCAGCAGGACCCCTTTTAAATCCACATGAAATCACTACGGCCATA
CAACAACTCAACGGCTATGATATCCCATCTGGGACACGGTGTCTCATCAACACACATGGGTAGGACGTAACCTTA

Рис. 4

AAGTGTGGGACAACATAGAGGATTTACCCTGAAAGACATTGGCGTGGATGGAAGTAGAGITGAGATTAGCC
ATGGGATCTGATTTAAAATATACCATTACTGCTGGGAAGAGAAGATGTCTGGGCCCACTTGGGGTGGTGT
TTGTGTTGATGGGATTGGCTACACTTTTCACTGCAATTGATTTGTTACCCACTGATGGAATGAAGGCAGAAGAAA
TTGATACTAAGGAAGTTATGGGATGACTATGCTAAAGCTAACCTTTAATGGCTTGGCTAAACCTAGGCCTIG
CTCATTTATATCTTGTATACATGTTCAATATGTGGTGACITATAAGCACAATAGACAAATACAAGTTGTATCG
ACTCTAACATCTTGTATTAGTATACTGCAACTGACAGTATGTAATTCTATAACTATAAACACAAGT
CATAACGCATTGTTGAAAAAGAGGTTACATTGCTTACACCACTAAA

B

TTAAAAAAATTAATTATTTCTTATACCTTATTAACCATTGAGGTTTATATTGACTCCAAATACTATTT
ATAGATCATGCCATGTTAATGACCAAACACTCTTCTCACACTTTTATAGGAGAAAAGTAGATCACTCACTAGCA
TATCATGACCRAGCAGAAACCAACCAACGCTTAATAGTTTATGCTTCCATTAGAGATAAGGTTAACTAATATAC
CATCTTGTGAAATTTGTGGATTTGGAACTTTAGCAATATATTACTGTGTGCACTTTGCTACCAAAATT
TTATACCAATGGCTCAAGTCTACTTACACACATAACAAACTCCCTCCGGCCACCAAGGTGGCCCTTATTT
GGAAACCTCCTCAACTAGGGCCACTTCCCACCGCGATTGCGCTCATTTGTGAAAAAATATGGGCTTGTAGTC
TACATAAGGCTGGTAATGGGATGCCATAACCTAATGTCAGAAATCATCGTGGAGATCTAGTTCGACAA
GATGATGTTATTGCGTCTCGTCCACATACCTTAGCGCGAACCCACTTGGCTTACAATAGTGGTGTGTC
GCTCCAATGGGACCAAAATGAAAAGAATGAGAAGGATATGCATGGAGCACTTGTCTACAACAGACGTTGAA
CTATTGTAAGTGTGAGTCATAGGGTGTGAGGGCACGACATTGGTCCAAGACGTATTAACCTGGTCCACAAAGATAAA
GTTGTTAATTGAGGGAAAGTGTAGGTGCTTATGAAATAACGTGACTAGAAATGTTGCTAGGGAAAGCAATAC
TTTGGGGCCGGGACGGGGCCCAAGAGGCTCTAGAGTTATGCTATATAACACATGAGTTGGTTACTA
GGCTTGATTTACTGGGTGATTATTGCTTTGGAGGTGGGTTGATCCATATGGATGTGAAAGAAAATGAGG
GAAGTTGAAAAAGGGTAGATGATTTCATCGCAAAATTATAGAGGAACATAGGAAGGAGAAGAAAAGGAAAGAA
GAAATGGGAGTGAATGAGGGTGAATGGGTTGAGATTTGTTGCGCTTGCCTGGTGAAGAATGGAAATGAG
CATATGGGATGTGAGATTTAAAGCTCTAATTCAAGGATATGAGACAGCAGCAACAGACATCAGCTGIAACC
AACGAATGGGCAATGGCAGAAGTAATAAACACCCACGGTGTGCTCCAAAGATCCAACAGAGCTTAACACRATA
STAGGACCCAACTGAATGGTACAGAAATCAGATCTCCCACATGAAATCACTACGCCATACAACATCAACGGCTATGAT
ATCCCCATCTGGGACACGTGCTICAICAAACACACATGGGTTAGGACGTAACCTTAAAGTGTGGGACAAACATAGAG
GATTTTACCTGAAAGACATTGGCGTGGATGGAAGTAGAGTTGAGATTAGGCCATGGATCTGATTAAAAAA
TTACCATTTAGTGTGGAGAGAAGATGCTCTGGGGCCCCACTTGGGGGGGTTGTTGIGTTGATGGGATTGGCT
ACACTTTTCTGCAATTGATGGTTACCCACTGATGGAATGAAGGCAGAAGAAAATTGATACTAAGGAAGTTAT
GGGATGACTATGCTAAACCTTTAATGGCTTGGCTAAACCTAGGCTTGTCTCATTTATCTTGTATAC
ATGTTCTATATGTGGTGCACTTATAAGCACAATACAAGTTGTATCGACTCTAACATGTTGTTAGT
ATTAGTATACTGCAACTCTACAAGTATGTAATTCTATAACTATAAACACAAGT

C

MDFGTIAIYLLCALFATKILYQNLKSYLYTTYKLPPGPRWPLFGNLLQLGPLPHRDFASFCEKYGPLVYIRLGN
VDAITTNDPEIREILVRQDDVFASRPHTLAATHLAYNSGDVALAPMPKWKMRMRICMEHLLTRRLLELFVSHR
ADEARHLVQDVLRSHKDVKVNLREVLGFASMNNTVRLMLLGKQYFGAGITAGPQEAELEMHITHELFWLLGLIYLG
DYLPGFWRNVDPYGCEKKMREVEKRVDDFRKIIIEHHRKEKKRKEEMGVNEGEMDFVDILLALPGENGNEHMDDAD
IKALIQDMIAATDTSAVINEWAMAEVIKHPVRLHKIQQELNTIVGPNRMVTESTDLPHLNLYLRCVVRETFRMHPA
GPFLIPHESLRHTTINGYDIPSGTRVFINTHGLGRNLKVWDNIEDFYPERHWPLDGSRVEISHGSDPKILPFSAG
KRRCPGAPLGVVFVLMGLATLFHAFDWLPPDGMAEEIDTKEVYGMTPKAQPLMALAKPRLAHLYL

Рис. 4 (продолжение)

D

GTATATCGTTGCCACATGTGCGTTGAATTTCCTTCCACTCCATATTCTCTCAAAAGTGTG
TAAAAATCCGACACACGAGTAGAATGGGATTGAAGTGGGTCAAGATCTGAAACCAATGGTCATGCCACAAAT
AAGGTAAAGGTTCTCGCACTAGCAAAAAAATAAGTTAGTTAGAGAAAAAATTATGAATAGTTGTTCTCGTGA
AGAGTTGTATACAAAAAAAGTCTAATTGATACATTTCCTTACATTATAAAGGATTGACCAATCATCCAAT
TACCAAATATTAGGATATAATCTTCAGATTACAACCCATATATGATACACTAAATTACATGAGGCAATGG
AGGATTGTCATGAATATCGAGGAGAGAAAAAATTAGTTACAAAACTGCATAATTATCCAAACCAATCAAGTC
AAGAAACAACGAACAATATTATCATTAGTACTATAAGTATATAATTAGGCTTAGAGCAAAGCCCTAACACCAC
ACTGCACACAAATGATAACTAGTAAGAGGAAATACAATTAAAGATTCAACATAGCAAATTATTCATGATTC
ATGATTCTGATTCATGATTCACTGATTCAAGAACATCAAGAATGCTAGTGTAAAGGACAATTAAACATAA
GTGTAAGGCTCGCACATCTCAATTATTCGCACTACTAGACCAATCTTACTTAGTACATGTTAGTACA
TGTGTTACTTCATGATGATTGATGTTGCAATGACATATCATGTTCACTTAATCTTAGGGCCATTAAAT
TATAACATGGAGAATAATAACCTTAAATTATGTGGCTCATCTCATTTCTAGATAATTAAACCTTAT
TTTGTATACATATATATTGCTTACATAGCAAACAAATATTGAAGGTATAACACCTTCCCTTTCTTACT
ACATGTTATGTTAGGTTTCGATTTACGATTGGTAAATTAAATTGTAATTGATCGGTTGTTGTTAGTCAA
GAAATGACGTATGAATCAATTAGGGCATGTTCTTCGGCATAAAACAGCTGAACGTGAAATTGAACTGAA
ATGAATAGTGTATGTGAGAGTAAAGTATTGTCAAGAGCTGAACGTGAACTGAACGGATCTGAACGT
TGATCTAATCTGAACTAATCTGAACTGAACCTGAATTGAACCTAGGAAACAGACCCCTA
CTACTATTATATAACCTCTTAAATATTAGGAAATTAAAAAATTATATTCTTTACTTTATTAAACCTA
TTATACAATAGTGGTGTGGCTTGGCTCAATGGGACCAAAATGGGAAAAGAATGAGAAGGATATGCGATGGAG
CACTTGCTCACAACTAGACGACTTGAACATTGTGAGTCATAGGCTGTAGGGCACGACATTGGTCCAAGAC
GTATTAACTCGTCCCACAAAGATAAAGTGTAAATTGAGGGAAAGTGTAGGTGATTTCTATGAATAACGTG
ACTAGAATGTTCTAGGGAAAGCAATACTTGGGGCCCCACGGGGCCCCACAAGAGGCTCTAGAGTTATGAT
ATAACACATGAGTTGTTACTAGGCTTGTATTACTTGGGTGATTATTCGCTTTTGAGGGTGGGTGAT
CCATATGGATGTGAAAAGAAAATGAGGGAAAGTTGAAAAAAAGGGTAGATGATTCCATCGAAAATTATAGAGGAA
CATAGGAAGGAGAAGAAAAGGAAAGAAGAAATGGGAGTGAATGAGGTGAAATGGATTGTAGATATTGTTG
GCTTGCCTGGTAAAATGGAAATGAGCATATGGATGATGCAAGATAATTAGCTAATTCAAGTAATTCACTGA
TAATTGAAATGTGATACAAAGTTGATAGAAAACATAATTGCAATAACATGGTACCCACTAGACCCAATAA
AAATACATAATTATTCGCTTGTCTAGTTGAAAGTTGAAACAAACCTAGCTTACATTGTTGATTATCATTAGCCA
ACCAAAACTATTCTGCACTCATATAATTGTTGAGAGTCGCCATTATTAACATTACTGTAAACATTIGAGC
AAACAAATTAAATATTGCAAGTATTGATGATACCTATATCTTAAATGAATCCCTTGGTTAT
GTACACTTGCCCTTCAAGGTACCAATTGAGCTAAATGTAATTACTATTAAACAAATTGATAAAATCTAATAAAT
ATGTAATATACGTTCACGGCACATATTGAAACAAAGATCACAAATGATAATGCAAAATAACTTATTGAACTAA
TGTTGTGAAAGTTAAATTGGAACAAAAGGTATATTGTTATTGCGGAAATTAAATTAAATTACTTATAACAAAC
AACTCAATATGTAACACTGTTAAGATGGAGTGTGGATAGATGAGGTGAGTACTTITACTAGTTACCACTCGA
AAATTGCAATTCCCTTGTGTTATAGTTGTTCTAATTCTATTATCAAAATAATTITGGACTTATTCAATGT
AAJTTTACAACGTTAATGTTAAATTAAATACATAATGTAACAAAGGATTCAAGGTCATGGTCAATTACACAA
TTATAGATATTCTTAAACAAACTTACTAATGCTCAATTAGTGTCTATAATTAGGATATGATAGCAGCAGCAAC
AGACACATCGCTGTAACCAACGAATGGCCATGGCAGAAAGTAAACACCCACCGTGTCCCTCACAAGATCCA
ACAAGAGCTAACACAATAGTAGGGACCAATGGAATGTAACAGAATCAGATCTTCCCACCTTAACCTACAG
TTGTTGCTTACGGCAAACGTTTGGGATGCTTACATCCAGCAGACGGCCCTTTTAATCCCATGAAATCATTACGCCATAC
AAACAAICACGGCTATGATAATCCCATCTGGACACGTTCTCATCACACACACATEGGTTAGGAGCTIAACCTTAA
AGTGTGGGACAAACATAGAGGGATTTTACCCCTGAAAGACATTGGGCCCTGGATGGAAAGTAGTGTCTGGGGCCCCACTTGGGGTGGTGT
TGGATCTGATTTAAATTAACATTAGTGTGCTGGGAAAGAGAAGATGTCCTGGGGCCCCACTTGGGGTGGTGT
TGTGTTGATGGGATTGGCTACACTTTTCTGATTTGATTTGTTACCCACTGATGGAATGAAGGCAGAAGAAAT
TGATACAAAGGAAGTTATGGGATGACTATGCTAAAGCTCAACCTTAAATGGCTTGGCTAAACCTAGGCTTGC
TCCCTCATTTATATCTTGTGATCATGTTGATGTTGTTGACTTAAAGCACACTGCTCAAATAGACAAATCA
AGCTTGATTTGACTCTAACATGTTATTAAATTAGCATACTGCAACTCTACAACAAAGTATGTAATCTCTAT
AAATTATAACATAAGTCATAACGCATTTGTTGAAAAAAAGCTACACTTTCTTACACCAAAACTGTTCC
GTACTAAATTCTCTGCTTATTTGAGATTAAATTTCATTTCTATTACTTATAACACAGATAAGAA
AAATCGACTAACAAATATTCAACACATTATCTGATAGTTGTTGAAATTAAATTTCATTTCAATAATTTCACAAAC

Рис. 4 (продолжение)

GCAAAAAGCTCTTGGATAATTACGTCAATTCAATAAAATCGCAAAATCTCCCTGCCATCAAGGA
ACAATTTCACATCAAACACTCTCTCCAAAAGTATACTTCTCATTGATTATGAAAAGAGTTGCATGTT
CTAACCATAGTATGTATTCGGATAAGAAGTTGATTACAGCTAAATAAAGATATAAAGCTTAGTTAAT
ATCGACCAATCCACAATATGTATCTCATATGTGTCCATCAGCGCTAACCTTGAGATGTAGATCATCTAAAGCA
CATTAACATGGCGTACTCCTCCTGTCATCAAATCAGGGTGTGAAACCAAACCTCTCCAAAATACTCCTCAGTA
ATGAAAGAGAAAAACTCACATTCTGAAAACATCAAACATCATGGTTATGGTCTAGGGTTATGACACTGGACTC
TGAATCCAGTAACCCGAGTT

Рис. 4 (продолжение)

